

# Calibration for Improving the Medium-Range Soil Forecast over Central Tibet: Effects of Objective Metrics' Diversity

Yakai Guo <sup>1,2,\*</sup>, Changliang Shao <sup>3,\*</sup>, Guanjun Niu <sup>4</sup>, Dongmei Xu <sup>2</sup>, Yong Gao <sup>5</sup> and Baojun Yuan <sup>1</sup>

<sup>1</sup> Henan Meteorological Bureau, China Meteorological Administration, Zhengzhou 450003, China

<sup>2</sup> Key Laboratory of Meteorological Disaster (KLME), Ministry of Education and Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), Nanjing University of Information Science and Technology (NUIST), Nanjing 210044, China; dmxu@nuist.edu.cn

<sup>3</sup> Meteorological Observation Centre, China Meteorological Administration, Beijing 100081, China; shaocl@cma.gov.cn

<sup>4</sup> Meteorological Development and Planning Institute, China Meteorological Administration, Beijing 100081, China; guanjun235@126.com

<sup>5</sup> Tibet Meteorological Observatory, China Meteorological Administration, Lhasa 850000, China; gy\_ynu2024@163.com

\* Correspondence: guoykhmb@126.com; shaocl@cma.gov.cn (C.S.); Tel.: +86-16603-990961 (Y.G.); +86-18210-986639 (C.S.)

**Abstract:** The high spatial complexities of soil temperature modeling over semiarid land have challenged the calibration–forecast framework, whose composited objective lacks comprehensive evaluation. Therefore, this study, based on the Noah land surface model and its full parameter table, utilizes two global searching algorithms and eight kinds of objectives with dimensional-varied metrics, combined with dense site soil moisture and temperature observations of central Tibet, to explore different metrics' performances on the spatial heterogeneity and uncertainty of regional land surface parameters, calibration efficiency and effectiveness, and spatiotemporal complexities in surface forecasting. Results have shown that metrics' diversity has shown greater influence on the calibration–predication framework than the global searching algorithm's differences. The enhanced multi-objective metric (EMO) and the enhanced Kling–Gupta efficiency (EKGE) have their own advantages and disadvantages in simulations and parameters, respectively. In particular, the EMO composited with the four metrics of correlated coefficient, root mean square error, mean absolute error, and Nash–Sutcliffe efficiency has shown relatively balanced performance in surface soil temperature forecasting when compared to other metrics. In addition, the calibration–forecast framework that benefited from the EMO could greatly reduce the spatial complexities in surface soil modeling of semiarid land. In general, these findings could enhance the knowledge of metrics' advantages in solving the complexities of the LSM's parameters and simulations and promote the application of the calibration–forecast framework, thereby potentially improving regional surface forecasting over semiarid regions.

**Keywords:** metrics diversity; Kling–Gupta efficiency; soil temperature modelling; spatial complexity; land surface parameters

## 1. Effects on Optimal Parameters

### 1.1 Spatial Heterogeneity

The optimal parameters of different sites for different types and optimizers are shown in Figures S1-1. For CCS, only the SBETA of the “General” type in the PSO's optimal parameter space seem to be less sensitive to sites (or spatial homogeneities) as indicated by fewer sites crossing with the upper reference limit (i.e., 0.67) when compared to other land parameters. Nevertheless, when compared to most parameters of the “Initial” type in the SCE's optimal parameter space (except STC1), all other parameters are less sensitive to sites as indicated by fewer crossing with the lower reference limit (i.e., 0.33).

Generally, 1 and 40 parameters are less sensitive to sites for PSO and SCE respectively. This indicates that PSO calibrations with CCS have larger parameter heterogeneity when compared to SCE.

For EKGE, the QTZ of the “Soil” type, the (SBETA, CZIL) of the “General” type, and the (STC1, SH2O1, SMC1, ALBEDO) of the “Initial” type in the PSO’s optimal parameter space seem to be less sensitive to sites compared to other land parameters. Meanwhile, the (QTZ, WLTSMC, SATDK) of the “Soil” type, the (SBETA, CZIL) of the “General” type, and the (STC1, SH2O1) of the “Initial” type in the SCE’s optimal parameter space are less sensitive to sites. Generally, 7 and 7 parameters are less sensitive to sites for PSO and SCE respectively. This indicates that PSO and SCE calibrations with EKGE have almost the same parameter heterogeneity.

For EMO, the (QTZ, SATDK) of the “Soil” type, the (SBETA, CZIL) of the “General” type, and the (STC1, SMC1, SMC2) of the “Initial” type in the PSO’s optimal parameter space seem to be less sensitive to sites compared to other land parameters. Meanwhile, the (QTZ, SATPSI) of the “Soil” type, the (SBETA, CZIL) of the “General” type, and the (SH2O1, SMC4) of the “Initial” type in the SCE’s optimal parameter space are less sensitive to sites. Generally, 7 and 6 parameters are less sensitive to sites for PSO and SCE respectively. This indicates that PSO calibrations with EMO have smaller parameter heterogeneity compared to SCE.

For MAES, the QTZ of the “Soil” type, the SBETA of the “General” type, and the (SMC3, SMC4) of the “Initial” type in the PSO’s optimal parameter space seem to be less sensitive to sites compared to other land parameters. Meanwhile, the QTZ of the “Soil” type, the SBETA of the “General” type, and the SMC4 of the “Initial” type in the SCE’s optimal parameter space are less sensitive to sites. Generally, 4 and 3 parameters are less sensitive to sites for PSO and SCE respectively. This indicates that PSO calibrations with MAES have smaller parameter heterogeneity compared to SCE.

For NSES, the (SMC3, SMC4) of the “Initial” type in the PSO’s optimal parameter space seem to be less sensitive to sites compared to other land parameters. Meanwhile, all the SCE’s optimal parameters have shown greater heterogeneities. This indicates that PSO calibrations with NSES have smaller parameter heterogeneity compared to SCE. However, for both PKGE and PMO, all the optimal parameter of both PSO and SCE have shown greater heterogeneities.

For RMSES, the (QTZ, SATPSI) of the “Soil” type, the SBETA of the “General” type, and the (SMC3, SMC4) of the “Initial” type in the PSO’s optimal parameter space seem to be less sensitive to sites compared to other land parameters. Meanwhile, the QTZ of the “Soil” type, and the SBETA of the “General” type in the SCE’s optimal parameter space are less sensitive to sites. Generally, 5 and 2 parameters are less sensitive to sites for PSO and SCE respectively. This indicates that PSO calibrations with RMSES have smaller parameter heterogeneity compared to SCE.

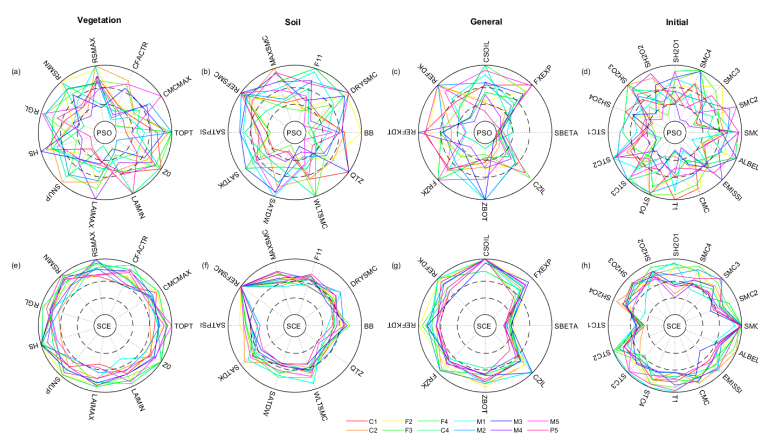


Figure S1-1-1. For the metric CCS, the optimal vegetation (a), soil (b), general (c), and initial (d) land surface parameters of the twelve sites for the PSO optimizer. (e–h) are the same as (a–d), but for the

SCE optimizer. The two dashed circles represent the reference limits of 0.33 (inner) and 0.67 (outer) in the parameter space.

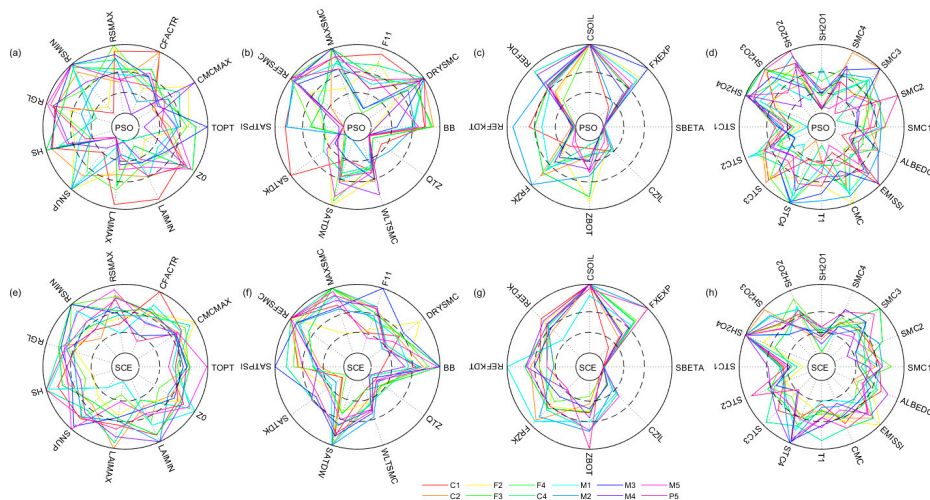


Figure S1-1-2. The same as Figure S1-1-1, but for the metric EKGE.

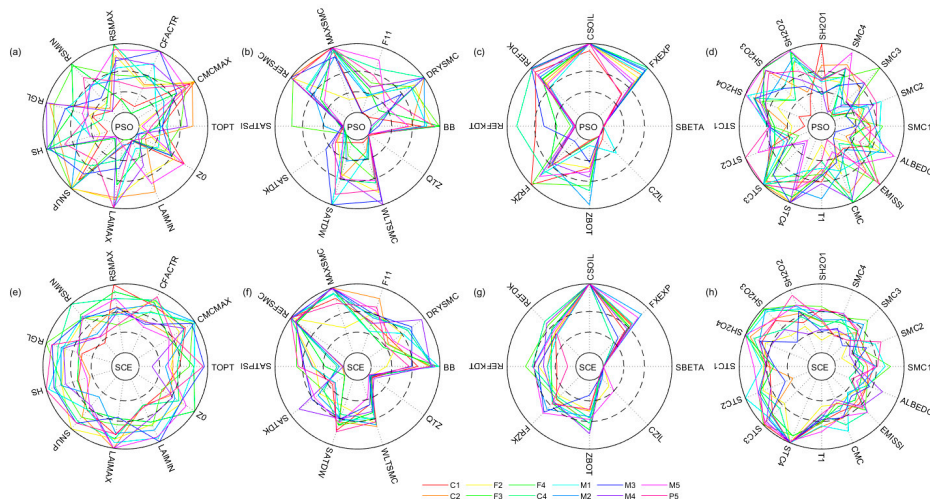


Figure S1-1-3. The same as Figure S1-1-1, but for the metric EMO.

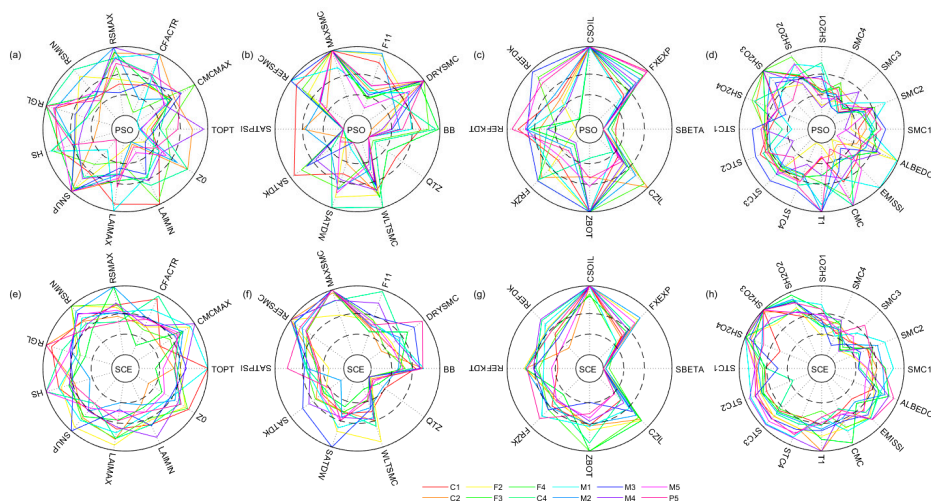


Figure S1-1-4. The same as Figure S1-1-1, but for the metric MAES.

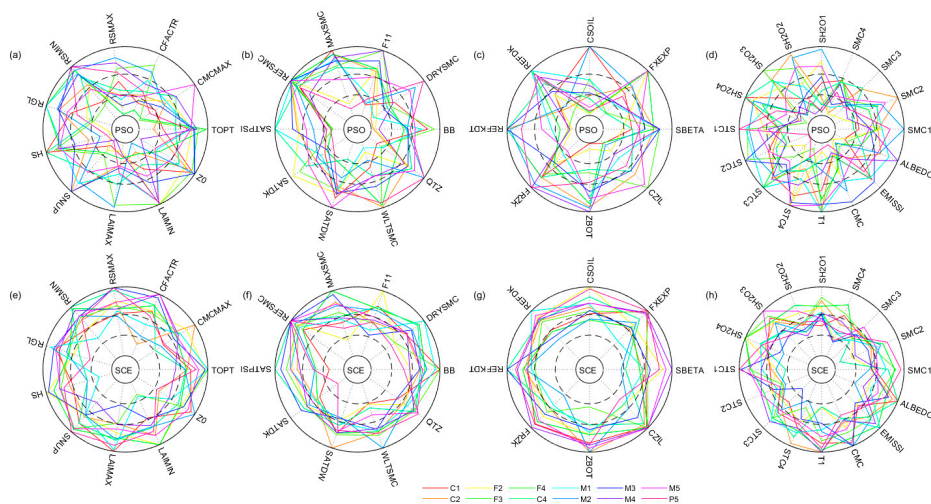


Figure S1-1-5. The same as Figure S1-1-1, but for the metric NSES.

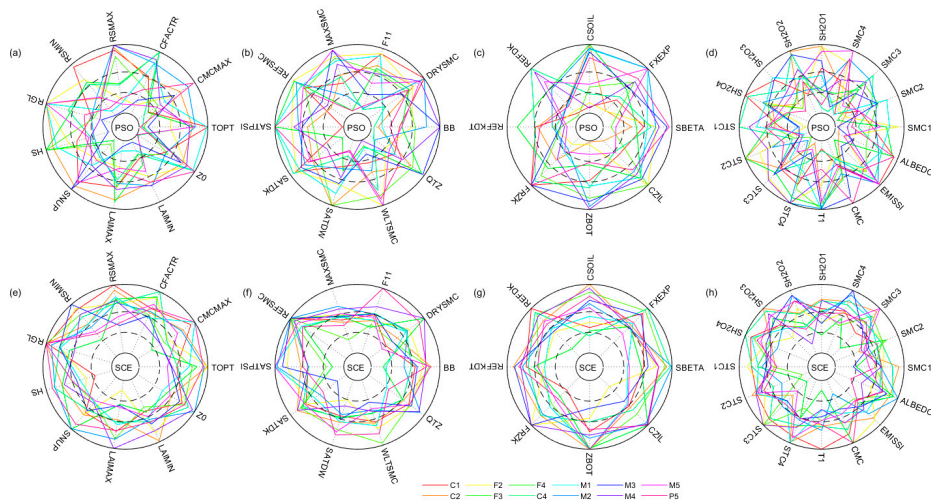


Figure S1-1-6. The same as Figure S1-1-1, but for the metric PKGE.

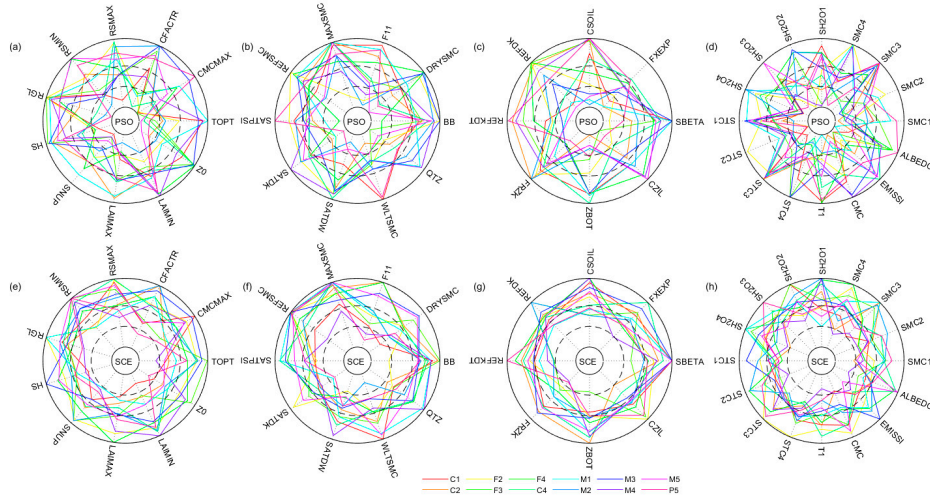


Figure S1-1-7. The same as Figure S1-1-1, but for the metric PMO.

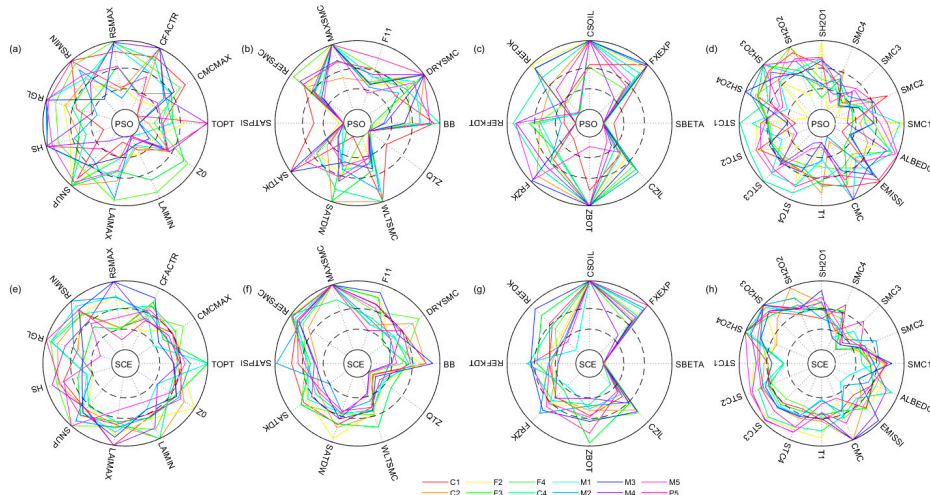


Figure S1-1-8. The same as Figure S1-1-1, but for the metric RMSES.

## 1.2 Spatial Uncertainty

The optimal parameter ranges and outliers against sites compared between PSO and SCE are shown in Figures S1-2. For CCS, except the SBETA parameter of the “General” type, all other PSO optimal parameters have larger ranges than SCE. Nevertheless, the optimal parameters of SCE likely have more outliers than those of PSO.

For EKGE, except the CFACTR and LAIMIN parameters of the “Vegetation” type, the WLTS MC and QTZ parameters of the “Soil” type, the SBETA and CSOIL parameters of the “General” type, and the SMC1 and T1 parameters of the “Initial” type, all other PSO optimal parameters have larger IQR ranges than SCE. Nevertheless, PSO also has more outliers than SCE in the “Vegetation”, “Soil” and “General” types, while this behaves oppositely in the “Initial” type.

For EMO, except the MAXSMC, SATPSI, and QTZ parameters of the “Soil” type, the SBETA, CSOIL, and CZIL parameters of the “General” type, and the SH2O1 parameter of the “Initial” type, all other PSO optimal parameters have larger IQR ranges than SCE.

Nevertheless, PSO also has more outliers than SCE in the “Soil”, “General” and “Initial” types, while this behaves oppositely in the “Vegetation” type.

For MAES, except the MAXSMC and QTZ parameters of the “Soil” type, and the SBETA, CSOIL and CZIL parameters of the “General” type, all other PSO optimal parameters have larger IQR ranges than SCE. Nevertheless, SCE has more outliers than PSO in all parameter types.

For NSES and PKGE, all PSO optimal parameters have larger IQR ranges than SCE, while SCE has more outliers than PSO in all parameter types. For PMO, except the WLTSML parameter of the “Soil” type, all other PSO optimal parameters have larger IQR ranges than SCE. Nevertheless, SCE has more outliers than PSO in the “General” type, while this behaves oppositely in the “Soil” type.

For RMSES, except the F11, MAXSMC, SATPSI and QTZ parameters of the “Soil” type, and the SBETA, CSOIL, and ZBOT parameters of the “General” type, all other PSO optimal parameters have larger IQR ranges than SCE. Nevertheless, SCE has more outliers than PSO in the “Vegetation” and “General” types, while this behaves oppositely in the “Soil” and “Initial” types.

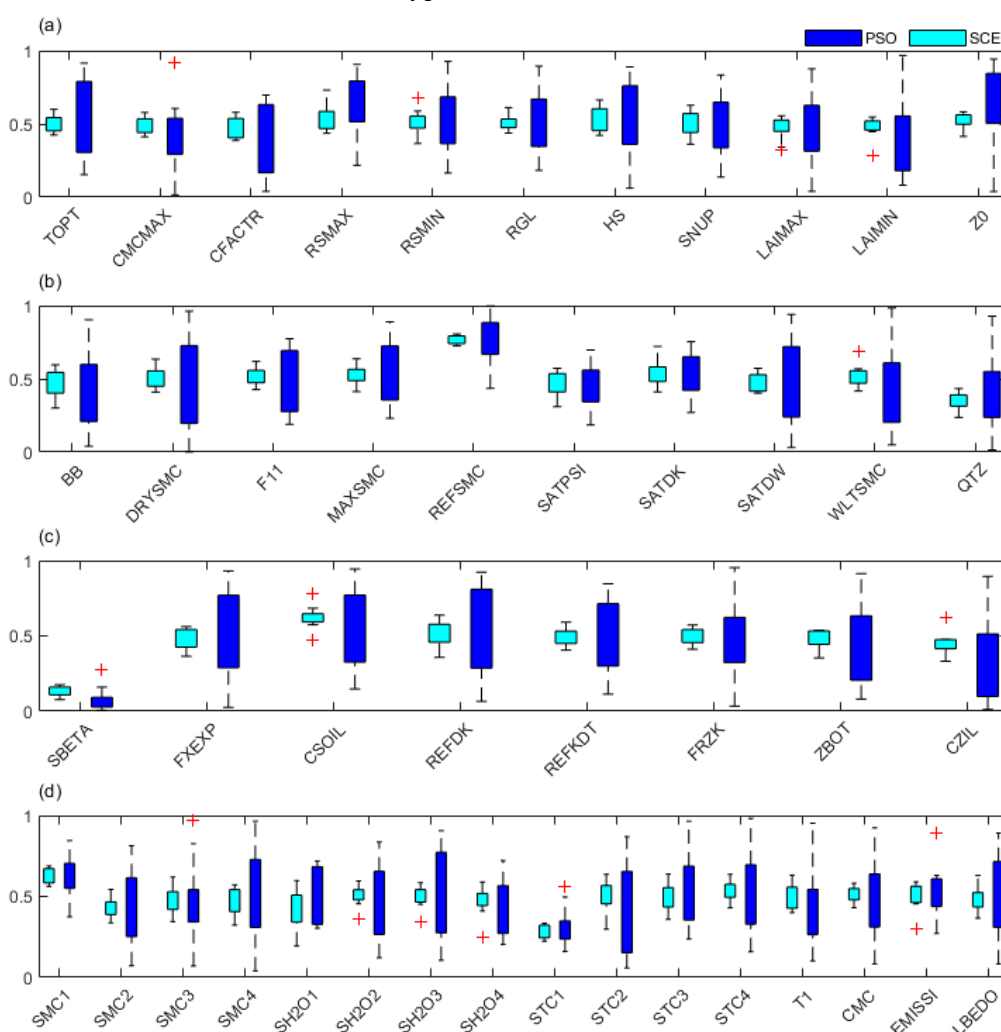


Figure S1-2-1. For the metric CCS, boxplot of optimal vegetation (a), soil (b), general (c), and initial (d) land surface parameters against stations (SCE in light blue and PSO in dark blue). Red crosses indicate outliers.

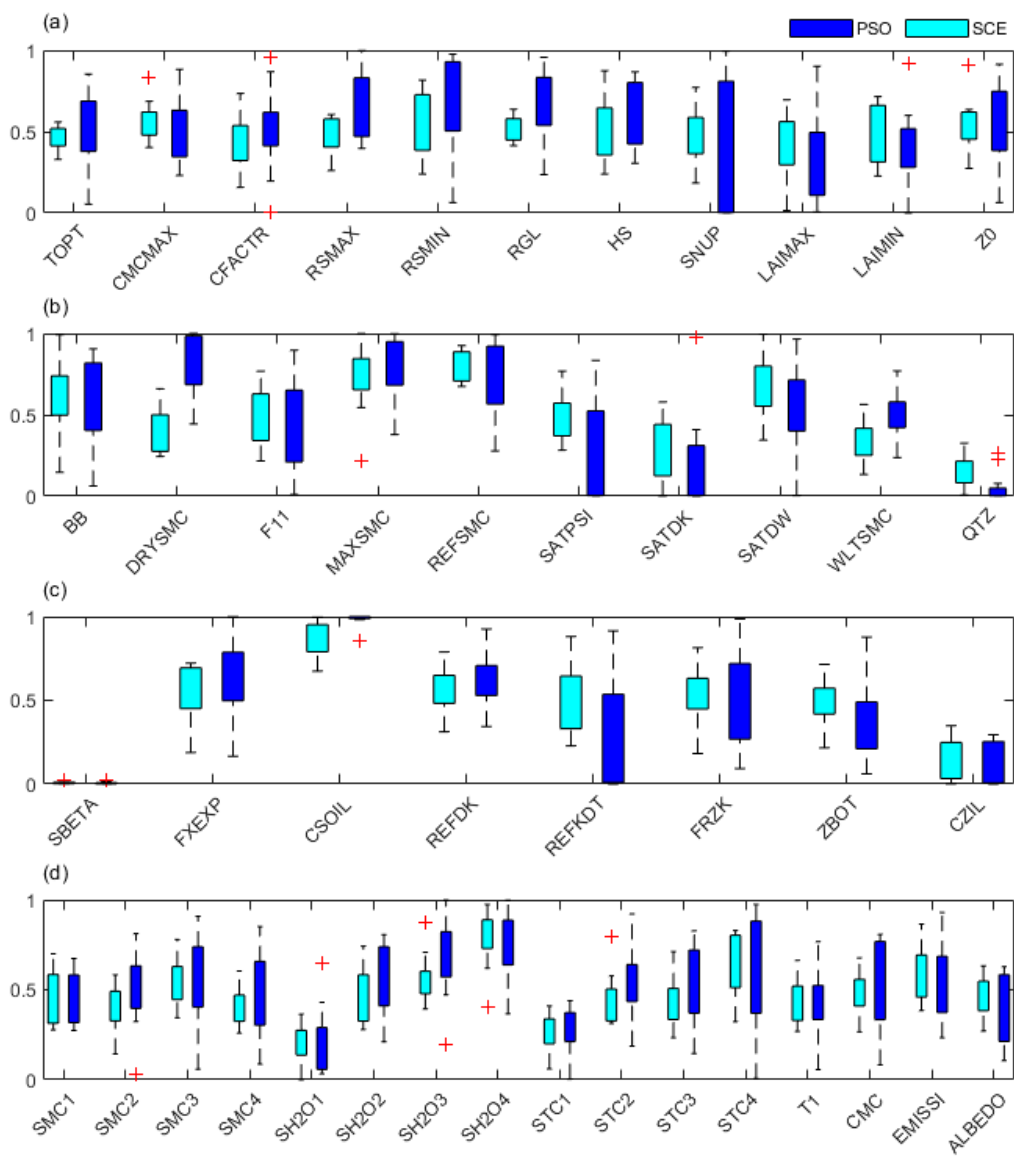


Figure S1-2-2. The same as Figure S1-2-1, but for the metric EKGE.

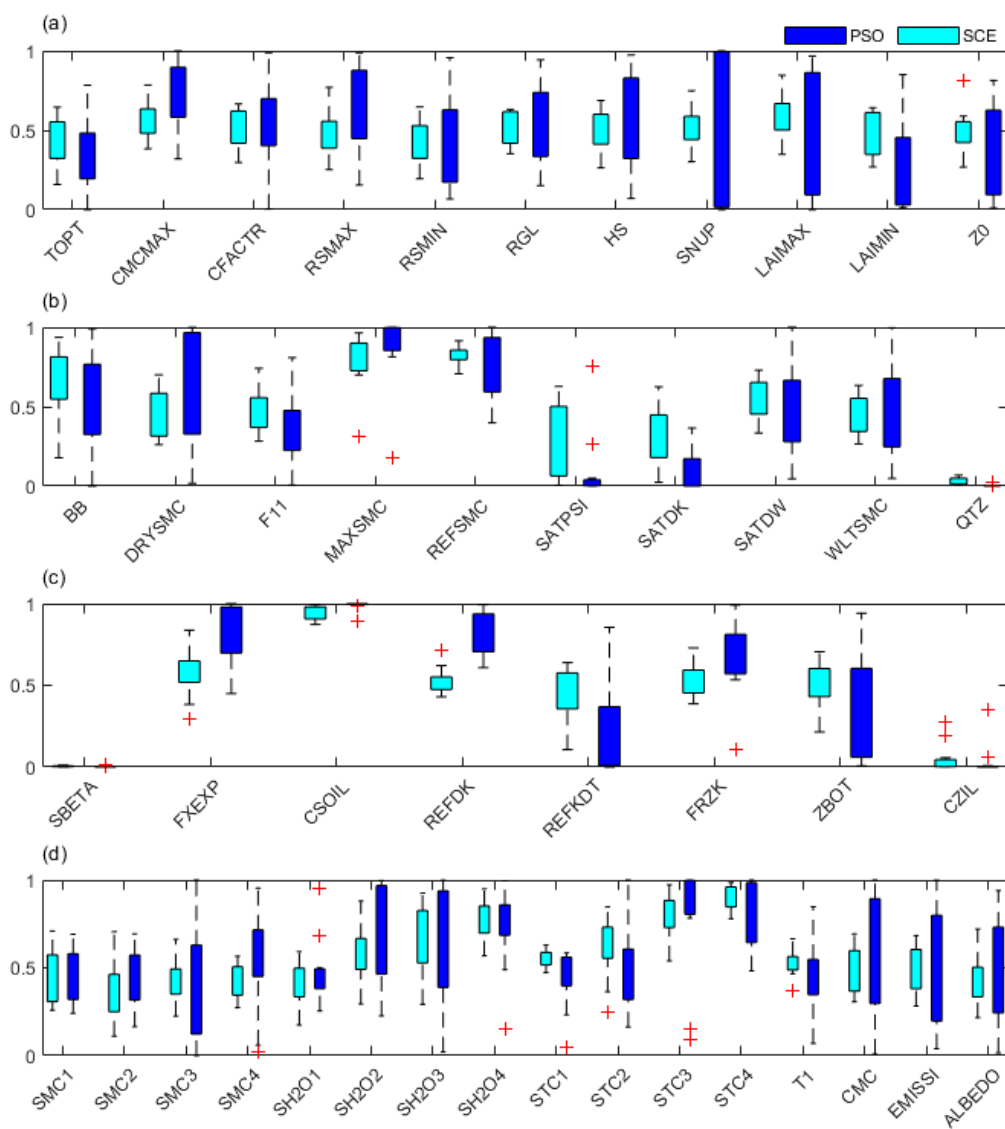


Figure S1-2-3. The same as Figure S1-2-1, but for the metric EMO.

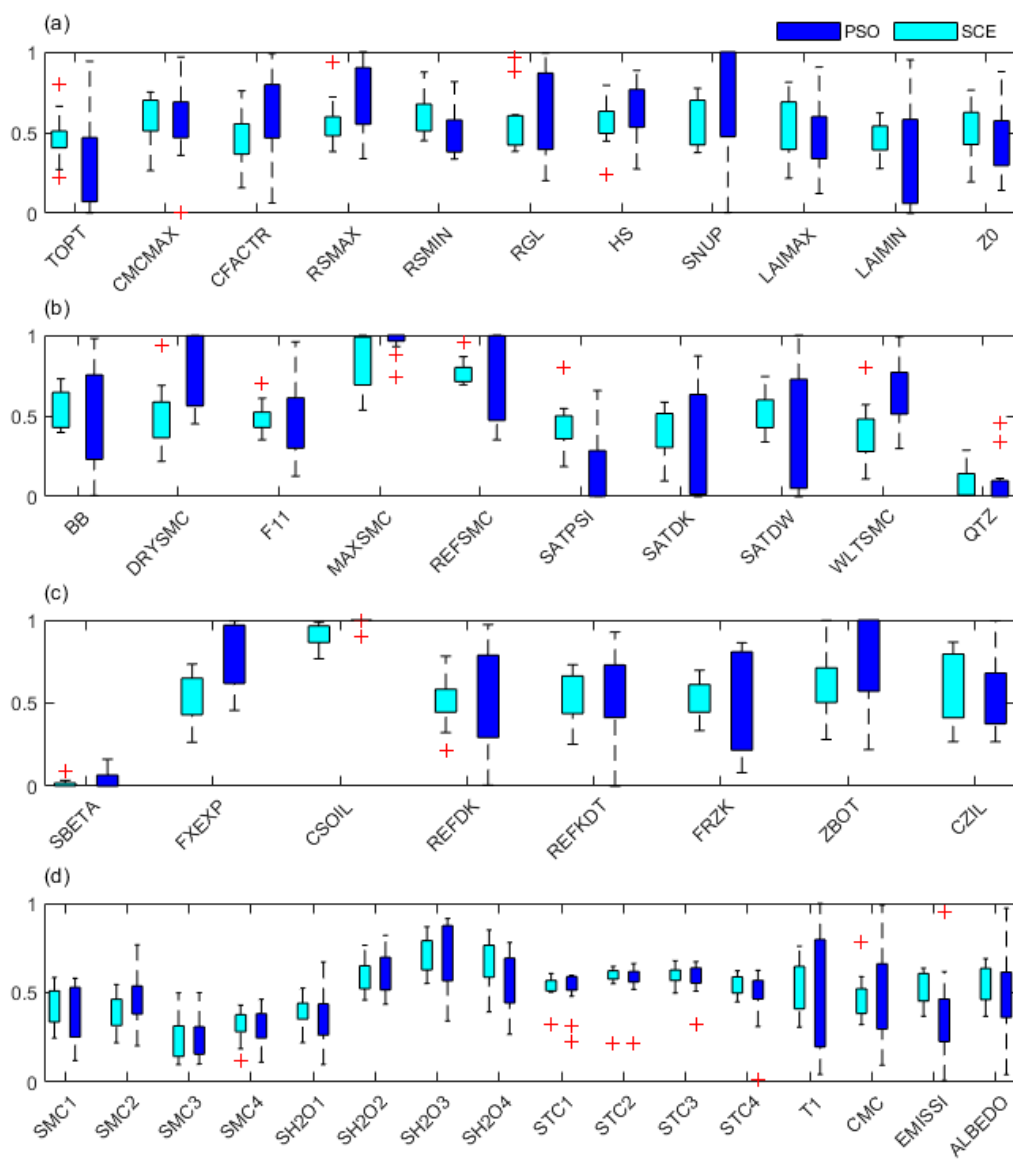


Figure S1-2-4. The same as Figure S1-2-1, but for the metric MAES.

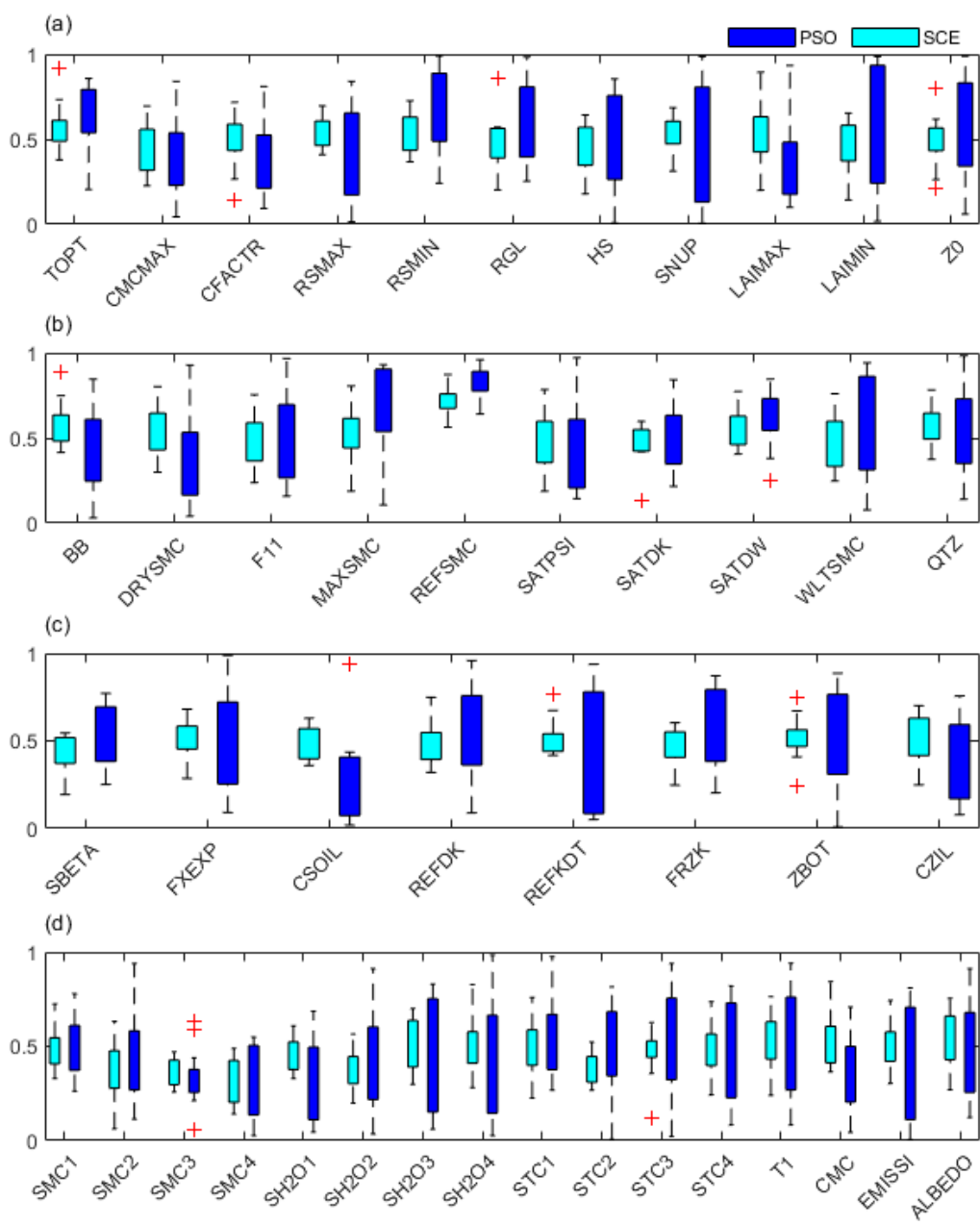


Figure S1-2-5. The same as Figure S1-2-1, but for the metric NSES.

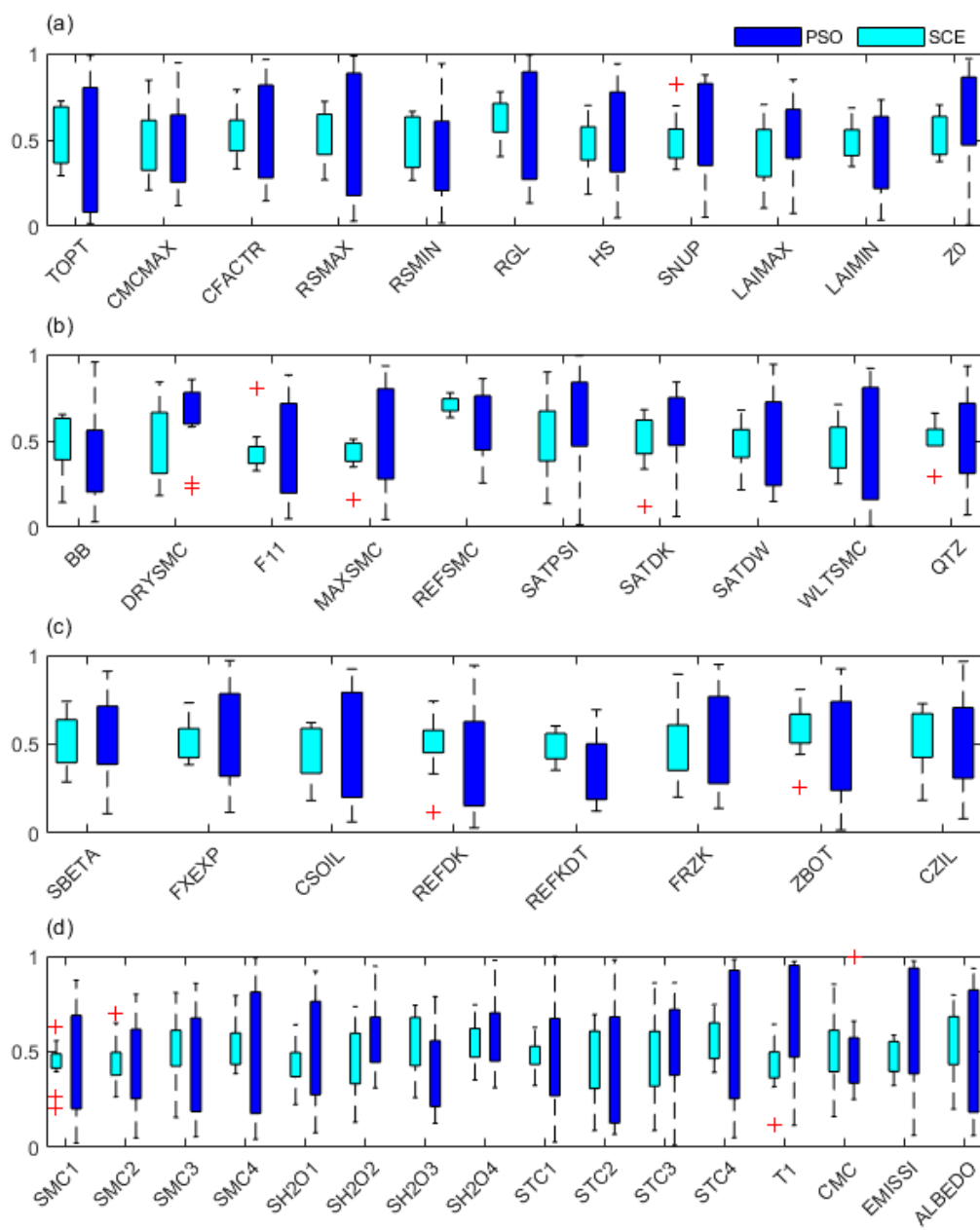


Figure S1-2-6. The same as Figure S1-2-1, but for the metric PKGE.

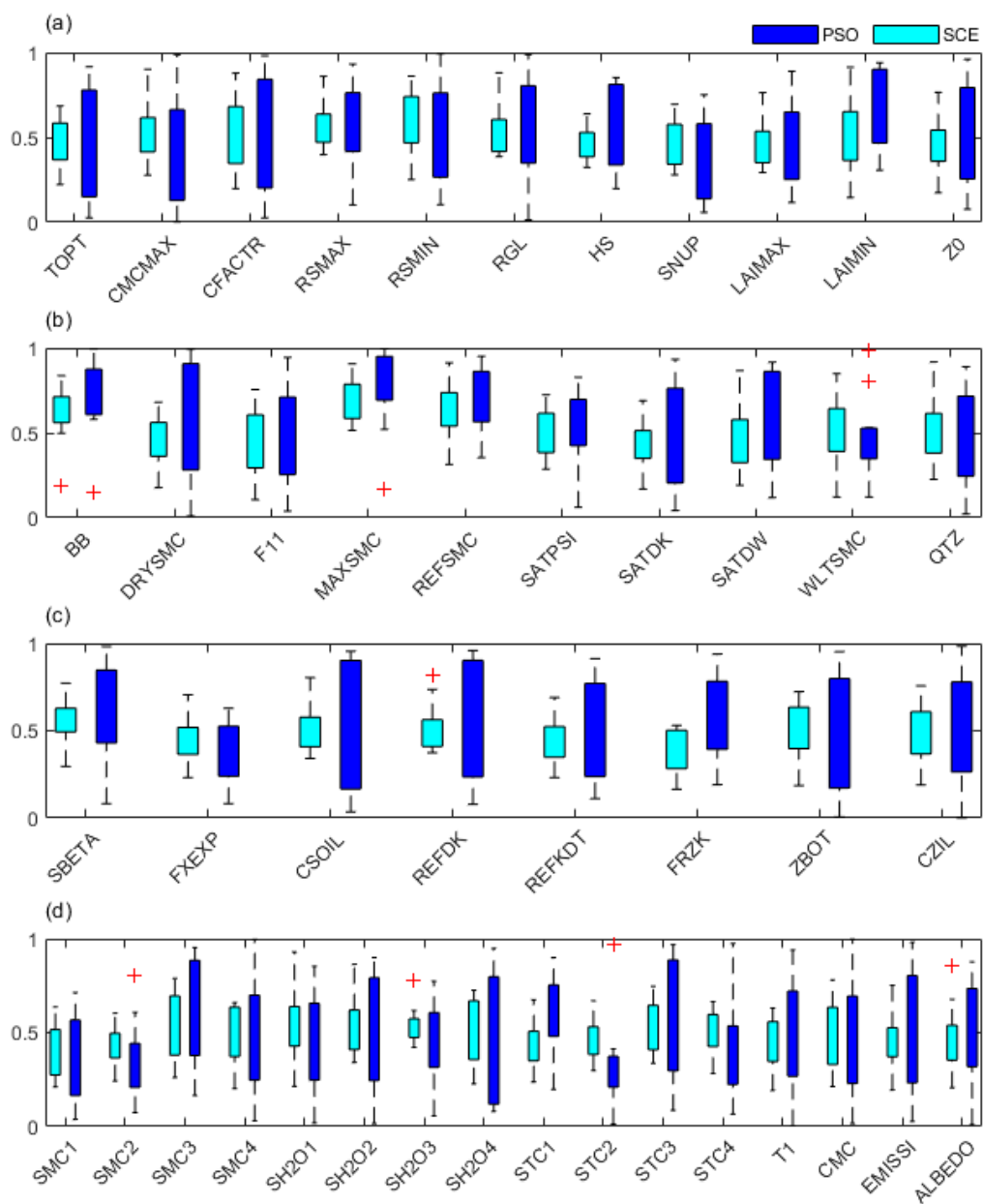


Figure S1-2-7. The same as Figure S1-2-1, but for the metric PMO.

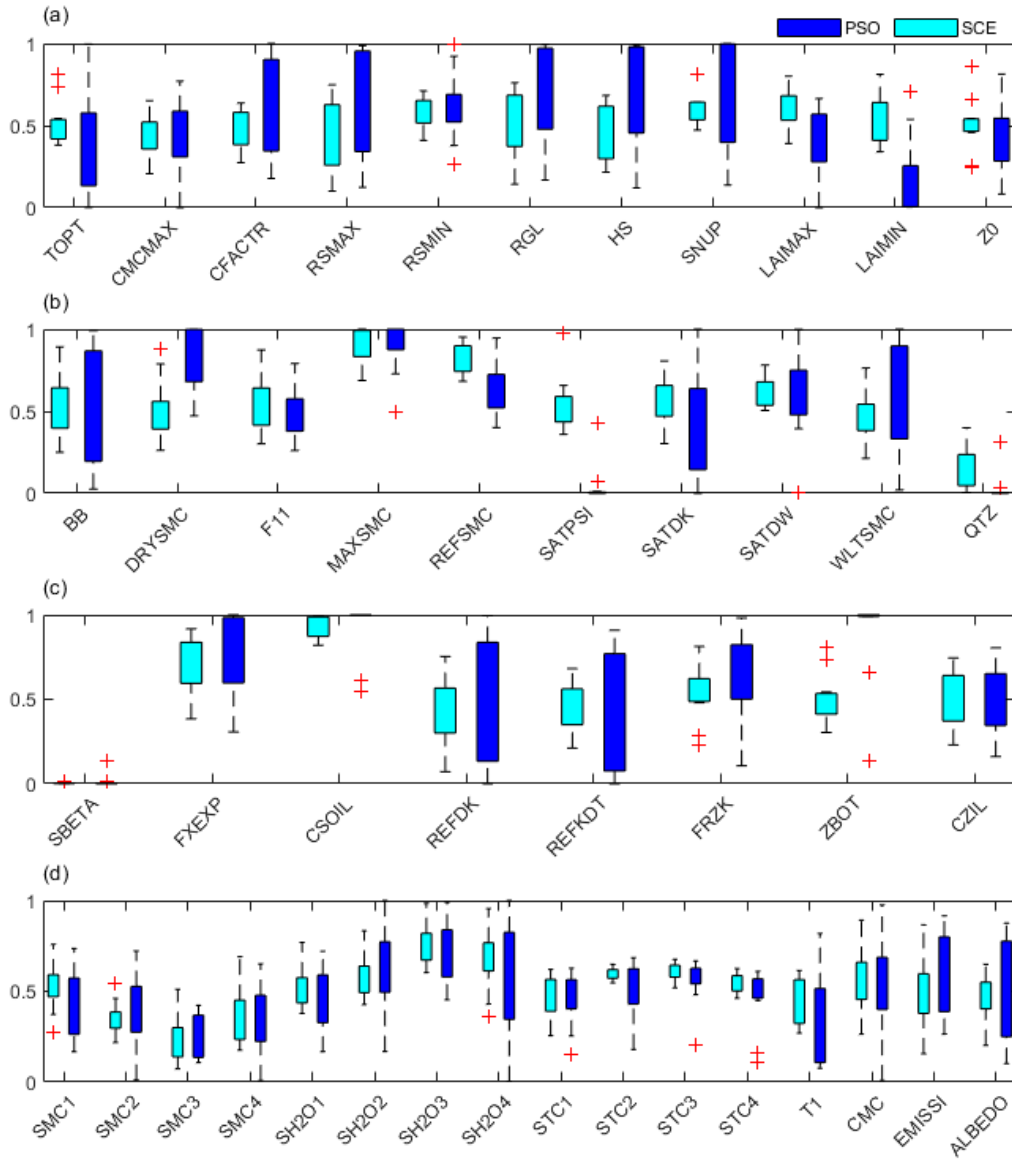


Figure S1-2-8. The same as Figure S1-2-1, but for the metric RMSES.

## 2. Effects on Optimal Simulations

### 2.1 Linear Fits

Figure S2-1-1 illustrates the varied statistical characteristic differences between simulations and observations of SM05cm and ST05cm under different metrics. For the linear fitting of SM05cm observations versus optimal simulations, the PSO linear fitting coefficients rank from highest to lowest as  $EMO > EKGE > RMSES > MAES > PMO > NSES > CCS > PKGE$ , whereas the coefficients of determination ( $r^2$ ) are ordered from highest to lowest as:  $EMO > EKGE > MAES > RMSES > NSES > PMO > CCS > PKGE$ . In contrast, for the SCE linear fitting, the coefficients rank from highest to lowest as  $EMO > PMO > EKGE > MAES > PKGE > NSES > RMSES > CCS$ , and the corresponding coefficients of determination ( $r^2$ ) are ordered from highest to lowest as:  $EMO > EKGE > MAES > PMO > NSES > RMSES > PKGE > CCS$ .

Moreover, for the linear fitting of ST05cm observations versus optimal simulations (Figure S2-1-2), the PSO linear fitting coefficients are ordered from highest to lowest as  $EKGE > EMO > MAES > RMSES > CCS > NSES > PKGE > PMO$ , whereas the coefficients of determination ( $r^2$ ) are ranked from highest to lowest as  $PMO > EMO / PKGE > MAES /$

RMSES / NSES > EKGE > CCS. In contrast, for the SCE linear fitting, the coefficients are ordered from highest to lowest as EKGE > EMO > CCS > RMSES > MAES > NSES > PMO > PKGE, and the corresponding coefficients of determination ( $r^2$ ) are ranked from highest to lowest as PKGE > PMO > NSES > EKGE > EMO > RMSES > CCS / MAES.

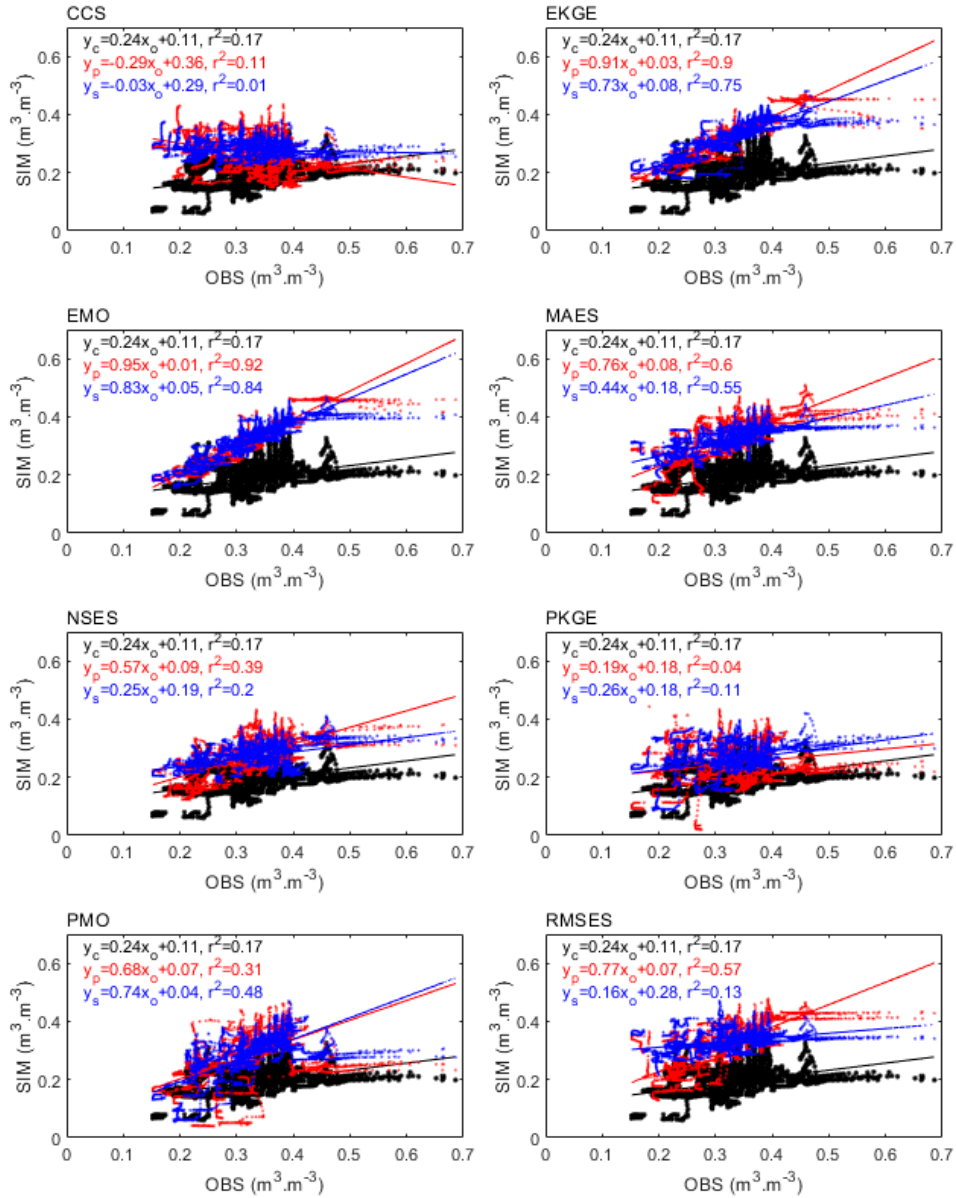


Figure S2-1-1. Different metrics' linear fits against sites for SM05cm during the calibration period. CRT, PSO and SCE are plotted in black, red and blue respectively.

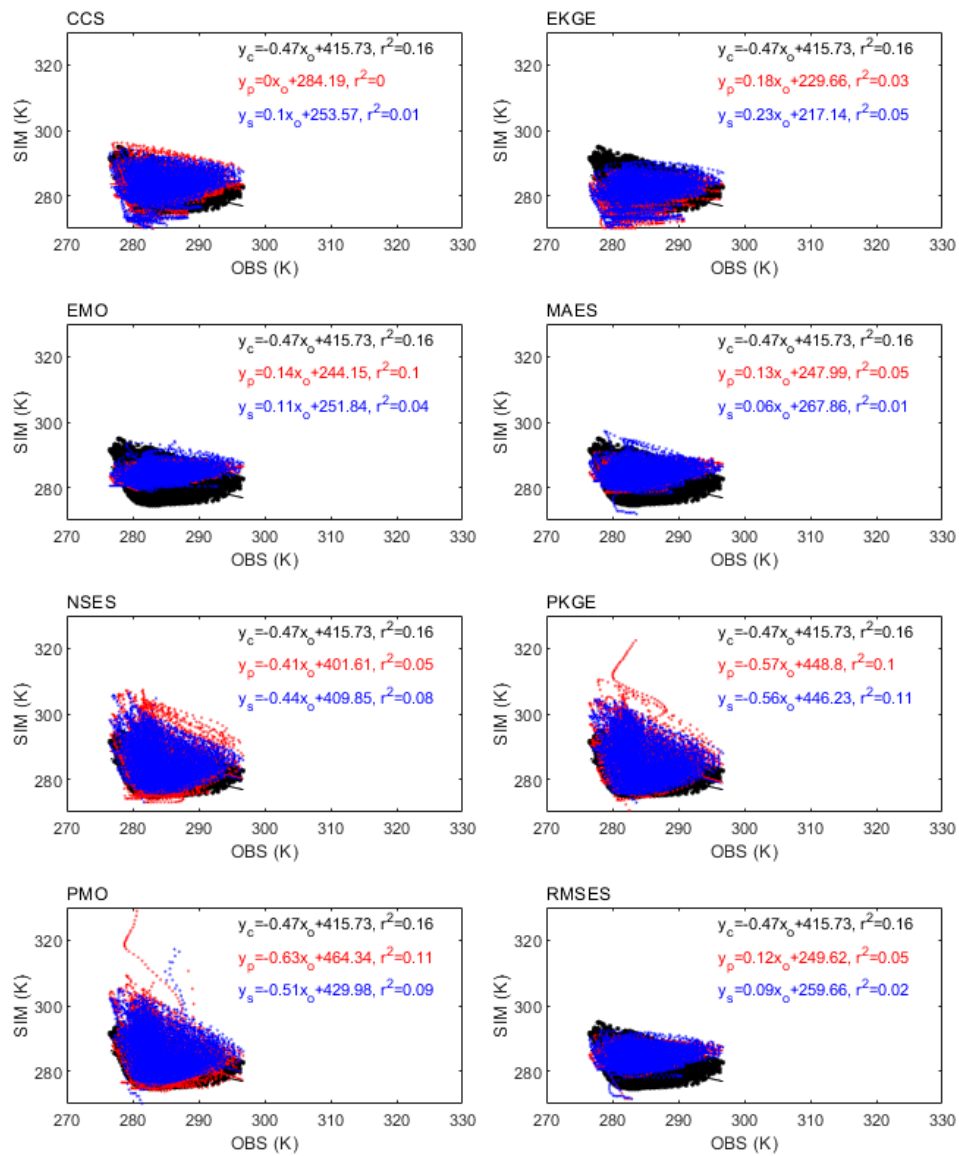


Figure S2-1-2. The same as Figure S2-1-1, but for ST05cm.

## 2.2 Gaussian Fits

The Gaussian fitting distribution characteristics of errors between SM05cm simulations and observations are depicted in Figure S2-2-1. The simulation errors for CTR exhibit a relatively wide distribution, centered approximately at 0.15 m3.m-3 with a frequency of around 297. For CCS, PSO and SCE show the wide distributions centered around -0.04 and 0.11 m3.m-3, respectively, with frequency of around 350 and 295, respectively. For EKGE, PSO and SCE both display a narrower error distribution centered at around 0 m3.m-3, with frequency of around 1276 and 608, respectively. For EMO, PSO and SCE exhibit a narrow error distribution centered at 0 and 0.01 m3.m-3, respectively, with frequency of around 1178 and 700, respectively. For MAES, PSO and SCE have shown the slightly wider error distributions centered around 0.01 and 0.02 m3.m-3 respectively, with the frequency of around 344 and 416 respectively. For NSES, PSO and SCE exhibit wide distributions both centered around 0.05 m3.m-3, but with the frequency of around 274 and 230, respectively. For PKGE, PSO and SCE displays wide error distributions centered around 0.08 and 0.11 m3.m-3 respectively, with the frequency of around 322 and 325 respectively. For PMO, PSO and SCE exhibit narrow error distributions centered at around 0.02 and 0.03 m3.m-3, respectively, with frequencies of around 480 and 444, respectively. For RMSES, PSO and SCE displays the narrow distribution centered around -0.02 and 0 m3.m-3 respectively, with the frequency of around 426 and 296 respectively.

Moreover, the Gaussian fitting distribution characteristics of errors between ST05cm simulations and observations are illustrated in Figure S2-2-2. The simulation errors for CTR exhibit a wide bimodal distribution with two centers located around 7.1 and -3.8 K respectively, with the frequency of around 192 and 134, respectively. For CCS, PSO and SCE display the wide distributions centered around 2.3 and 1.1 K respectively, with the frequency of around 216 and 167 respectively. For EKGE, PSO and SCE show wide distributions centered around 1.3 and 2.5 K respectively, with the frequency of around 200 and 203 respectively. For EMO, PSO and SCE exhibit the distributions with centers around 0.85 and 1.23 K respectively, with the frequency of around 170 and 207, respectively. For MAES, PSO and SCE display the distributions centered around -0.06 and 0.88 K respectively, with the frequency of around 200 and 230 respectively. For NSES, PSO and SCE show wide distributions with centers located around 5.86 and 5.03 K respectively, with the frequency of around 169 and 213, respectively. For PKGE, PSO and SCE exhibit wide distributions centered around 4.91 and 5.01 K respectively, with the frequency of around 237 and 152 respectively. For PMO, PSO and SCE display wide distributions centered around at 6.1 and 5.19 K respectively, with the frequency of around 300 and 224, respectively. For RMSES, PSO and SCE show the distributions centered around 0.16 and 1.29 K respectively, with the frequency of around 200 and 206 respectively.

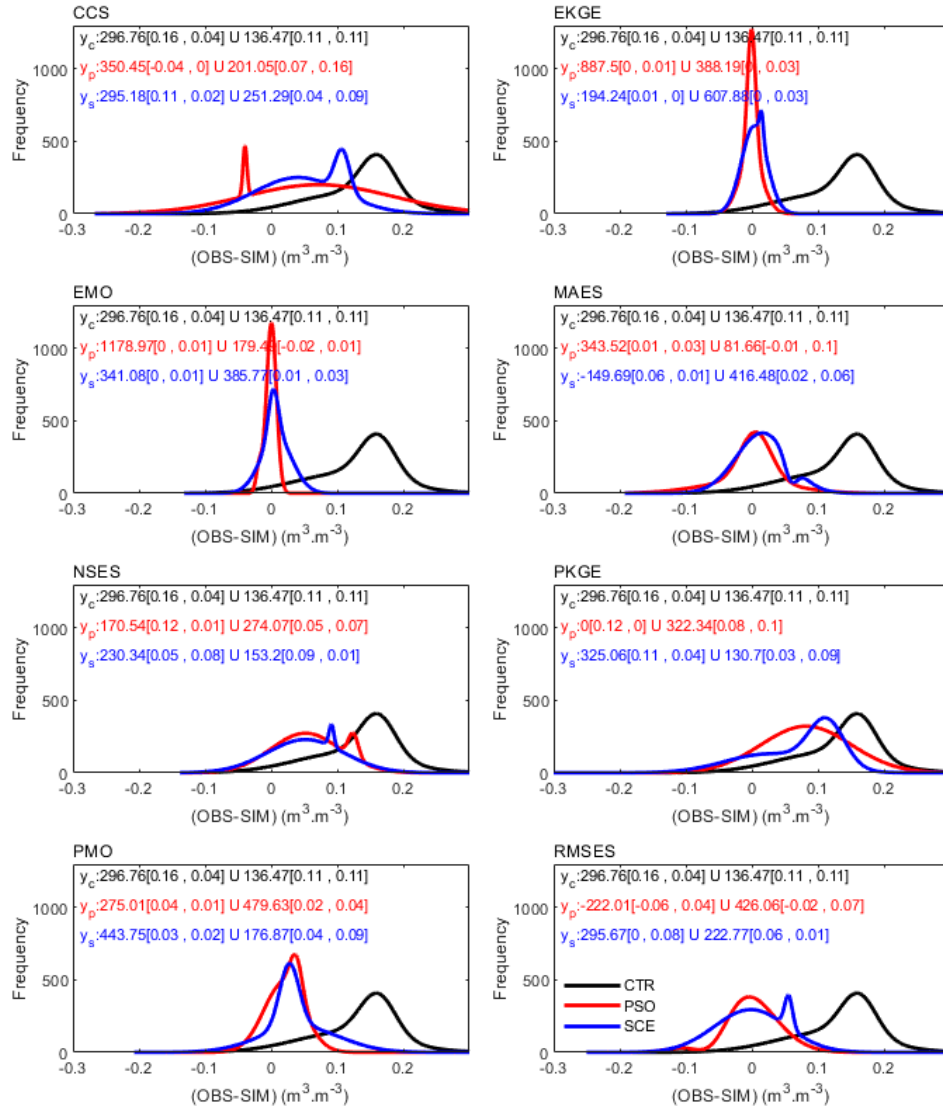


Figure S2-2-1. Different metrics' Gaussian fits against sites for SM05cm during the calibration period. CRT, PSO and SCE are plotted in black, red and blue respectively. Also, the two typically characterized "amplitude [peak position, peak width]" in Gaussian fitting are displayed together. Note that two amplitudes with one same peak could be summed to one amplitude.

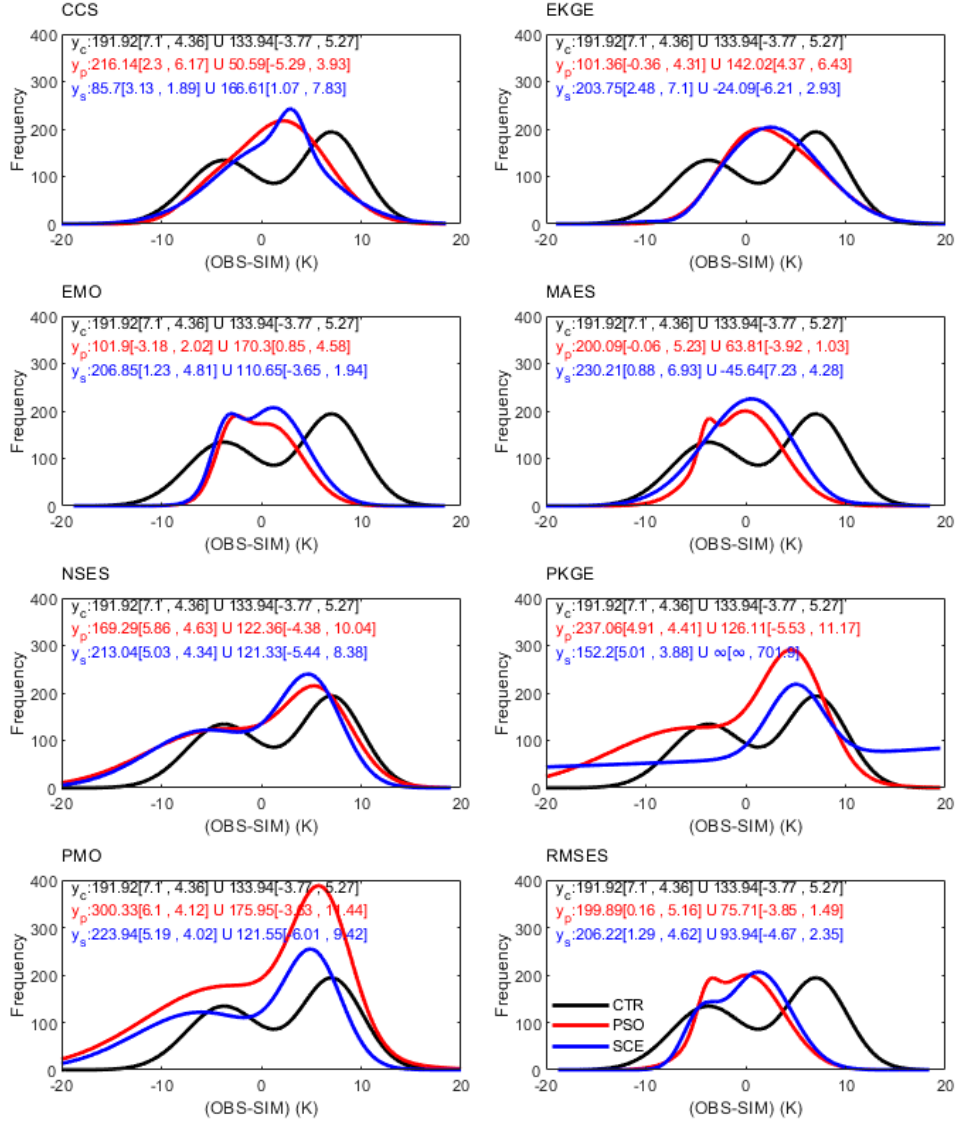


Figure S2-2-2. The same as Figure S2-2-1, but for ST05cm.

### 2.3 Spatial Complexity

For ST05cm, the RMSE of CTR exhibits a pronounced diurnal variation pattern, fluctuating around 8K on average. Due to the significant overlap in the diurnal range of high-to-low error values across different metrics, the performance appears more complex relative to SM05cm. Notably, for metrics such as NSES, PKGE, and PMO, the RMSE values at their peaks even exceed 14K (the maximum RMSE for CTR), indicating a notably inferior performance compared to CTR. In contrast, for MAES and RMSES, the maximum RMSE values are both around 8K, outperforming CTR. Additionally, for EKGE and EMO, except for the initial period (i.e., July 1st to 2nd), the extreme RMSE values are also around 8K, also outperforming CTR. Clearly, for ST05cm, the RMSEs of different target metrics are best for MAES and RMSES, followed by EKGE and EMO, while CCS, NSES, PKGE, and PMO exhibit the poorest performance.

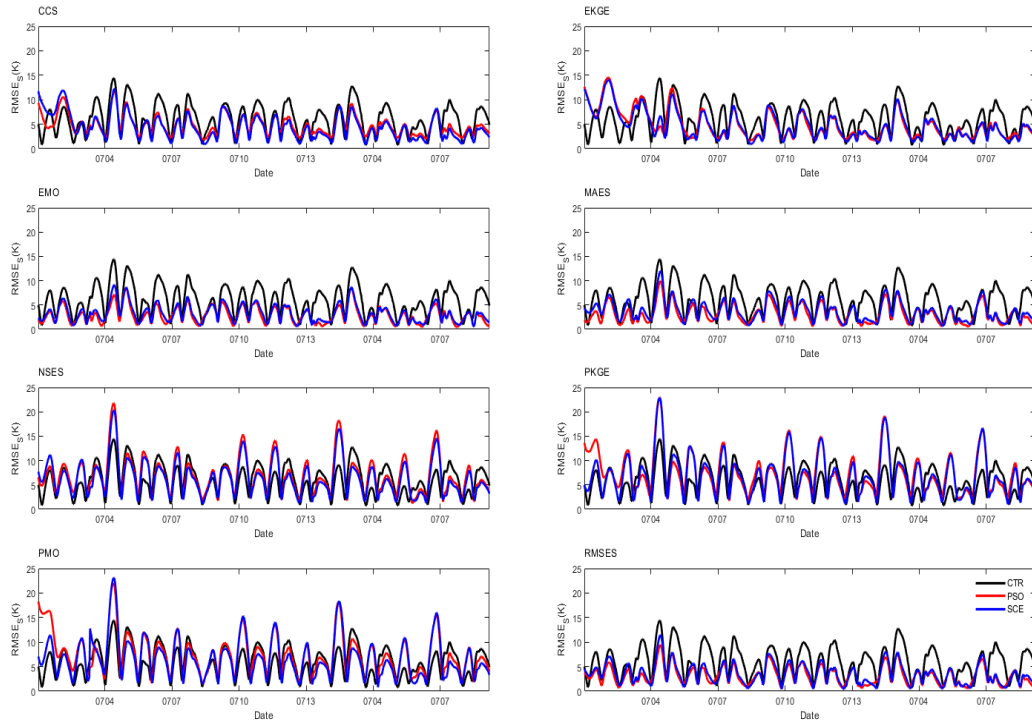


Figure S2-3-1. Different metrics' spatial root-means-square-error (RMSES) against time for ST05cm during the calibration period. CTR, PSO and SCE are plotted in black, red and blue lines respectively.

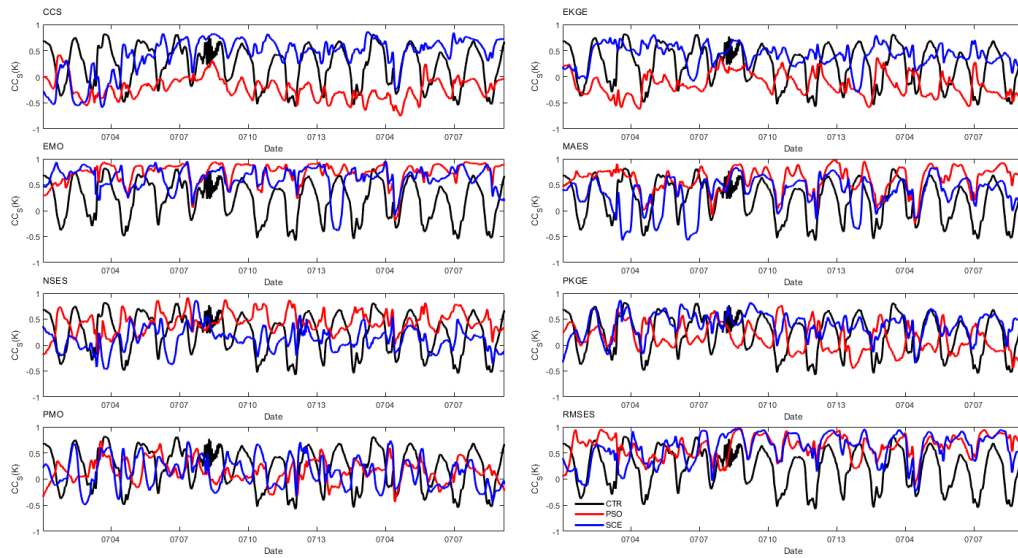


Figure S2-3-2. The same as Figure S2-2-1, but for spatial correlation coefficients (CCs).

### 3. Effects on Forecasts' Improvement

#### 3.1 Linear Fits

For the linear fit between observed and optimal simulated SM05cm values (Figure S3-1-1), the PSO linear fit coefficients, from highest to lowest, are EKGE > EMO > MAES > RMSES > NSES > PMO > CCS > PKGE, while the coefficient of determination ( $r^2$ ) follows the order of EKGE > EMO > MAES > RMSES > NSES > PMO > CCS > PKGE. In contrast, for SCE, the linear fit coefficients, from highest to lowest, are EMO > EKGE > PMO > MAES

> NSES > RMSES > CCS > PKGE, and the coefficient of determination ( $r^2$ ) ranks as EKGE > EMO > MAES > PMO > NSES > PKGE > RMSES > CCS.

Moreover, for the linear fit between observed and optimal simulated ST05cm values (Figure S3-1-2), the PSO linear fit coefficients, from highest to lowest, are MAES > RMSES > EMO > EKGE / CCS > NSES > PKGE > PMO, while the coefficient of determination ( $r^2$ ) ranks as PMO > PKGE > EMO > RMSES > MAES / NSES > EKGE > CCS. For SCE, the linear fit coefficients, from highest to lowest, are MAES / CCS > RMSES > EMO / EKGE > NSES > PMO > PKGE, and the coefficient of determination ( $r^2$ ) ranks as PKGE > RMSES / PMO > MAES / EMO > NSES > CCS / EKGE.

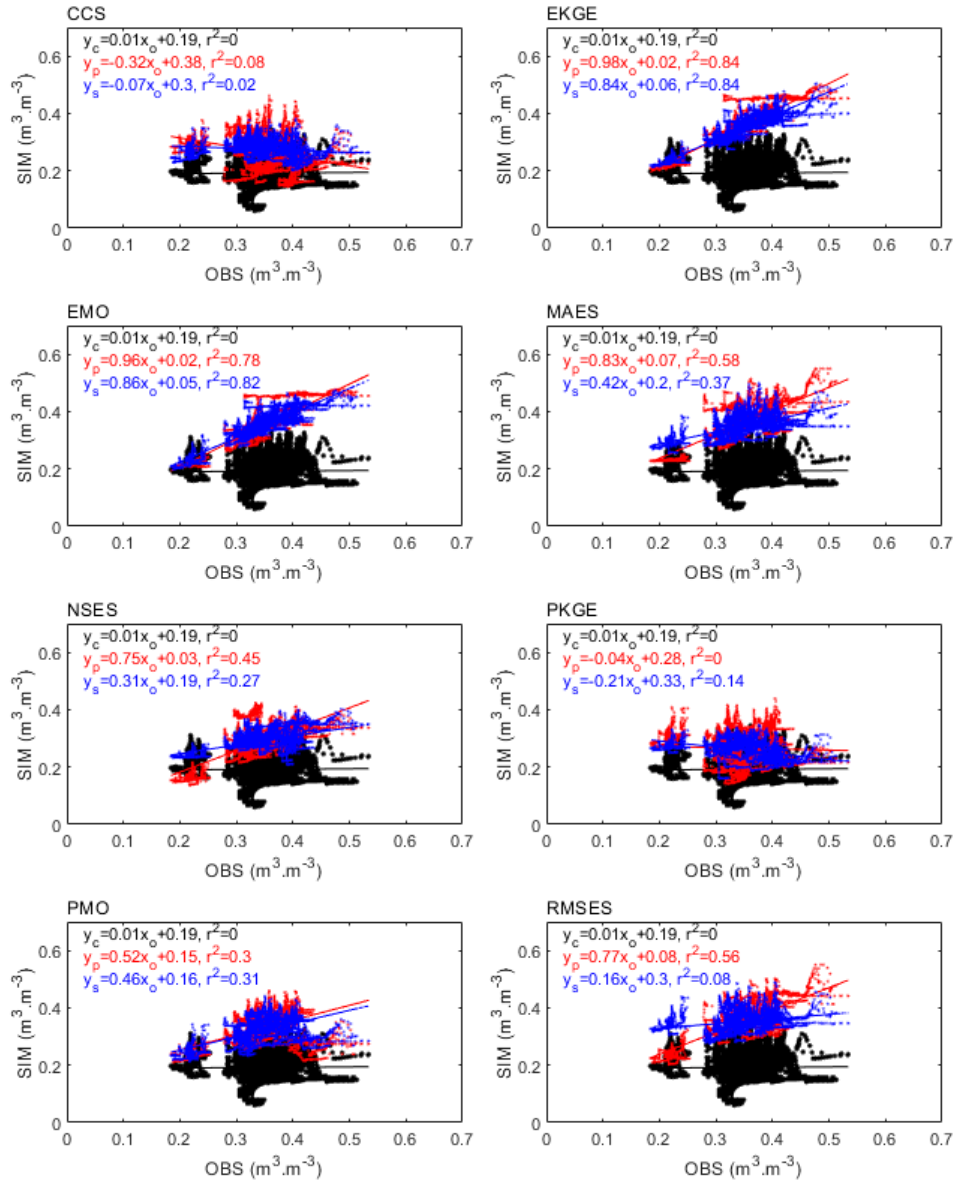


Figure S3-1-1. The same as Figure S2-1-1, but for the validation period.

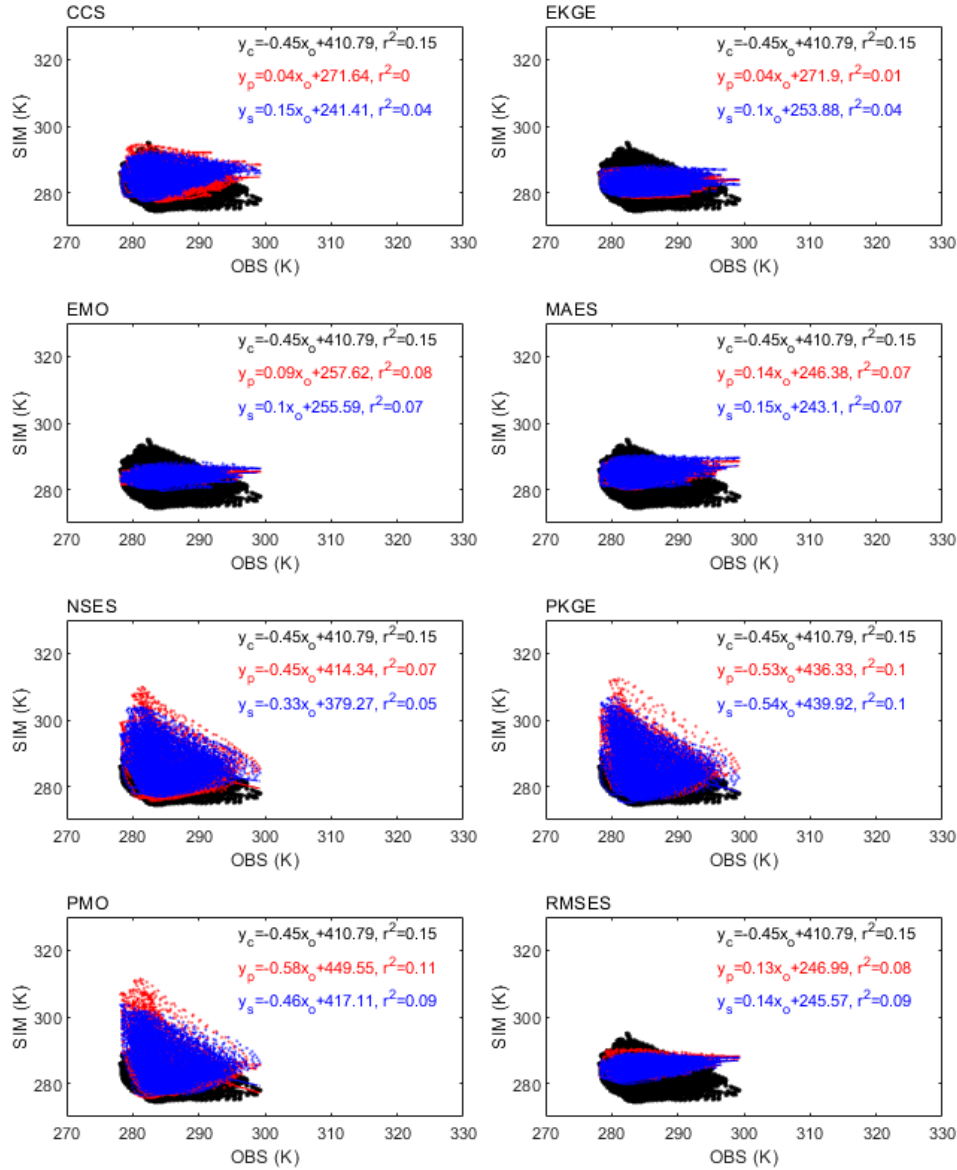


Figure S3-1-2. The same as Figure S2-1-2, but for the validation period.

### 3.2 Gaussian Fits

The Gaussian fitting distribution characteristics of errors between simulated and observed SM05cm values are presented in Figure S3-2-1. The SM05cm simulation errors for CTR exhibit a wide distribution centered around 0.19 m3.m-3 with a frequency of around 272. For CCS, PSO and SCE show wide distributions centered around 0.15 and 0.07 m3.m-3, respectively, with the frequency of around 189 and 225 respectively. For EKGE, PSO and SCE both demonstrate narrow distributions centered around 0 m3.m-3, but with the frequency of around 383 and 363 respectively. For EMO, PSO and SCE both exhibit the distributions centered around 0 m3.m-3, but with the frequency of around 416 and 359 respectively. For MAES, PSO and SCE display the distributions centered around -0.01 and 0 m3.m-3 respectively, with the frequency of around 359 and 284 respectively. For NSES, PSO and SCE display the distributions centered around 0.06 and 0.05 m3.m-3 respectively, with the frequency of around 343 and 322 respectively. For PKGE, PSO exhibits a wide bimodal distribution with centers of around 0.13 and 0.04 m3.m-3, and frequencies of around 234 and 220 respectively, whereas SCE also displays wide bimodal distribution

with centers of around 0.16 and 0.06 m<sup>3</sup>.m<sup>-3</sup>, and frequencies of around 199 and 365 respectively. For PMO, PSO and SCE show wide distribution centered around 0.01 and 0.04 m<sup>3</sup>.m<sup>-3</sup> respectively, with the frequency of around 367 and 323 respectively. For RMSES, PSO and SCE exhibit the distributions centered around -0.02 and 0.01 m<sup>3</sup>.m<sup>-3</sup> respectively, with the frequency of around 293 and 326 respectively.

Moreover, the Gaussian fitting distribution characteristics of errors between simulated and observed ST05cm values are depicted in Figure S3-2-2. The simulation errors for CTR exhibit a relatively wide, bimodal distribution centered around 7.28 and -3.57 K respectively, with the frequencies of around 211 and 160 respectively. For CCS, PSO and SCE display wider distributions centered around 3.2 and -0.38 K respectively, with the frequency of around 187 and 181 respectively. For EKGE, PSO and SCE show the distributions centered around -0.09 and 3.39 K respectively, with the frequency of around 143 and 189 respectively. For EMO, PSO and SCE exhibit the distributions centered around -1.41 and -0.98 K respectively, with the frequency of around 175 and 148 respectively. For MAES, PSO and SCE demonstrate the distributions centered around 0.49 and 0.29 K respectively, with the frequency of around 181 and 206 respectively. For NSES, PSO and SCE display wide distributions centered around 5.81 and 4.56 K respectively, with the frequency of around 204 and 210 respectively. For PKGE, PSO and SCE show wide distributions centered around 4.9 and 5.7 K respectively, with the frequency of around 214 and 217 respectively. For PMO, PSO and SCE exhibit wide distributions centered around 6.17 and 5.47 K respectively, with the frequency of around 221 and 187 respectively. For RMSES, PSO and SCE display the distributions centered around 0.55 and 0.32 K respectively, with the frequency of around 194 and 198 respectively.

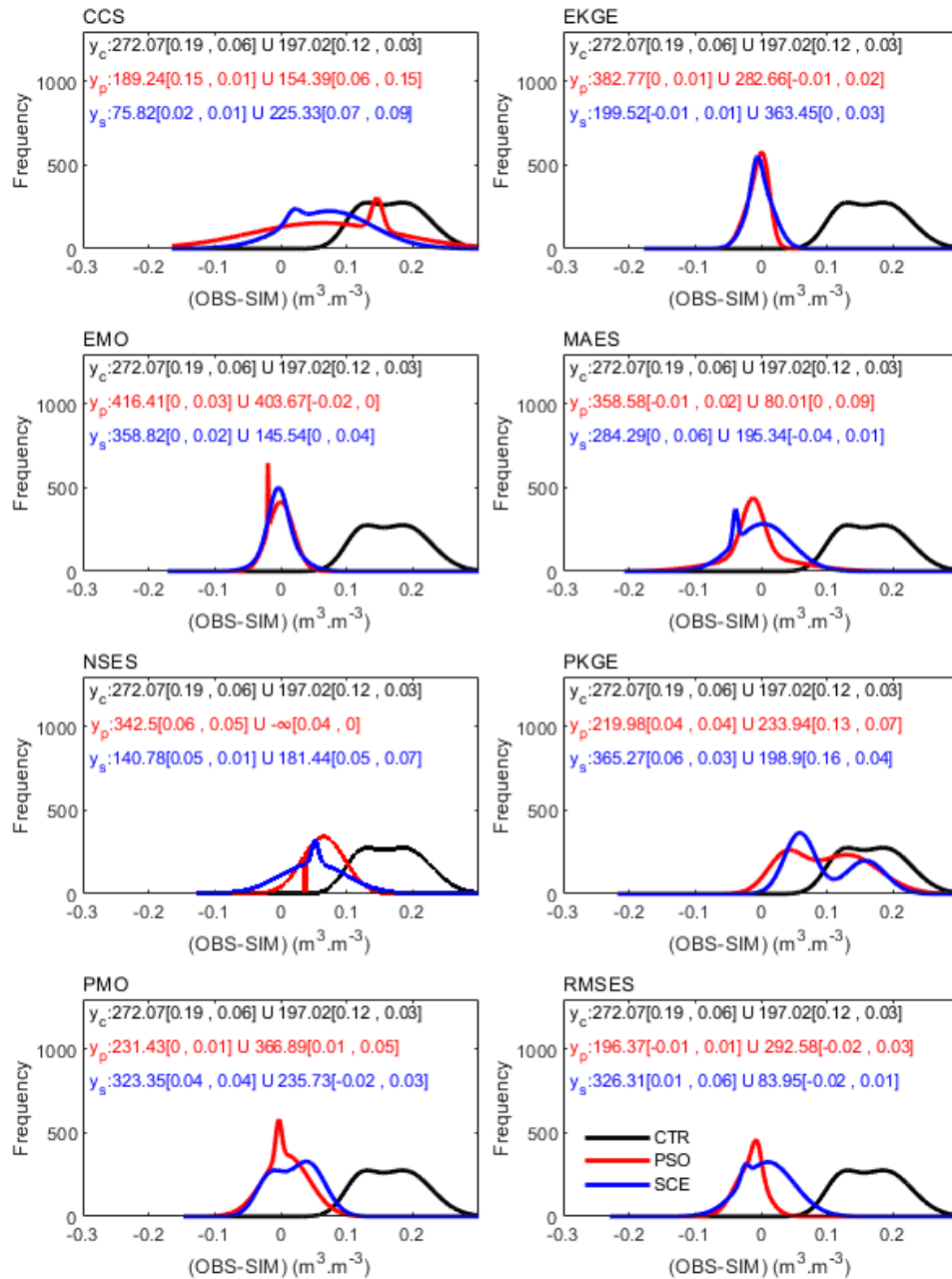


Figure S3-2-1. The same as Figure S2-2-1, but for the validation period.

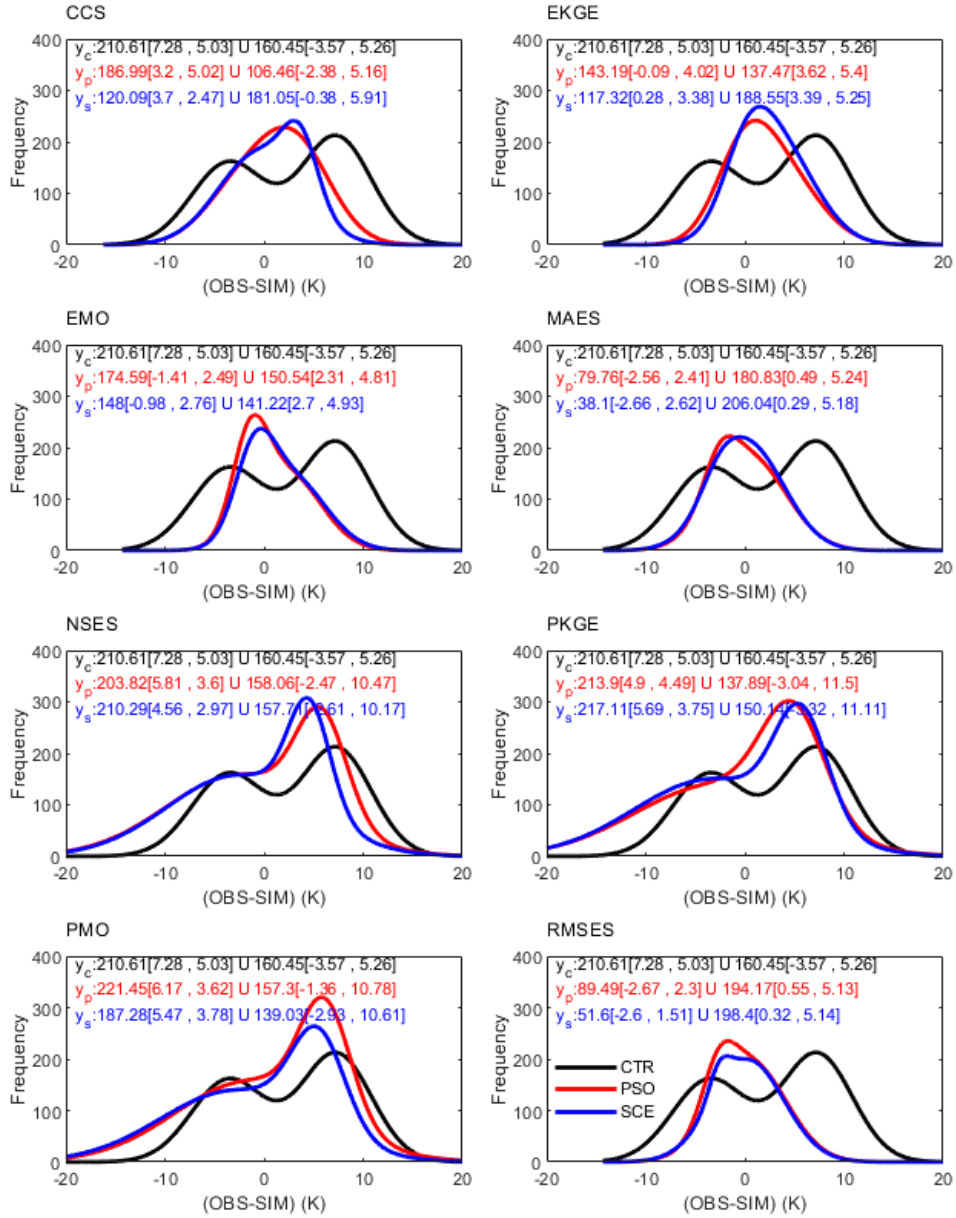


Figure S3-2-2. The same as Figure S2-2-2, but for the validation period.

### 3.3 Spatial Complexity

For ST05cm, the RMSE of CTR exhibits a pronounced diurnal variation, fluctuating around 10K overall. Due to significant overlaps in the diurnal amplitudes of high-low error values among various metrics, the performance is relatively more complex compared to SM05cm. Notably, for metrics such as NSES, PKGE, and PMO, the RMSE values at their peaks even exceed 15K (the maximum RMSE of CTR), indicating inferior performance compared to CTR. In contrast, for EMO, MAES, and RMSES metrics, the maximum RMSEs are all below 7K, demonstrating better performance than CTR. Additionally, for CCS and EKGE metrics, except for the initial period (i.e., July 1st to 2nd), the extreme RMSE values are around 8K, also outperforming CTR. Clearly, for ST05cm, the RMSEs of different target metrics exhibit a hierarchy, with EMO, MAES, and RMSES performing the best, followed by CCS and EKGE, and NSES, PKGE, and PMO performing the worst.

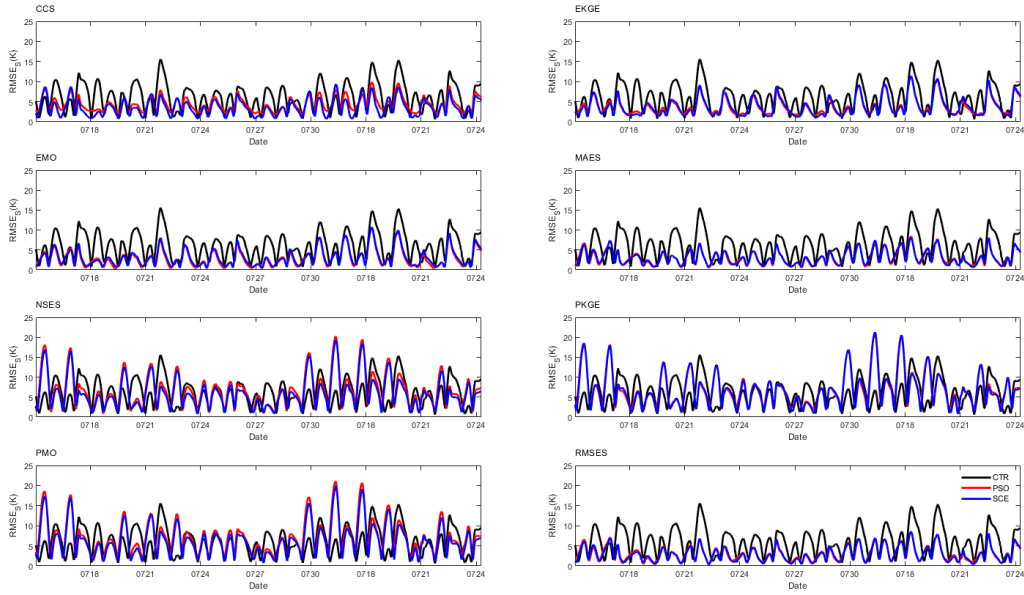


Figure S3-3-1. Different metrics' spatial root-means-square-error (RMSES) against time for ST05cm during the forecasting period. PSO and SCE are plotted in solid and dotted lines respectively.

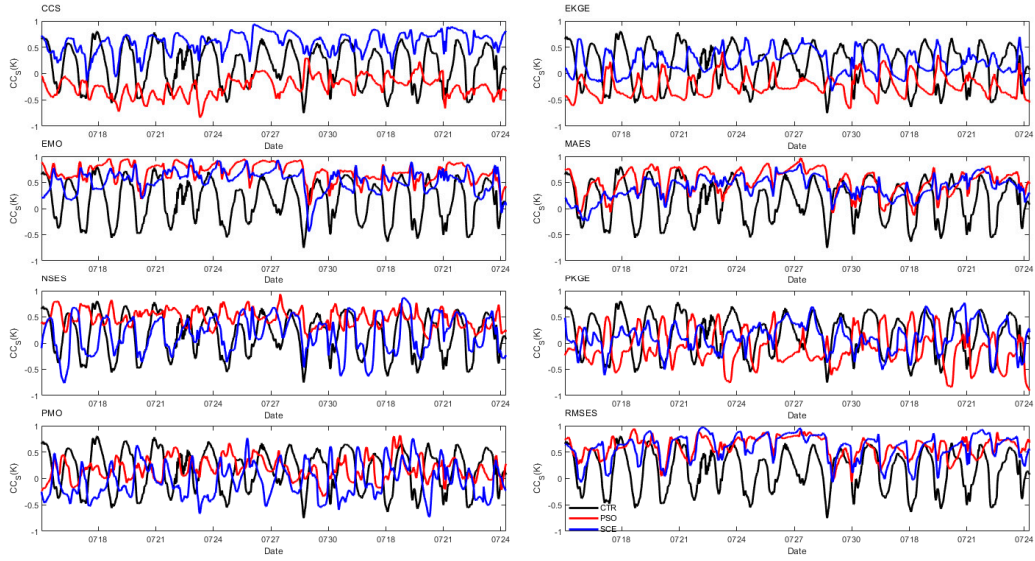


Figure S3-3-2. The same as Figure S3-3-1, but for spatial correlation coefficients (CCs).