Supplementary Materials

Article

Modelling Soil Temperature by Tree-Based Machine Learning Methods in Different Climatic Regions of China

Jianhua Dong [1], Guomin Huang [1], Lifeng Wu[1,2,*], Fa Liu [3], Sien Li [4], Yaokui Cui [5], Yicheng Wang [2], Menghui Leng [6] Jie Wu [7] and Shaofei Wu [1]

[1] School of Hydraulic and Ecological Engineering, Nanchang Institute of Technology, Nanchang 330099, China; djh0530dyz@whu.edu.cn (J.D.); g.huang@nit.edu.cn (G.H.); wsf17@nit.edu.cn (S.W.)
[2] State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China; wangych@iwhr.com
[3] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Science and Science and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; liufa@igsnrr.ac.cn
[4] Center for Agricultural Water Research in China, China Agricultural University, Beijing 100083, China; lisien@cau.edu.cn
[5] Institute of RS and GIS, School of Earth and Space Sciences, Peking University, Beijing 100871, China; yaokuicui@pku.edu.cn
[6] Jiangxi Key Laboratory of Hydrology-Water Resources and Water Environment, Nanchang Institute of Technology, Nanchang 330099, China; 18279130275@163.com
[7] School of Civil Engineering and Architecture, Wuhan Polytechnic University, Wuhan 430023, China; wujiemc@whpu.edu.cn
* Correspondence: wulifeng@nit.edu.cn

Figures

Figure S1. Performance of data-driven models (M5P2: **a**, **b**; RF2: **c**, **d**; XGBoost2: **e**, **f**) for the estimation of T_s with T_{mean} as inputs.

Figure S2. Performance of data-driven models (M5P3: **a**, **b**; RF3: **c**, **d**; XGBoost3: **e**, **f**) for the estimation of T_s with T_{mean} and RH as inputs.

Figure S3. Performance of data-driven models (M5P4: **a**, **b**; RF4: **c**, **d**; XGBoost4: **e**, **f**) for the estimation of T_s with T_{mean} and R_s as inputs.

Figure S4. Performance of data-driven models (M5P5: **a**, **b**; RF5: **c**, **d**; XGBoost5: **e**, **f**) for the estimation of T_s with T_{max} and T_{min} as inputs.

Figure S5. Performance of data-driven models (M5P6: **a**, **b**; RF6: **c**, **d**; XGBoost6: **e**, **f**) for the estimation of T_s with T_{mean} and U_2 as inputs.

Figure S6. Scatter plot of estimated T_s versus measured T_s from the three machine learning models with T_{max} and T_{min} as inputs.

Figure S7. Scatter plot of estimated T_s versus measured T_s from the three machine learning models with T_{mean} and U_2 as inputs.

Figure S8. Combined indicator values for each model at all stations in the IM region.

Figure S9. Combined indicator values for each model at all stations in the NEC region.

Figure S10. Combined indicator values for each model at all stations in the NC region.

Figure S11. Combined indicator values for each model at all stations in the CC region.

Figure S12. Combined indicator values for each model at all stations in the SC region.

Figure S13. Combined indicator values for each model at all stations in the QTP region.

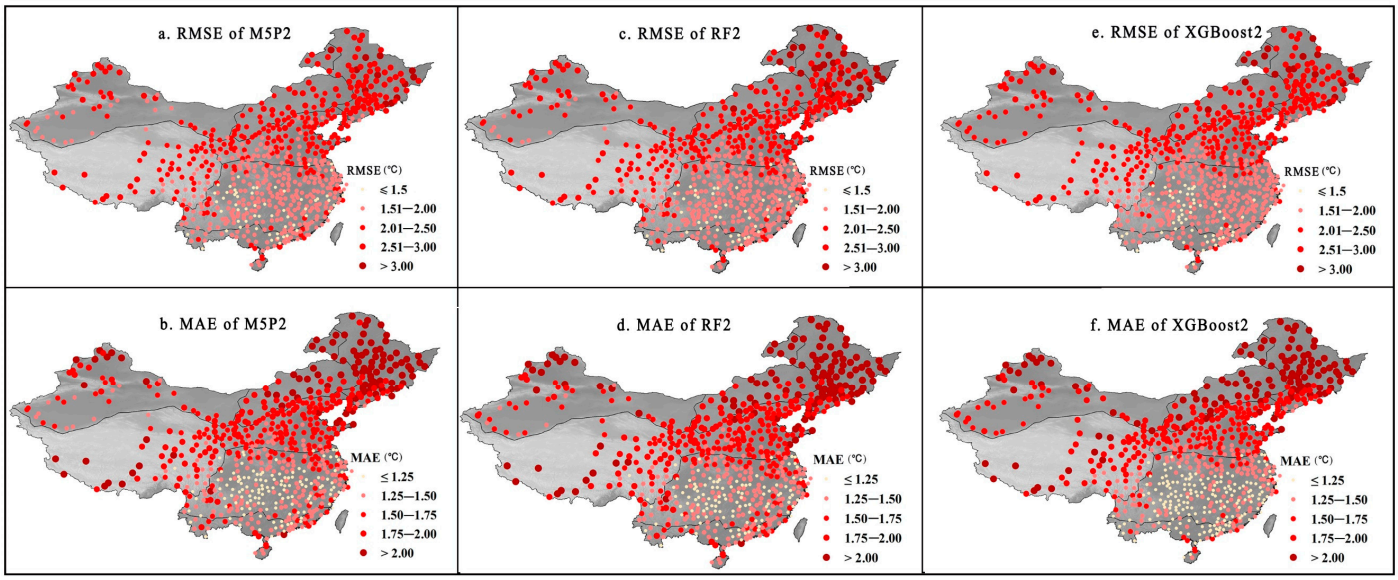

Figure S1. Performance of data-driven models (M5P2: **a**, **b**; RF2: **c**, **d**; XGBoost2: **e**, **f**) for the estimation of T_s with T_{mean} as inputs.

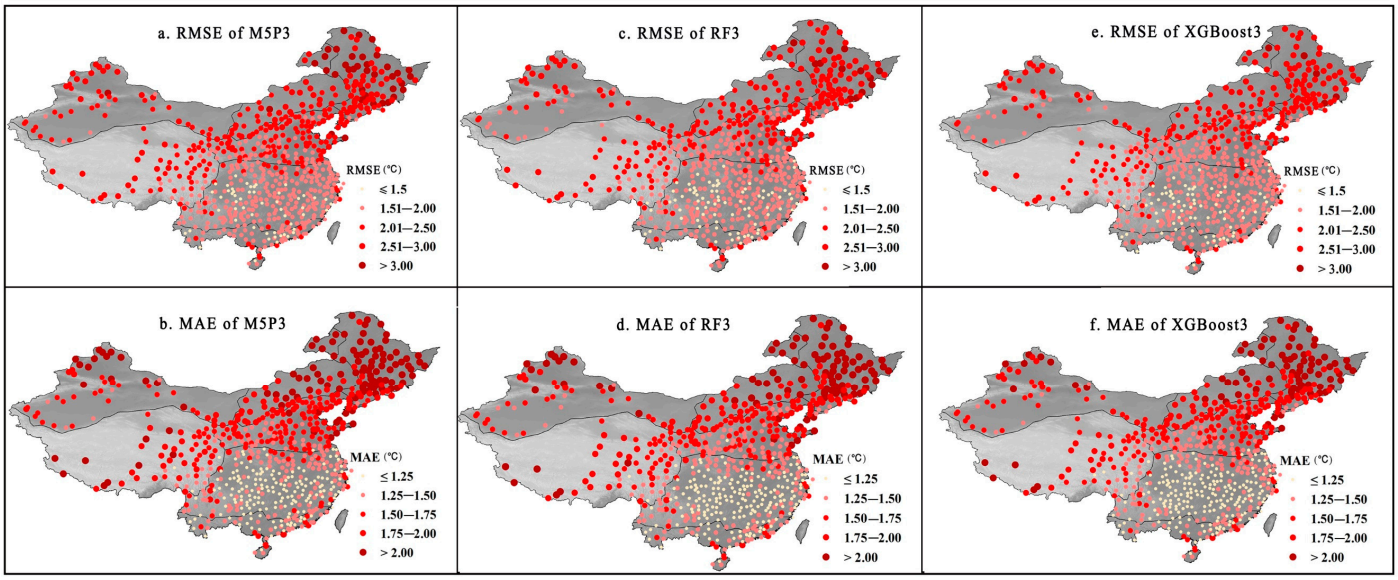

Figure S2. Performance of data-driven models (M5P3: **a**, **b**; RF3: **c**, **d**; XGBoost3: **e**, **f**) for the estimation of T_s with T_{mean} and RH as inputs.

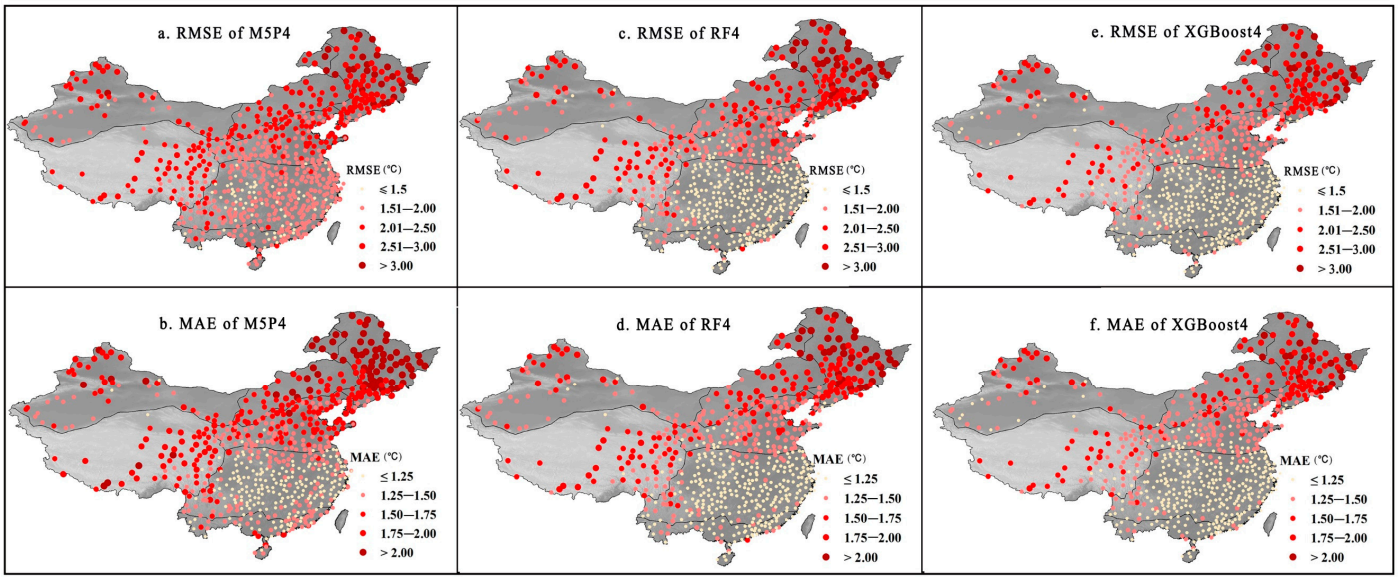

Figure S3. Performance of data-driven models (M5P4: **a**, **b**; RF4: **c**, **d**; XGBoost4: **e**, **f**) for the estimation of T_s with T_{mean} and R_s as inputs.

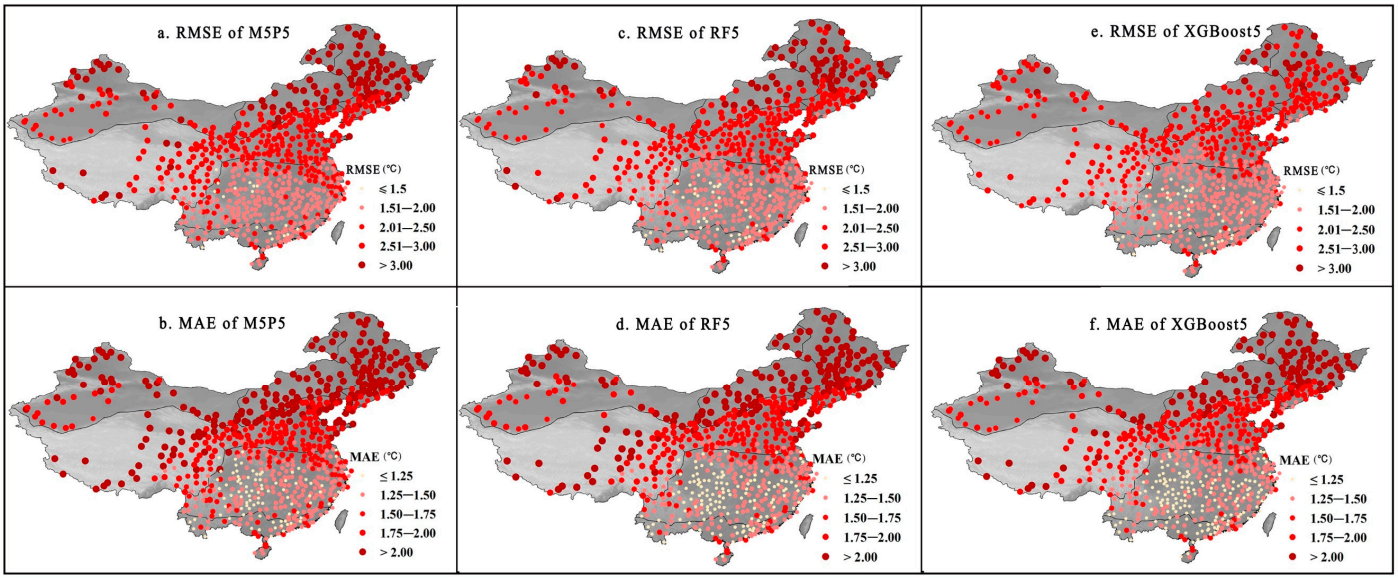

Figure S4. Performance of data-driven models (M5P5: **a**, **b**; RF5: **c**, **d**; XGBoost5: **e**, **f**) for the estimation of T_s with T_{max} and T_{min} as inputs.

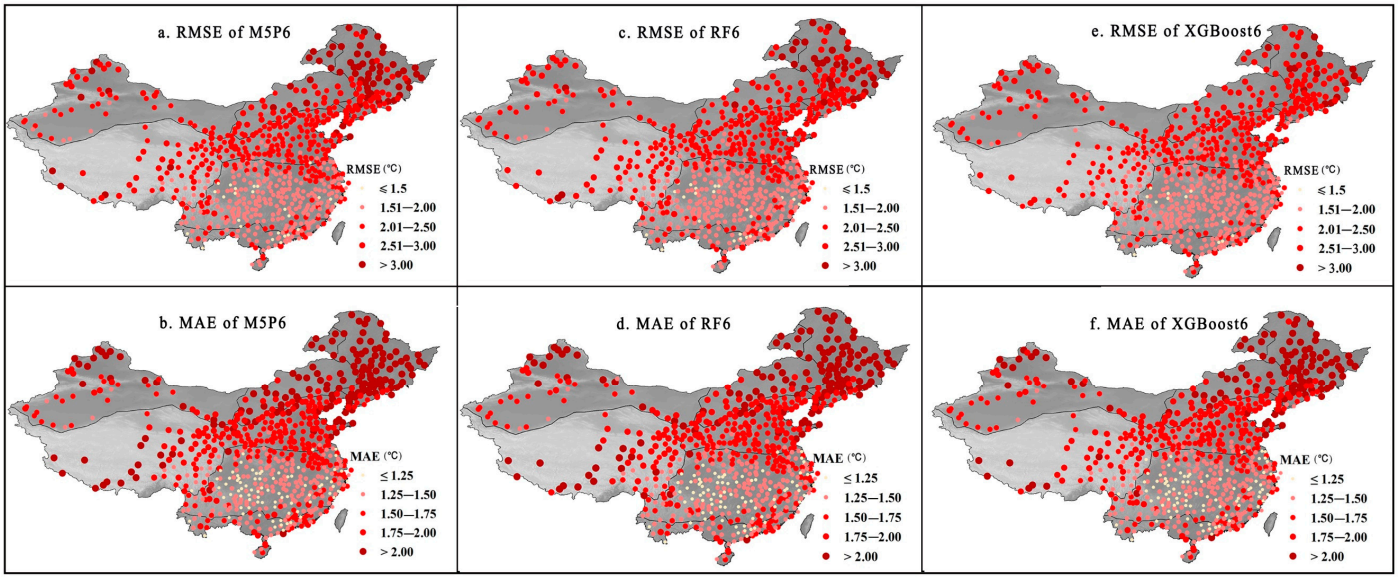

Figure S5. Performance of data-driven models (M5P6: **a**, **b**; RF6: **c**, **d**; XGBoost6: **e**, **f**) for the estimation of T_s with T_{mean} and U_2 as inputs.

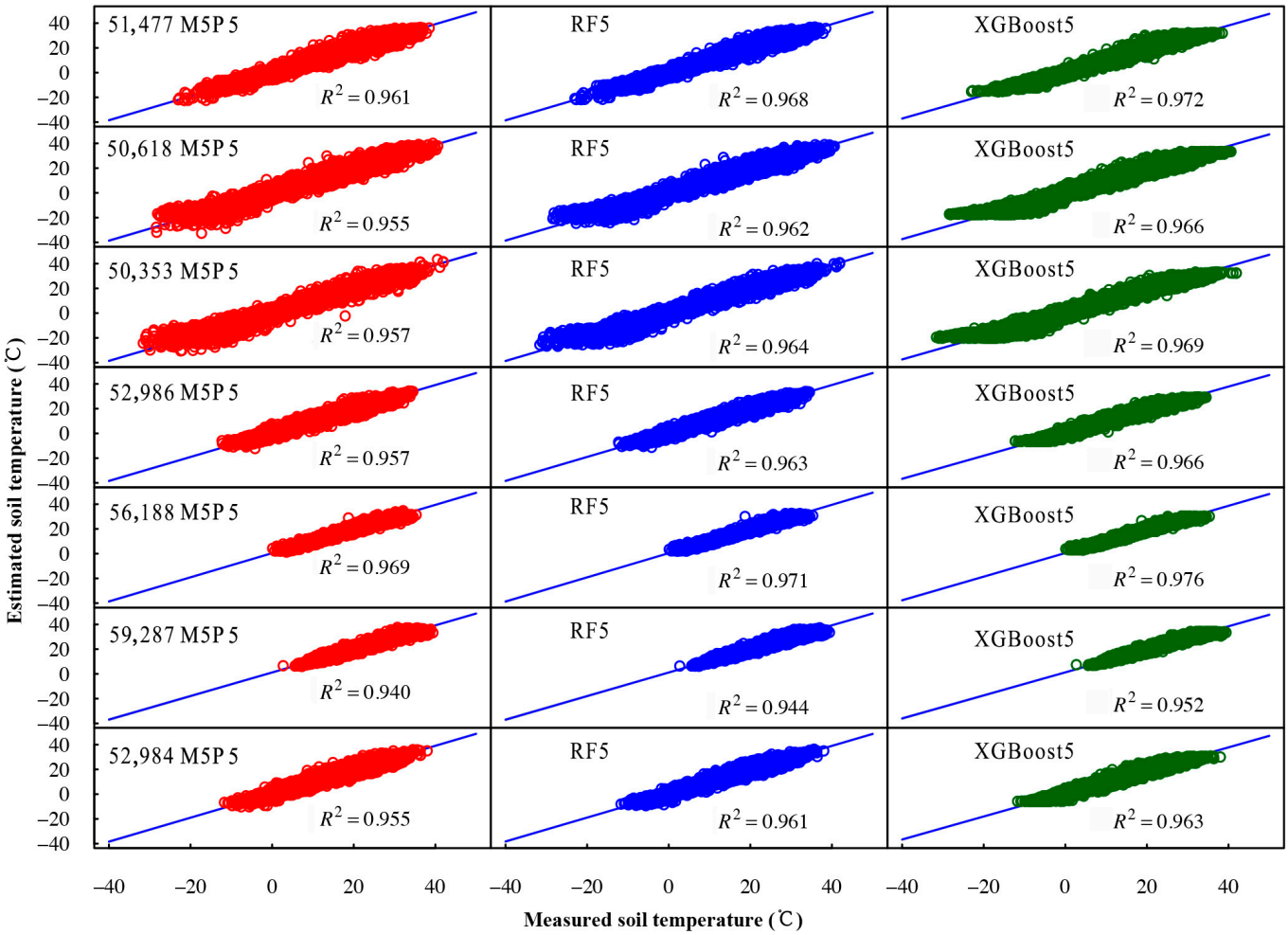

Figure S6. Scatter plot of estimated T_s versus measured T_s from the three machine learning models with T_{max} and T_{min} as inputs.

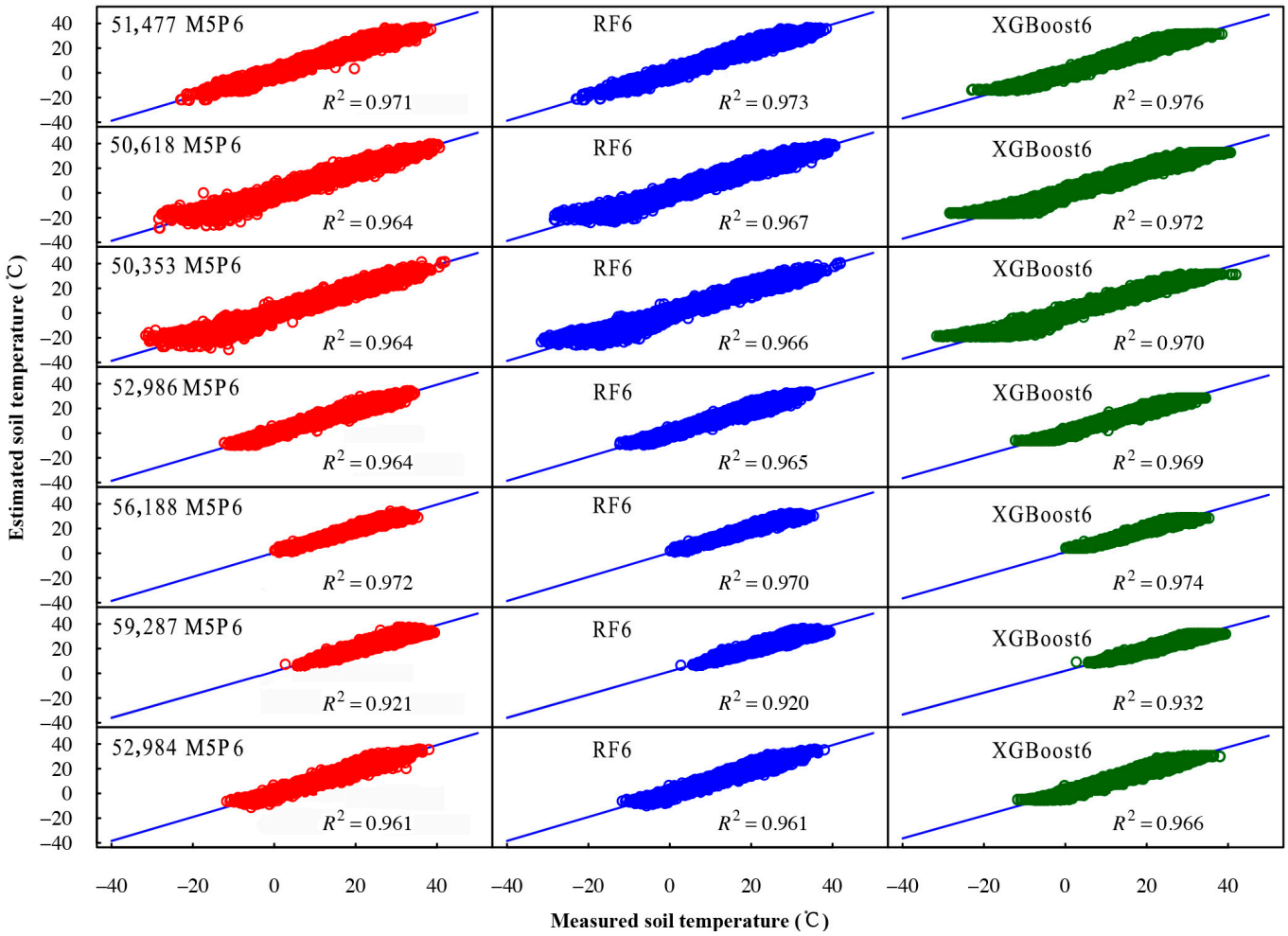

Figure S7. Scatter plot of estimated T_s versus measured T_s from the three machine learning models with T_{mean} and U_2 as inputs.

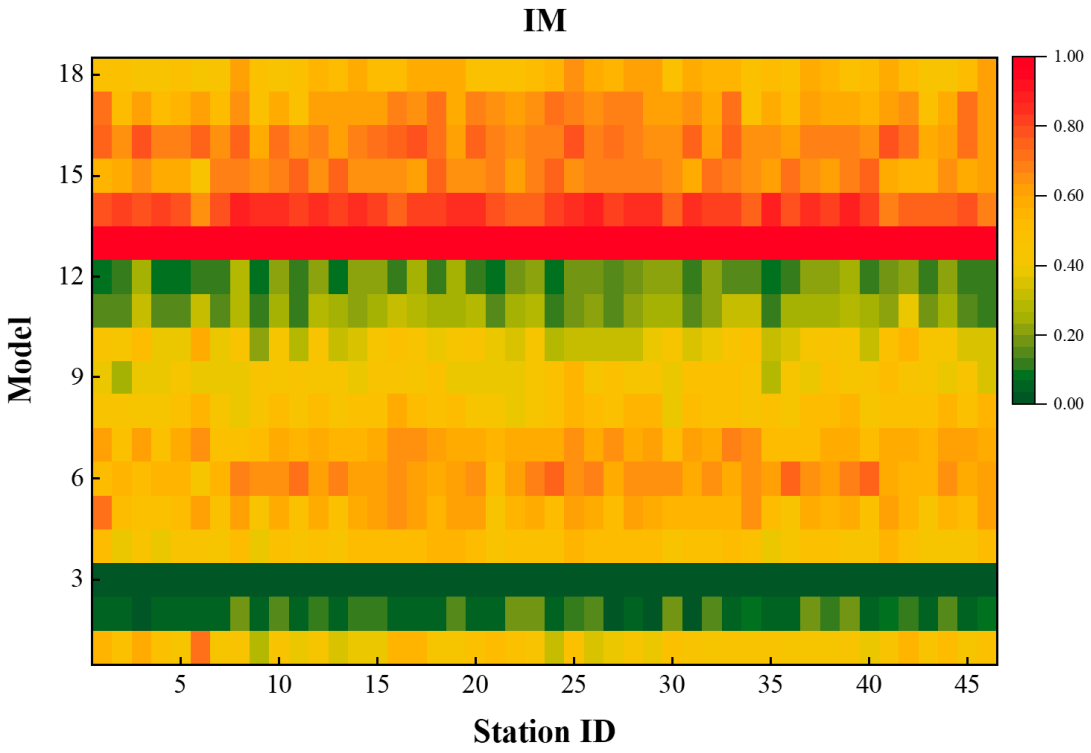

Figure S8. Combined indicator values for each model at all stations in the IM region.

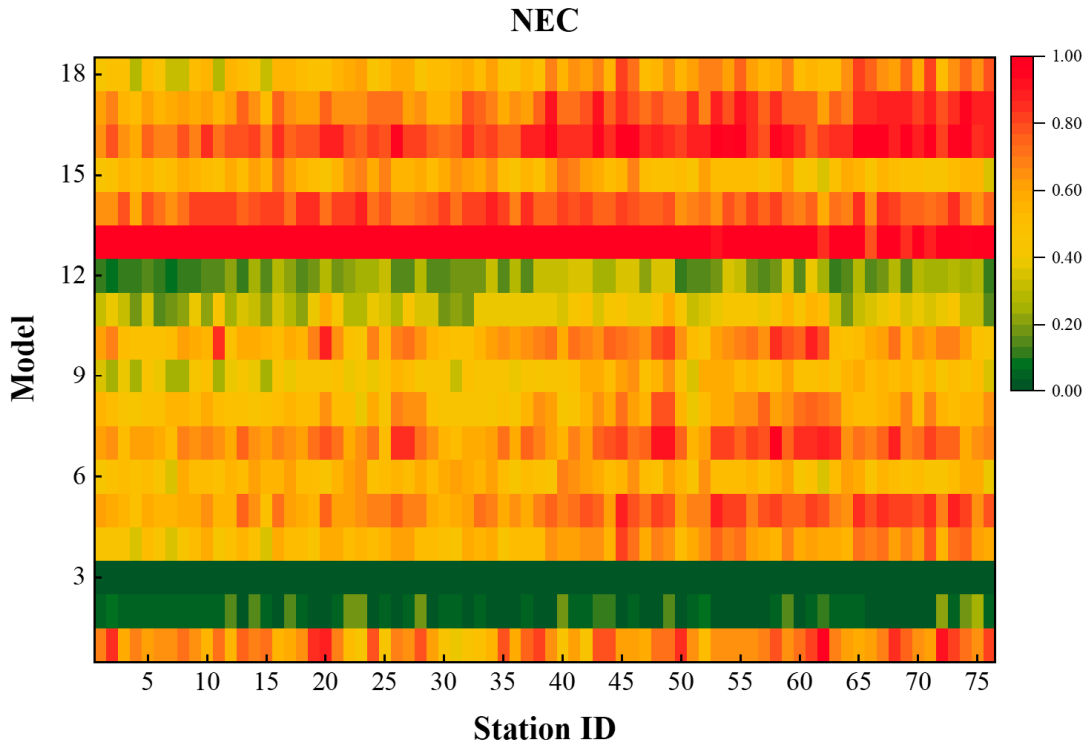

Figure S9. Combined indicator values for each model at all stations in the NEC region.

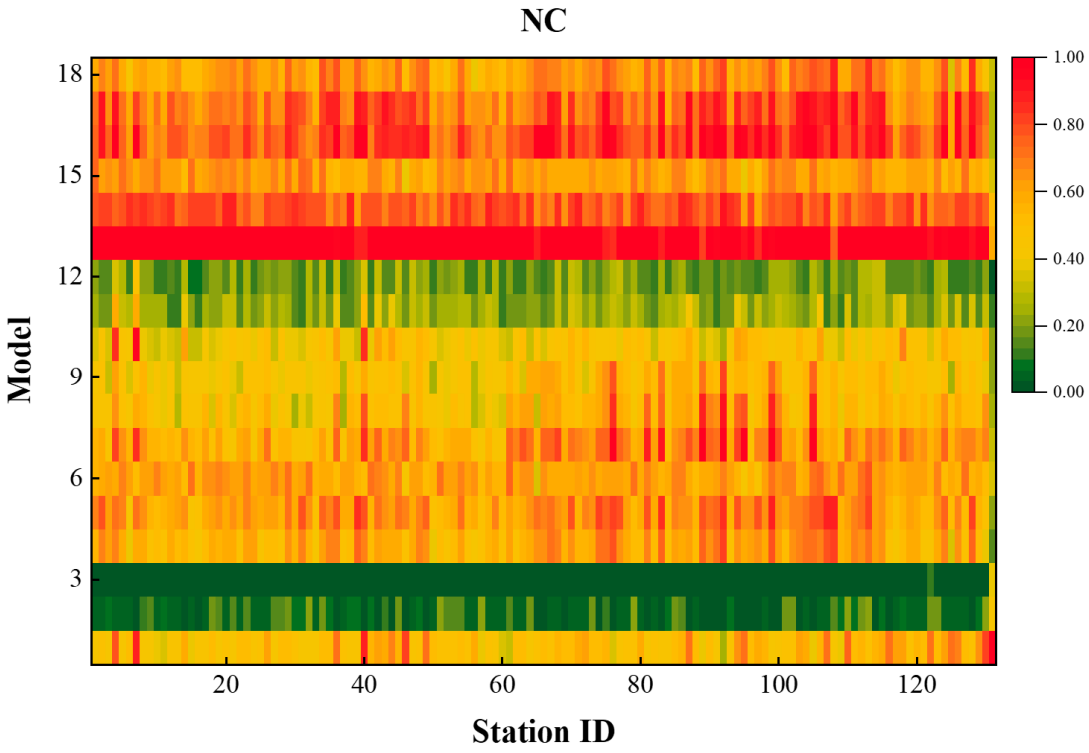

Figure S10. Combined indicator values for each model at all stations in the NC region.

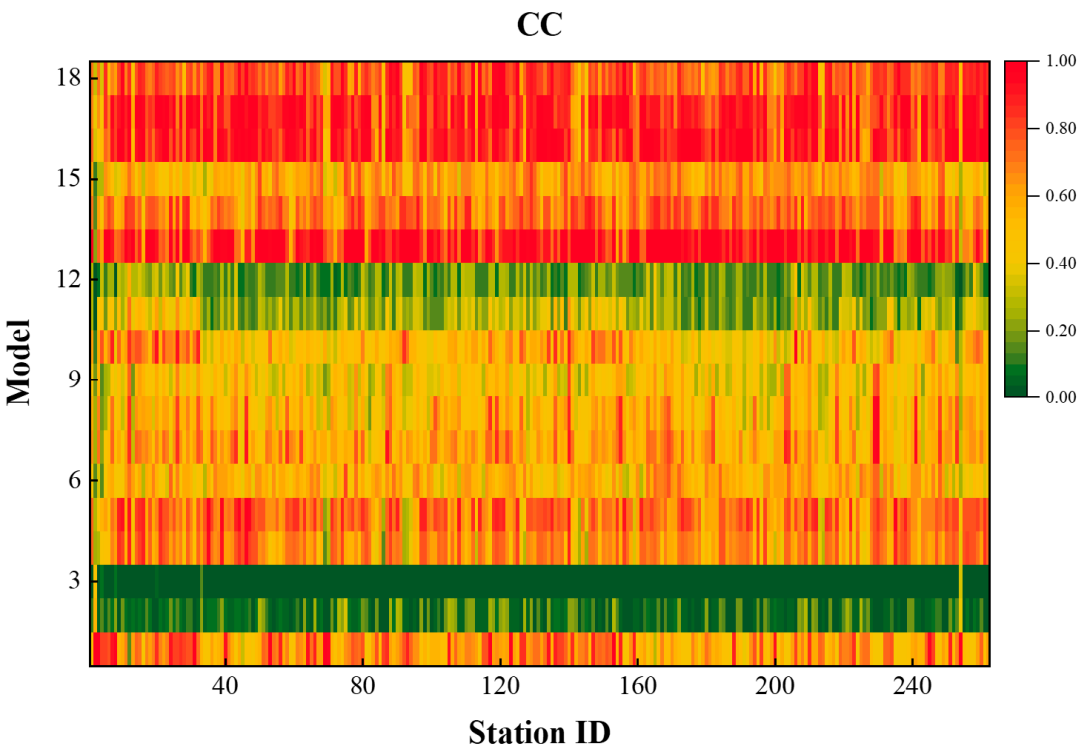

Figure S11. Combined indicator values for each model at all stations in the CC region.

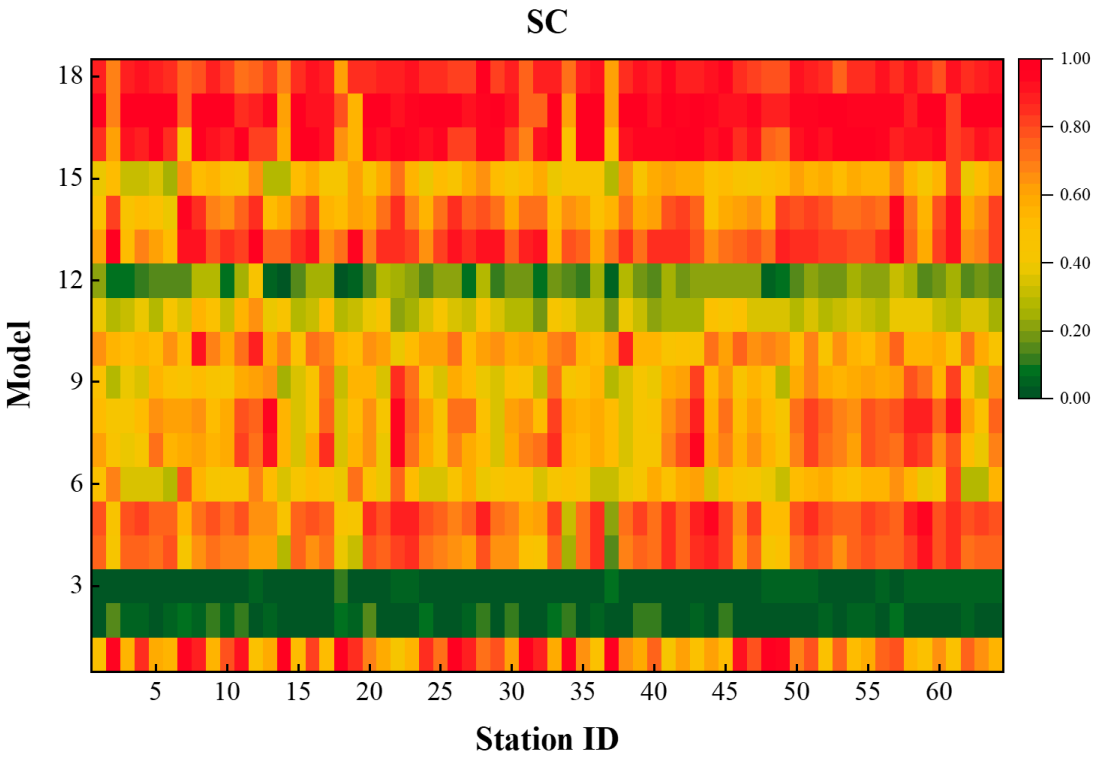

Figure S12. Combined indicator values for each model at all stations in the SC region.

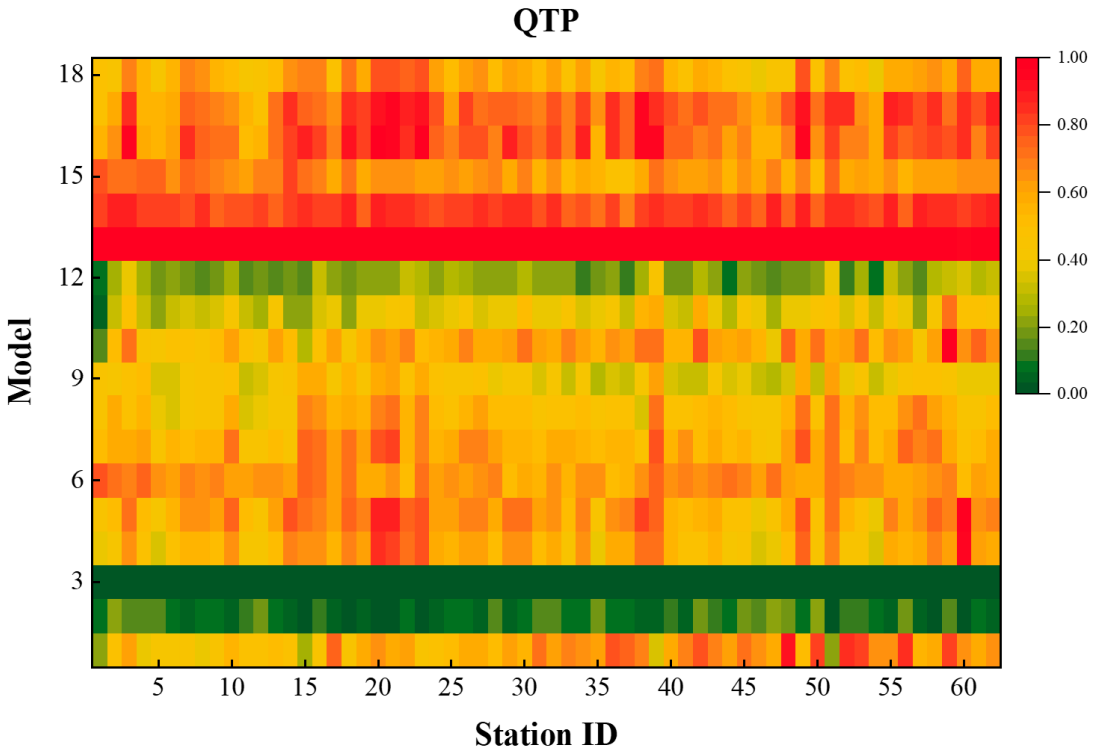

Figure S13. Combined indicator values for each model at all stations in the QTP region.