

Supplementary Materials to paper

This supplementary material consists of details in the following **2 Sections**.

Section S1 refers to the principle of boosted regression trees (BRT) model.

Section S2 includes the differences in overall prediction accuracy, daily prediction accuracy, and spatial coverage between the optimized strategies obtained and the single strategies using the BRT model.

Supplementary Material Section S1.

The principle of Boosted Regression Trees model

Boosted regression tree (BRT) combines boosting and regression trees to improve predictive performance compared to single models. Boosting is a forward and stagewise process that iteratively generates a series of regression trees by randomly selecting subsets of data (Rätsch et al., 2001). In BRT, each tree is trained based on the performance of its predecessor, progressively correcting the errors made by previous trees, which distinguishes it from bagging methods like RF. During this process, stochastic gradient boosting is introduced to further enhance model accuracy and reduce the risk of overfitting (Elith et al., 2008). The iterative process is based on recursive binary splitting of tree models, allowing the identification of poorly modeled observations until the model deviation (i.e., loss function) is minimized (Groemping, 2006). The final model's fitted values are calculated by aggregating the outputs from all trees and multiplying them by the learning rate (Elith et al., 2008).

The advantages of BRT include its ability to handle complex nonlinear relationships between covariates without the need for data normalization. The algorithm also automatically incorporates interactions between variables without explicit specification. In BRT modeling process, two main parameters are the learning rate (LR) and tree complexity (TC). LR determines each tree's contribution to the final model, while TC controls tree depth and whether to account for interactions between variables. In this study, the default parameters of the BRT algorithm were used, with the learning rate set at 0.01 and tree complexity at 1. The BRT algorithm was implemented using "gbm" package in R (Ridgeway, 2024).

Supplementary Material Section S2.

Prediction results based on Boosted Regression Trees model

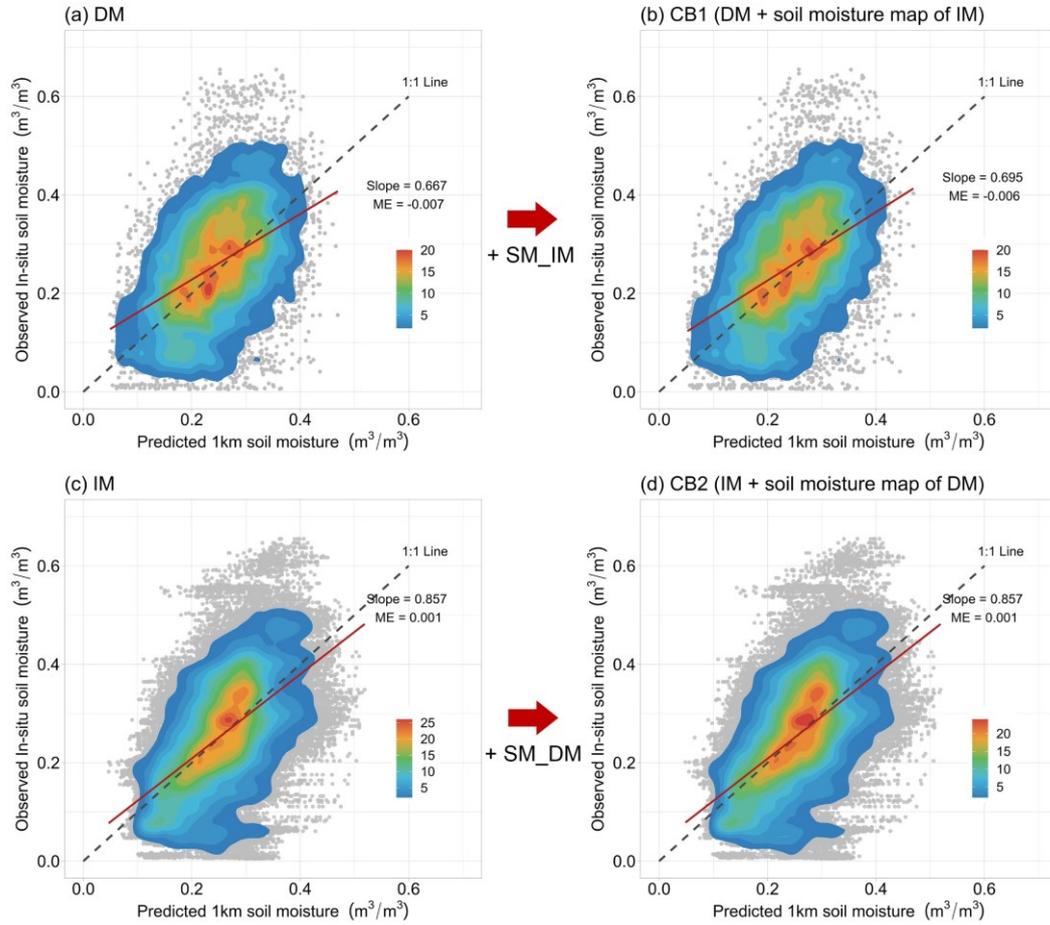


Figure S1. Scatter density plot of soil moisture based on BRT model comparing the overall accuracy of observation and prediction of (a) DM, (b) CB1, (c) IM, and (d) CB2. Red solid lines represent the linear regression fitting for the observed and predicted values, while black dashed line indicates the 1:1 line. The gray dots represent the collection of predicted values and corresponding observed values from the test set across all available time period and sites in the independent dataset validation. The captions SM_IM and SM_DM refer to the soil moisture maps generated by the IM and DM, respectively.

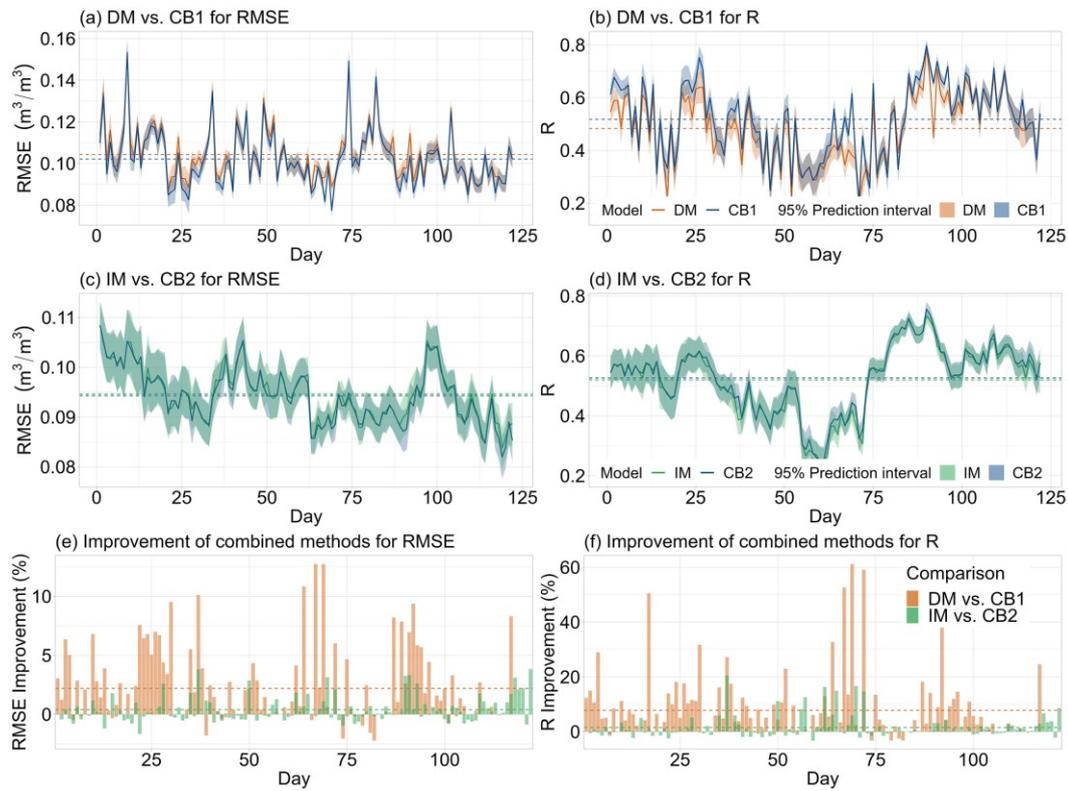


Figure S2. Daily prediction performance and differences for the single methods (DM and IM) and the combined methods (CB1 and CB2) using BRT. (a) and (b): daily comparisons of RMSE and R between DM and CB1; (c) and (d): daily comparisons of RMSE and R between IM and CB2; (e) and (f): percentage improvement in RMSE and R of CB1 and CB2 relative to their respective single method (DM and IM). Day 1 corresponds to June 1, 2016. An average value over the available time periods is represented by a horizontal dashed line, and shaded areas represent the 95% prediction intervals.

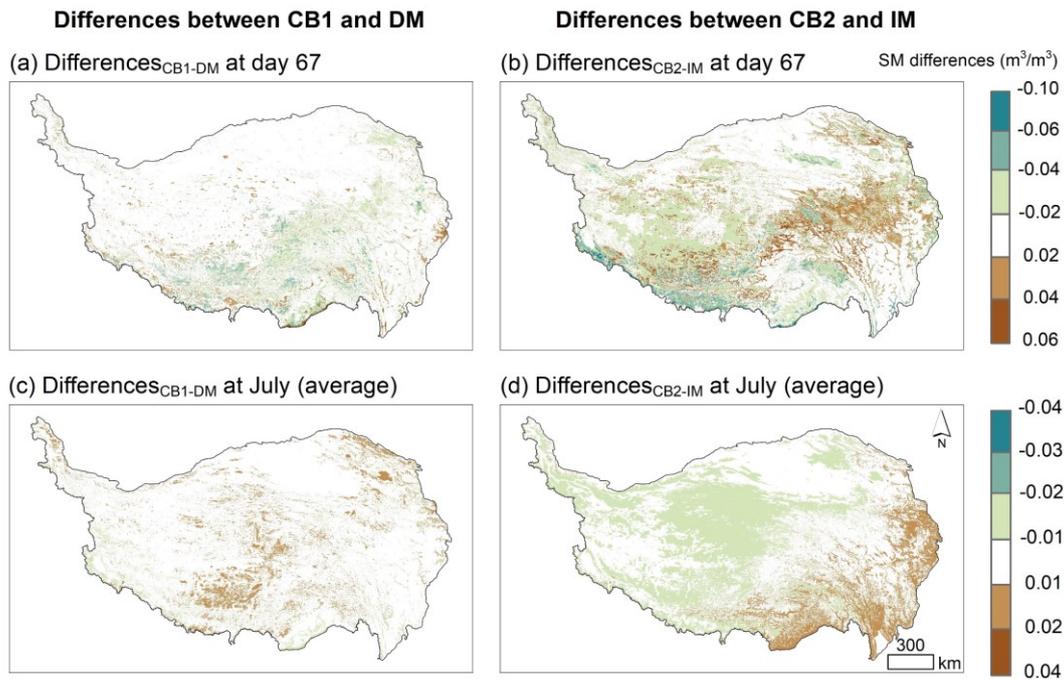


Figure S3. Difference mapping for spatial coverage in 1-km downscaled or interpolated soil moisture maps generated using single (DM and IM) and combined (CB1 and CB2) methods across different temporal resolutions based on BRT: (a) and (b) daily scales; (c) and (d) monthly scales.

References

Elith, J., Leathwick, J., Hastie, T., 2008. A working guide to boosted regression trees. *The Journal of animal ecology* 77, 802-813.

Groemping, U., 2006. Relative importance for linear regression in r: The package relaimpo. *Journal of Statistical Software* 17(1), 1 - 27.

Rätsch, G., Onoda, T., Müller, K.R., 2001. Soft margins for adaboost. *Machine Learning* 42(3), 287-320.

Ridgeway, G., 2024. *gbm: Generalized boosted regression models*, R package version 2.2.2 ed.