Supplement of Earth Syst. Sci. Data, 14, 3053–3073, 2022
https://doi.org/10.5194/essd-14-3053-2022-supplement

Supplement of

Daily soil moisture mapping at 1 km resolution based on SMAP data for desertification areas in northern China

Pinzeng Rao et al.

Correspondence to: Fang Wang (657563390@qq.com)

Part 1: Implementation of machine learning methods

The regression models based on machine learning approaches are constructed in R language. The main R packages used are shown in Table 1.

Table. S1 Mainly used R packages d to implement the algorithm

Packages	Functions	Description
'stats'	lm	Building the MLR regression model
'e1071'	svm	Building the SVM regression model
'nnet'	nnet	Building the ANN regression model
'randomForest'	randomForest	Building the RF regression model
'xgboost'	xgboost	Building the XGB regression model
'e1071'	tune	Hyperparameter optimization for SVM, ANN and RF
'caret'	train	Hyperparameter optimization for XGB

Part 2: Hyperparameter optimization results

In addition to MLR, SVM, ANN, RF and XGB all contain some important hyperparameters (Table S2). Among them, some hyperparameters are very critical to the simulation accuracy, and they are tuned using grid search 10-fold CV. The optimization result is shown in Table S3.

Table S2 Important hyperparameters and set value ranges for SVM, ANN, RF and XGB

Algorithm	Key parameters	Range	Description
SVM	kernel	radial	The kernel used in training and predicting.
	gamma	(0.001, 0.5)	The parameter needed for kernels

	cost	(1, 5)	Cost of constraints violation
ANN	size	(5, 20)	Number of units in the hidden layer.
	decay	(0, 0.1)	The parameter for weight decay
	maxit	500	Maximum number of iterations
RF	mtry	(5, 10)	Number of variables randomly sampled as candidates at each split
	ntree	(50, 500)	Number of trees to grow
	nodesize	(5, 10)	Minimum size of terminal nodes
XGB	nrounds	(50, 500)	Max number of boosting iterations.
	max_depth	(5, 10)	Maximum depth of a tree
	eta	(0.1, 0.5)	The parameter control the learning rate
	gamma	0	Minimum loss reduction required to make a further partition on a leaf node of the tree
	colsample_bytree	1	Subsample ratio of columns when constructing each tree
	min_child_weight	1	Minimum sum of instance weight (hessian) needed in a child
	subsample	1	Subsample ratio of the training instance

Table S3 The optimization results of hyperparameters

DOY	SVM		ANN		RF		XGB		
	gamma	cost	size	decay	ntree	mtry	nrounds	max_depth	eta
2015081	0.2	3	18	0.02	500	5	95	5	0.2
2015097	0.3	3	10	0.02	230	6	140	6	0.1
2015113	0.4	3	17	0.02	320	5	95	9	0.1
2015129	0.3	4	16	0.01	365	5	140	5	0.1

2015145	0.3	2	20	0.03	410	5	95	8	0.1
2015161	0.3	2	12	0.01	320	5	140	6	0.1
2015177	0.5	2	19	0.02	455	5	185	5	0.1
2015193	0.5	4	17	0.01	275	5	320	6	0.1
2015209	0.3	2	14	0.02	455	5	140	7	0.1
2015225	0.3	5	16	0.03	365	5	500	9	0.1
2015241	0.3	2	18	0.02	410	5	140	5	0.1
2015257	0.3	3	20	0.03	185	7	140	6	0.1
2015273	0.4	5	19	0.01	230	9	500	7	0.1
2015289	0.3	4	14	0.01	500	5	320	10	0.1
2015305	0.2	5	16	0.01	455	5	230	9	0.1
2015321	0.3	5	13	0.03	320	6	230	5	0.1
2015337	0.1	5	10	0.02	230	10	95	7	0.1
2015353	0.1	5	11	0.01	95	7	50	5	0.1
2016001	0.1	5	14	0.01	50	9	500	6	0.1
2016017	-	-	-	-	-	-	-	-	-
2016033	0.1	5	15	0.01	95	6	185	5	0.1
2016049	0.1	4	12	0.01	185	5	185	7	0.2
2016065	0.2	5	16	0.02	410	8	230	9	0.1
2016081	0.3	4	20	0.03	455	5	185	5	0.1
2016097	0.3	4	16	0.03	410	5	140	7	0.1
2016113	0.3	5	20	0.02	455	9	500	6	0.1
2016129	0.3	4	17	0.03	455	5	185	5	0.1
2016145	0.4	3	18	0.03	455	5	95	6	0.1
2016161	0.2	3	18	0.03	455	7	50	8	0.1
2016177	0.3	3	16	0.02	500	5	140	6	0.1
2016193	0.3	3	18	0.02	410	6	140	5	0.1
2016209	0.4	3	20	0.01	455	5	140	5	0.1
2016225	0.3	2	15	0.03	320	5	140	7	0.1
2016241	0.5	3	19	0.05	365	5	275	7	0.1
2016257	0.4	3	18	0.03	410	5	185	7	0.1
2016273	0.4	4	20	0.04	410	5	230	8	0.1
2016289	0.3	4	16	0.03	500	7	140	6	0.1
2016305	0.3	5	20	0.04	320	6	365	7	0.1
2016321	0.2	5	10	0.03	95	6	500	6	0.2
2016337	0.1	5	10	0.02	140	5	95	6	0.1
2016353	0.1	5	18	0.02	365	5	95	5	0.1
2017001	0.1	4	16	0.01	320	6	185	6	0.1
2017017	0.1	5	17	0.02	320	5	95	5	0.1
2017033	0.1	5	10	0.03	140	5	95	9	0.1
2017049	0.2	4	10	0.02	185	5	95	5	0.2
2017065	0.4	5	16	0.03	365	5	140	8	0.1

2017081	0.2	3	12	0.02	455	7	95	5	0.1
2017097	0.3	4	10	0.01	230	8	185	5	0.1
2017113	0.4	4	18	0.05	320	6	185	6	0.1
2017129	0.4	4	16	0.04	230	6	185	6	0.1
2017145	0.2	4	16	0.01	500	5	365	7	0.1
2017161	0.3	3	19	0.02	500	5	185	5	0.1
2017177	0.3	4	15	0.01	365	5	140	8	0.1
2017193	0.3	2	17	0.02	320	5	95	6	0.1
2017209	0.4	2	18	0.02	320	5	140	5	0.2
2017225	0.3	2	16	0.03	455	5	185	6	0.1
2017241	0.4	3	20	0.02	365	6	230	5	0.1
2017257	0.4	3	18	0.02	500	5	230	6	0.1
2017273	0.4	5	18	0.01	185	7	455	8	0.2
2017289	0.4	5	16	0.02	455	6	320	6	0.1
2017305	0.3	5	16	0.02	365	5	275	7	0.2
2017321	0.2	5	16	0.04	365	6	140	5	0.1
2017337	0.2	5	10	0.03	320	5	95	6	0.1
2017353	0.1	3	20	0.02	365	7	185	5	0.2
2018001	0.2	4	15	0.04	140	9	50	6	0.1
2018017	-	-	-	-	-	-	-	-	-
2018033	0.1	4	16	0.02	320	6	95	5	0.1
2018049	0.1	3	10	0.03	455	5	230	5	0.1
2018065	0.2	4	10	0.02	320	5	230	7	0.1
2018081	0.2	3	18	0.04	410	5	230	7	0.2
2018097	0.2	4	18	0.02	365	5	230	6	0.1
2018113	0.4	5	14	0.02	455	5	275	8	0.1
2018129	0.4	3	20	0.01	95	9	95	7	0.1
2018145	0.2	4	13	0.02	500	5	185	6	0.1
2018161	0.3	2	18	0.03	275	5	95	10	0.1
2018177	0.2	4	16	0.01	230	6	140	5	0.1
2018193	0.2	2	20	0.04	365	5	140	5	0.1
2018209	0.4	3	15	0.02	365	5	185	5	0.1
2018225	0.2	3	20	0.03	185	5	140	5	0.1
2018241	0.3	3	14	0.02	365	5	95	7	0.1
2018257	0.3	3	18	0.02	140	5	320	6	0.1
2018273	0.4	4	14	0.02	365	8	365	6	0.1
2018289	0.3	5	13	0.04	275	5	275	10	0.1
2018305	0.3	5	10	0.02	455	10	320	8	0.1
2018321	0.2	5	14	0.03	365	6	185	8	0.1
2018337	0.1	3	10	0.02	320	8	185	7	0.3
2018353	-	-	-	-	-	-	-	-	-
2019001	-	-	-	-	-	-	-	-	-

2019017	0.1	5	13	0.01	95	7	95	6	0.1
2019033	0.2	3	10	0.02	500	5	50	6	0.3
2019049	0.3	5	12	0.02	410	5	140	7	0.1
2019065	0.2	4	14	0.03	500	5	185	8	0.1
2019081	0.3	5	18	0.02	455	10	275	8	0.1
2019097	0.2	4	12	0.04	455	5	185	7	0.1
2019113	0.3	2	16	0.02	230	5	95	6	0.1
2019129	0.2	5	16	0.06	410	5	140	6	0.1
2019145	0.2	4	18	0.02	455	7	230	9	0.1
2019161	0.1	2	14	0.02	455	6	95	10	0.1
2019177	-	-	-	-	-	-	-	-	-
2019193	0.2	3	18	0.04	275	5	140	8	0.1
2019209	0.4	4	20	0.02	500	5	185	6	0.1
2019225	0.2	2	16	0.02	410	5	95	5	0.1
2019241	0.5	3	18	0.02	455	5	365	5	0.1
2019257	0.5	5	16	0.01	140	5	410	7	0.1
2019273	0.3	4	20	0.03	500	6	275	7	0.1
2019289	0.4	3	20	0.04	320	6	275	6	0.1
2019305	0.4	5	18	0.03	455	7	275	8	0.1
2019321	0.2	5	12	0.02	275	5	230	5	0.2
2019337	0.1	5	10	0.02	95	7	230	9	0.1
2019353	0.1	5	10	0.03	140	7	185	5	0.2
2020001	0.1	2	15	0.03	95	6	95	6	0.4
2020017	0.1	2	11	0.03	320	5	95	6	0.3
2020033	0.1	2	10	0.02	500	5	230	9	0.2
2020049	0.2	5	11	0.02	455	5	185	8	0.1
2020065	0.2	5	10	0.01	455	6	275	9	0.1
2020081	0.2	5	12	0.02	140	5	320	6	0.1
2020097	0.3	5	17	0.03	320	7	185	8	0.1
2020113	0.4	5	20	0.04	365	5	500	8	0.1
2020129	0.4	4	12	0.02	455	8	140	6	0.1
2020145	0.4	4	16	0.01	320	6	140	5	0.1
2020161	0.3	3	14	0.02	500	5	95	5	0.1
2020177	0.4	2	18	0.03	275	5	140	5	0.1
2020193	0.3	3	16	0.03	500	5	140	6	0.1
2020209	0.4	3	20	0.01	365	5	140	5	0.1
2020225	0.5	3	18	0.01	140	5	230	5	0.1
2020241	0.5	3	13	0.02	500	5	275	6	0.1
2020257	0.4	5	18	0.03	410	6	365	6	0.1
2020273	0.3	4	16	0.04	500	6	275	6	0.1
2020289	0.4	5	15	0.02	275	9	320	8	0.1
2020305	0.4	5	15	0.02	365	6	275	7	0.1

2020321	0.2	5	16	0.02	500	6	95	5	0.1
2020337	0.1	5	10	0.03	140	6	185	7	0.3
2020353	0.1	5	12	0.02	140	9	230	7	0.3

Part 3: Auxiliary evaluation results

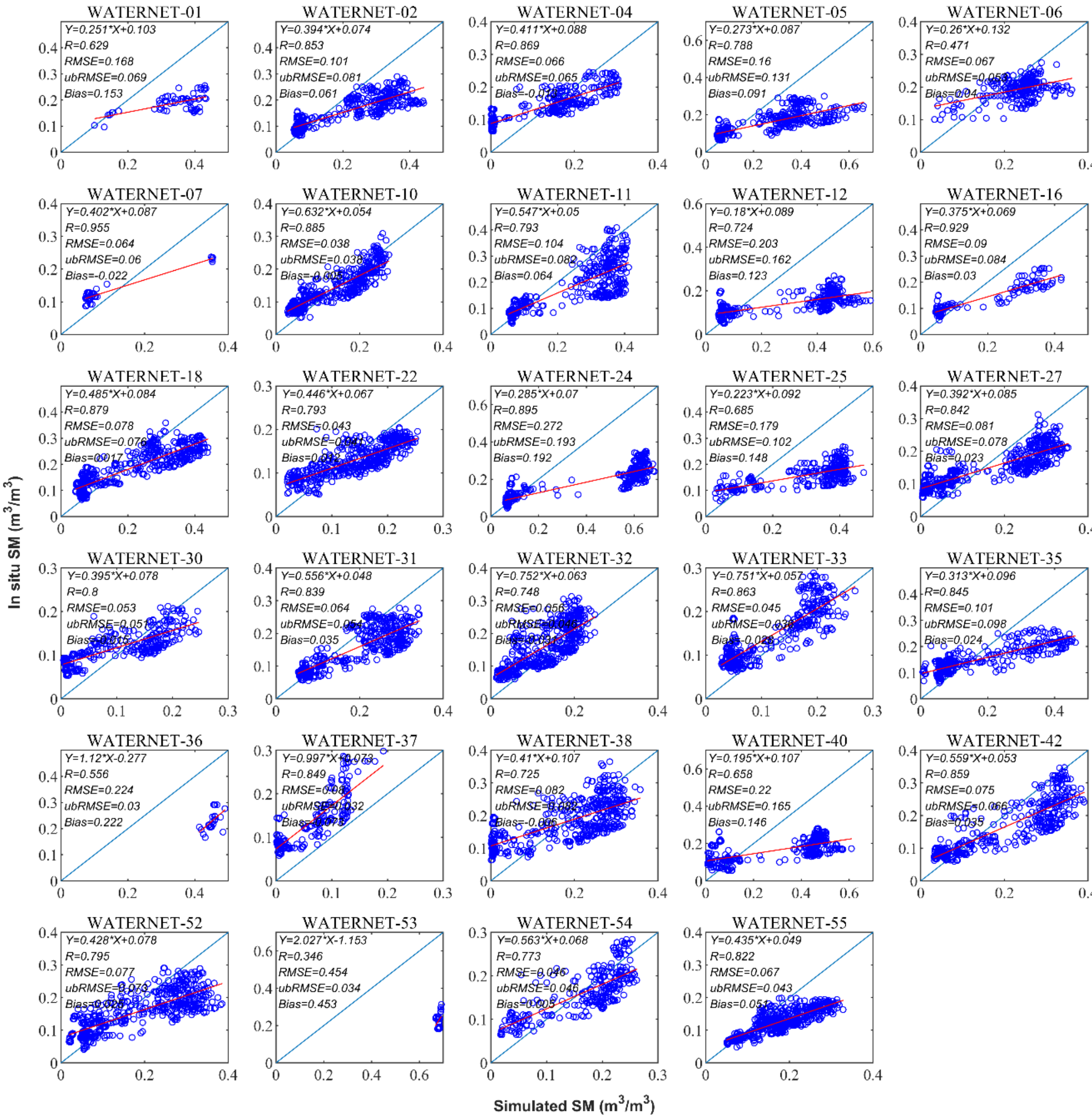

Figure S1: Comparison between the downscaled SM and in situ SM of Babao Network.

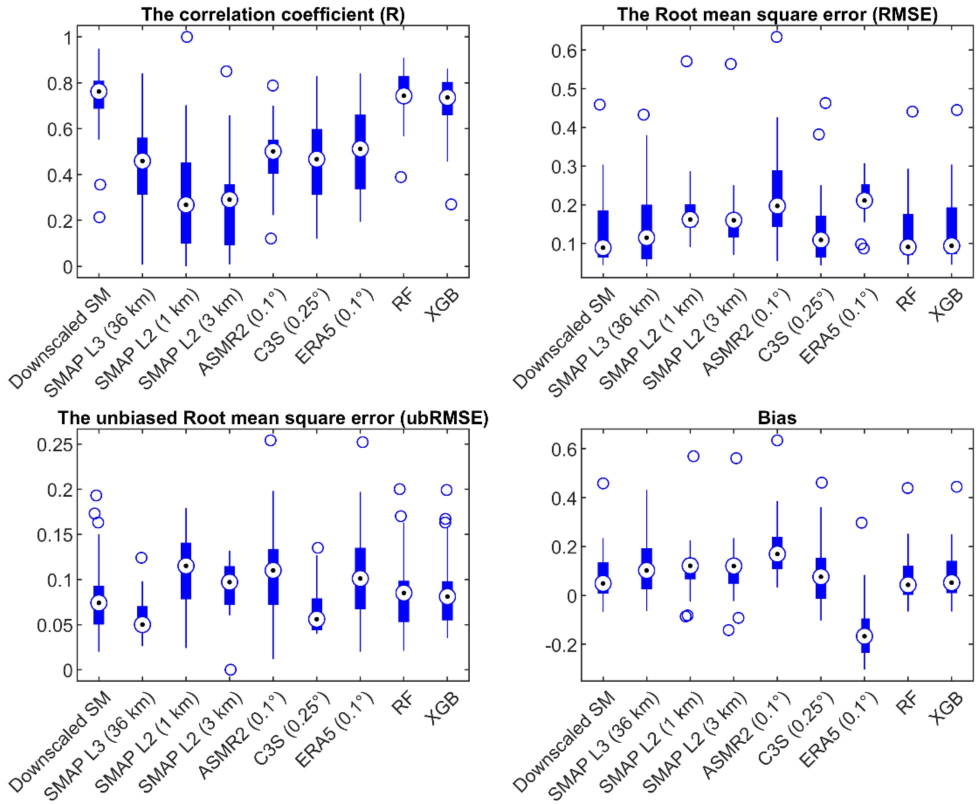

Figure S2: Comparison of gridded products and in situ observation SM of the Babao Network.

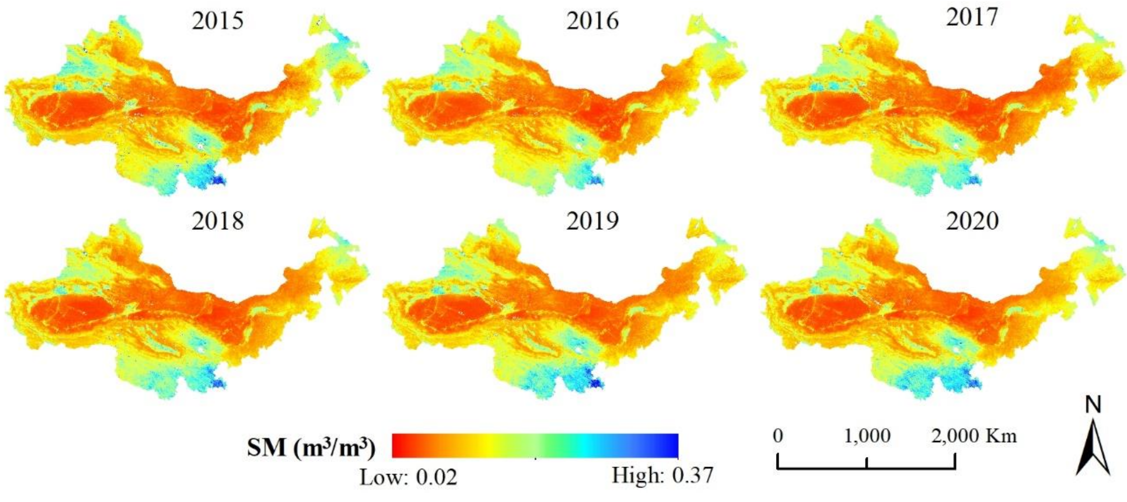

Figure S3: Annual average SM in the study area. Note: The SM data for 2015 are only from 2015/3/22.