Supplementary materials for: "Downscaling satellite-derived soil moisture in the Three North region using ensemble machine learning and multiple-source knowledge integration"
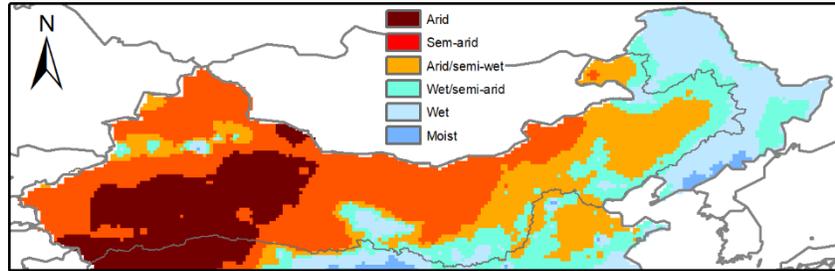
Figure S1. Climate zones across Northern China are delineated based on the China Humidity Index map, a product of the National Earth System Science Data Center, National Science & Technology Infrastructure of China.
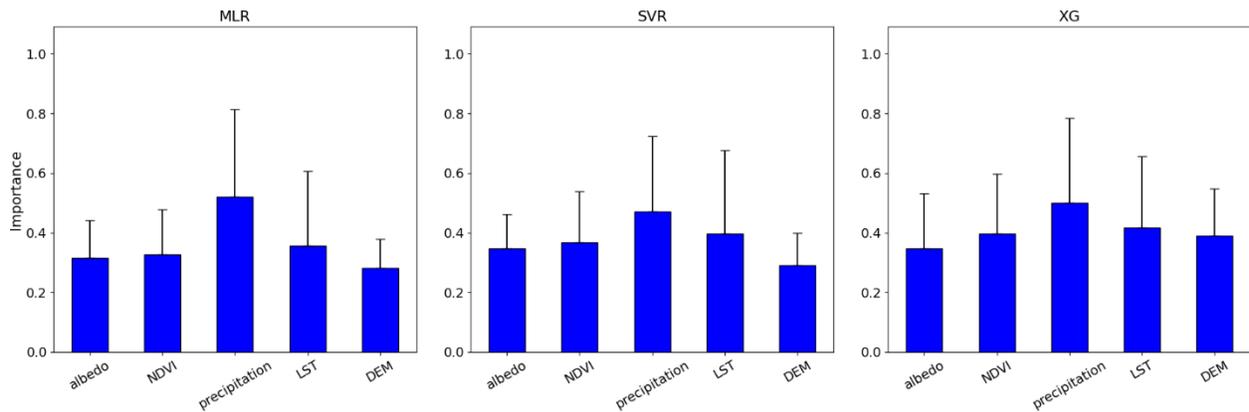
Figure S2. Average (blue bar) and standard deviation (error bar) of permutation-based importance of explanatory variables for soil moisture: comparison across multiple linear regression (MLR), support vector regression (SVR), and extreme gradient XG Boost (XG).
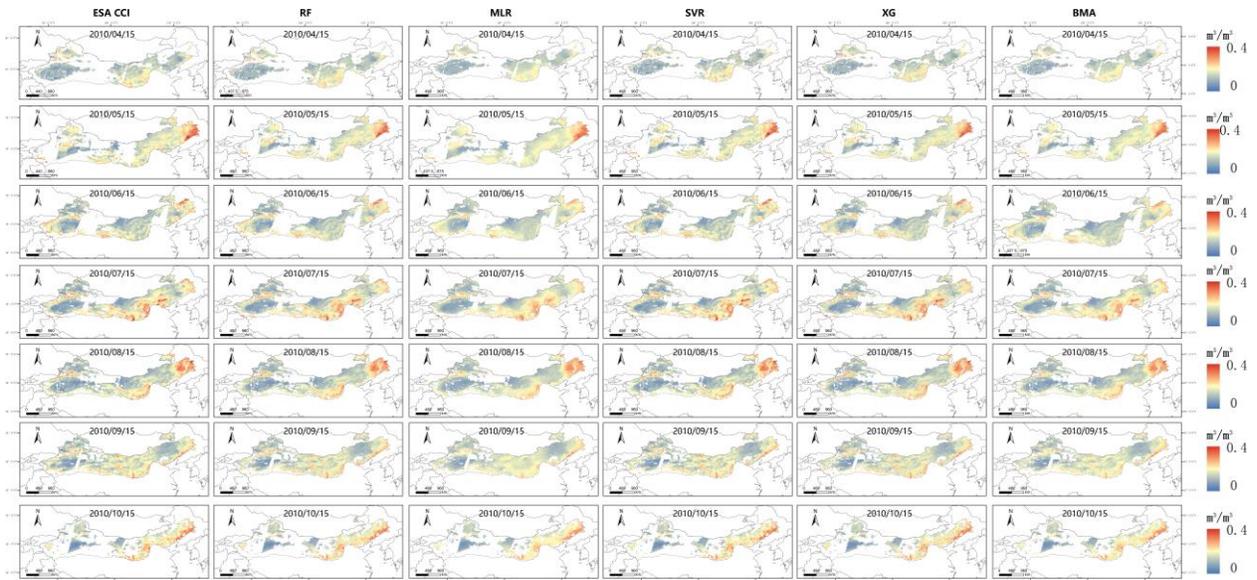
Figure S3. Spatial distribution of soil moisture across six data sources, representing the 15th day of April-October 2010. Columns, from left to right, show the 25-km ESA CCI soil moisture and the 1-km downscaled soil moisture derived from random forest (RF), multiple linear regression (MLR), support vector regression (SVR), extreme gradient XG Boost (XG), and Bayesian model averaging (BMA), respectively.
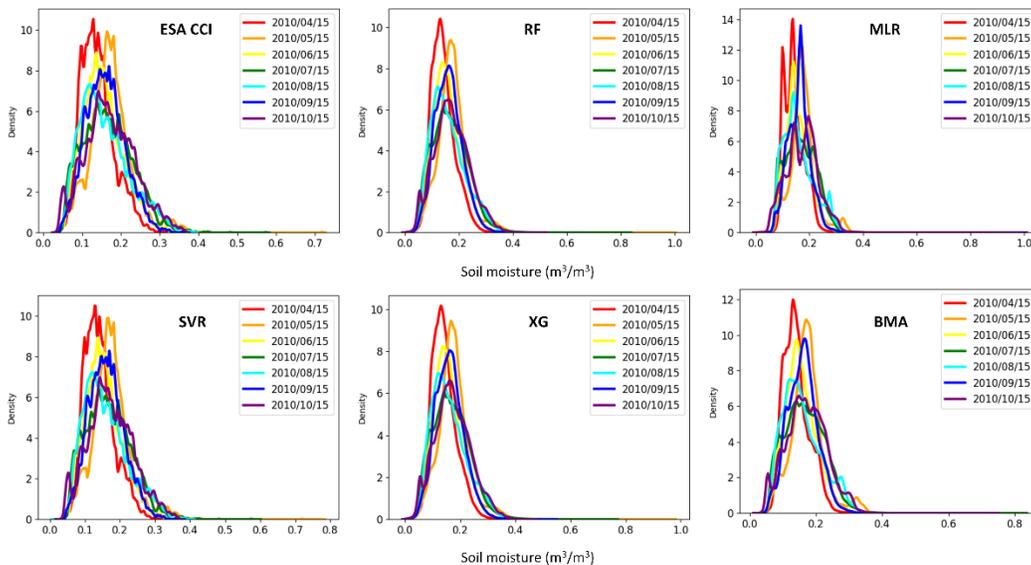
Figure S4. Histograms depicting soil moisture data on the 15th day of April-October 2010. Histograms show the comparison between 25-km ESA CCI soil moisture and 1-km downscaled soil moisture using random forest (RF), multiple linear regression (MLR), support vector regression (SVR), extreme gradient XG Boost (XG), and Bayesian model averaging (BMA).
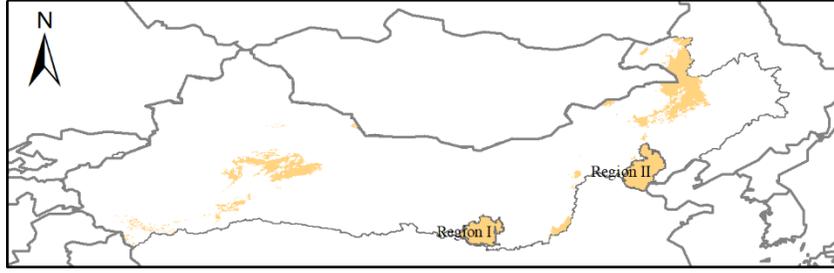
Figure S5. Delineation of typical arid zones based on the Palmer Drought Severity Index (PDSI) and highlights two priority arid zones, namely Region I and Region II.
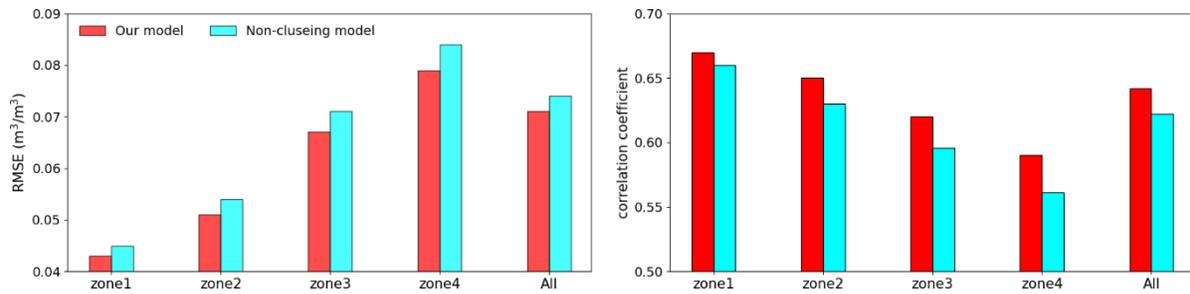
Figure S6. Comparison of accuracy between our proposed model and a model without clustering before fitting, for four divided zones and all regions.
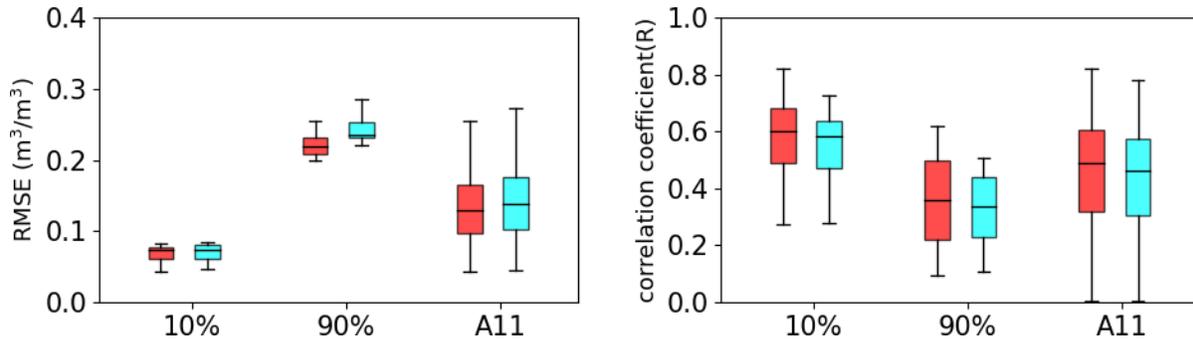
Figure S7. Average model accuracy for our proposed model (red) and the global model without spatiotemporal searching window (cyan) across all sites, considering 10% and 90% soil moisture quantiles.
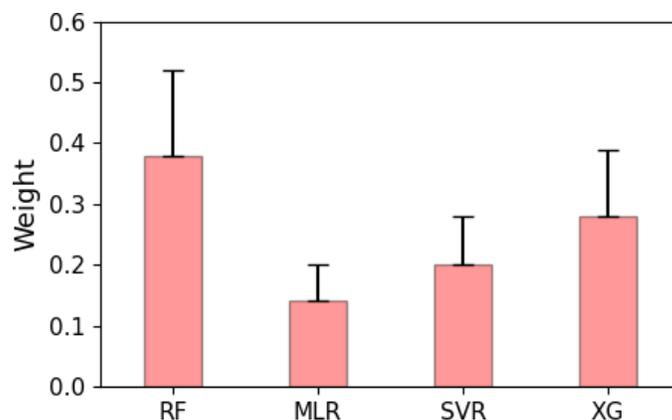
Figure S8. Average (blue bar) and standard deviation (error bar) of weights assigned to random forest (RF), multiple linear regression (MLR), support vector regression (SVR), and extreme gradient XG Boost (XG) in Bayesian model averaging.

Table S1. Information of Chinese Crop Growth and Development and Soil Moisture in Agricultural Land Dataset

Stations	Longitude	Latitude	Stations	Longitude	Latitude
50525	119.75E	49.15N	53646	109.7E	38.23N
50658	125.88E	48.05N	53663	111.81E	38.93N
50739	123.18E	47.33N	53665	111.65E	38.28N
50742	124.48E	47.8N	53704	105.18E	37.53N
50755	126.1E	47.6N	53705	105.67E	37.48N
50756	126.97E	47.43N	53707	105.25E	36.93N
50844	123.42E	46.4N	53723	107.4E	37.78N
50851	126.1E	46.68N	53725	107.58E	37.58N
50854	125.32E	46.38N	53754	110.21E	37.5N
50867	127.35E	46.08N	53759	110.81E	37N
50928	120.33E	45.07N	53764	111.1E	37.5N
50934	121.55E	45.4N	53769	111.76E	37.25N
50936	122.83E	45.63N	53777	112.15E	37.53N
50945	124.27E	45.5N	53806	105.65E	36.57N
50946	124.83E	45.18N	53810	105.91E	36.98N
50949	124.87E	45.08N	53817	106.26E	36N
50953	126.77E	45.75N	53821	107.3E	36.58N

50954	125.08E	45.5N	53845	109.5E	36.6N
50955	126.3E	45.38N	53853	110.95E	36.7N
50958	126.95E	45.52N	53857	110.18E	36.07N
51076	88.08E	47.73N	53859	110.18E	36.1N
51133	83E	46.73N	53863	111.93E	37.05N
51238	82.06E	44.9N	53868	111.5E	36.06N
51334	82.9E	44.62N	53881	106.48E	37.28N
51346	84.66E	44.43N	53903	105.71E	35.96N
51352	85.25E	44.85N	53906	105.72E	35.52N
51353	86.1E	45.02N	53908	105.4E	35.11N
51358	85.81E	44.28N	53914	106.12E	35.62N
51359	86.2E	44.32N	53915	106.67E	35.55N
51368	87.43E	44.01N	53916	106.32E	35.5N
51379	89.57E	44.02N	53923	107.63E	35.73N
51431	81.33E	43.95N	53924	107.4E	35.15N
51436	83.3E	43.45N	53925	107.18E	35.68N
51437	81.13E	43.15N	53926	107.71E	35.4N
51463	87.62E	43.78N	53930	107.98E	36.45N
51469	87.12E	43.28N	53937	108E	35.42N
51572	89.25E	42.83N	53938	108.33E	35.11N
51573	89.2E	42.93N	53941	109.58E	35.18N
51628	80.23E	41.16N	53942	109.5E	35.81N
51629	80.23E	41.27N	53948	109.58E	34.95N
51642	84.25E	41.78N	53955	110.45E	35.46N
51644	82.95E	41.72N	53956	110.83E	35.4N
51656	86.13E	41.75N	54026	120.9E	44.57N
51709	75.98E	39.47N	54027	119.4E	43.83N
51716	78.57E	39.8N	54049	123.97E	44.25N
51730	81.05E	40.5N	54064	125.16E	44.41N
51777	88.17E	39.03N	54069	125.8E	44.17N

51810	77.63E	38.92N	54072	126.51E	44.81N
51811	77.27E	38.43N	54076	126.93E	44.42N
51828	79.93E	37.13N	54080	127.15E	44.9N
51855	85.55E	38.15N	54102	116.12E	43.95N
51931	81.66E	36.86N	54134	121.28E	43.6N
52101	93E	43.6N	54135	122.27E	43.6N
52203	93.52E	42.82N	54154	124.3E	43.35N
52418	94.68E	40.15N	54156	124.8E	43.52N
52533	98.48E	39.77N	54213	119.03E	42.95N
52546	99.83E	39.37N	54218	118.97E	42.27N
52652	100.43E	38.93N	54223	120.65E	42.85N
52656	100.82E	38.45N	54236	122.53E	42.42N
52674	101.96E	38.23N	54237	121.65E	42.03N
52679	102.66E	37.91N	54243	124.12E	42.78N
52681	103.08E	38.63N	54304	115.28E	40.96N
52737	97.37E	37.37N	54305	115.28E	41.9N
52749	99.58E	37.08N	54308	116.63E	41.22N
52765	101.61E	37.38N	54311	117.75E	41.93N
52784	102.9E	37.48N	54318	117.73E	41.31N
52797	104.05E	37.18N	54319	118.66E	41N
52818	94.9E	36.42N	54324	120.45E	41.55N
52825	96.41E	36.43N	54326	119.7E	41.38N
52836	98.1E	36.3N	54333	122.8E	41.98N
52853	100.98E	36.92N	54335	122.01E	41.68N
52855	101.08E	36.75N	54337	121.12E	41.13N
52856	100.62E	36.27N	54404	115.83E	40.9N
52863	101.95E	36.82N	54405	115.5E	40.4N
52868	101.43E	36.03N	54408	115.21E	40.38N
52876	102.85E	36.32N	54416	116.87E	40.38N
52885	103.25E	36.75N	54420	117.33E	40.93N

52895	104.68E	36.56N	54423	117.93E	40.97N
52896	104.18E	36.55N	54425	117.48E	40.41N
52980	103.3E	35.97N	54429	117.95E	40.2N
52983	104.15E	35.87N	54430	118.15E	40.76N
52984	103.18E	35.58N	54431	116.63E	39.92N
52985	103.35E	35.43N	54432	118.48E	40.6N
52986	103.86E	35.36N	54434	118.3E	40.15N
52993	105.08E	35.68N	54437	118.66E	39.5N
52995	104.6E	35.53N	54439	118.71E	40N
53289	113.85E	42.25N	54452	119.82E	40.8N
53357	110.05E	41.03N	54454	120.35E	40.35N
53368	111.45E	41.1N	54502	115.96E	39.48N
53378	112.61E	41.3N	54506	115.85E	39.33N
53384	113.18E	41.45N	54510	116.98E	39.88N
53392	114.6E	41.85N	54512	116.28E	39.43N
53397	113.95E	41.06N	54515	116.7E	39.5N
53399	114.7E	41.15N	54518	116.38E	39.12N
53420	107.13E	40.9N	54519	116.48E	39.31N
53433	108.65E	40.73N	54520	117.08E	39.96N
53455	110.53E	40.55N	54521	116.98E	39.76N
53464	111.15E	40.68N	54522	117.73E	39.88N
53478	112.45E	40N	54525	117.28E	39.7N
53483	113.86E	40.88N	54531	118.75E	39.73N
53486	113.77E	40.37N	54532	118.1E	39.8N
53487	113.33E	40.1N	54533	118.12E	39.55N
53491	114.4E	40.68N	54534	118.15E	39.67N
53492	114.16E	40.11N	54539	118.9E	39.42N
53498	115.03E	40.56N	54540	119.16E	39.71N
53505	105.38E	39.08N	54612	116.45E	38.86N
53513	107.4E	40.77N	54619	116.91E	38.91N

53547	109.03E	39.1N	54636	116.1E	38.98N
53553	110.86E	39.66N	56092	104.65E	35N
53564	111.15E	39.38N	57002	105.67E	34.87N
53578	112.43E	39.3N	57004	104.88E	34.73N
53593	114.57E	39.83N	57006	105.75E	34.58N
53611	106.55E	38.91N	57012	106.2E	34.98N
53614	106.22E	38.48N	57025	107.38E	34.51N
53615	106.68E	38.81N	57030	108.15E	34.7N
53618	106.23E	38.25N	57033	108.82E	34.55N
53619	106.3E	38.12N	57034	109.91E	34.88N

Table S2. Hyperparameters selection ranges of machine learning models

Model	Hyperparameters	Represents	List values for tuning
Random Forest (RF)	n_estimators	number of trees	[50, 100, 200]
	max_depth	maximum depth of each tree	[None, 5, 10, 15]
	min_samples_split	minimum number of samples required to split an internal node	[2, 5, 10]
	min_samples_leaf	minimum number of samples required to be in a leaf node	[1, 2, 4]
Support Vector Machine (SVR)	kernel	type of decision boundary	[Radial basis function (RBF), Sigmoid]
	C	trade-off between achieving a smooth decision boundary and classifying the training points correctly	[0.1, 0.3, 1, 3, 10, 100]
Extreme gradient boosting (XG)	n-estimators	number of decision trees	[50, 1000]
	max depth	Maximum depth of the tree	[3, 10]
	learning rate	Shrinkage factors	[0.01, 0.3]
	subsample	The proportion of training sub-samples	[0.5, 1.0]
	colsample bytree	Column sampling	[0.5, 1.0]
	gamma	Regularization parameter of T	[0, 5]
	reg-alpha	L1 regularization parameter	[0, 3]
	reg-lambda	L2 regularization parameter	[0, 3]