KADF：A knowledge-augmented deep fusion method for estimating near-surface air temperature

Fengrui Chen[a,b*], Xi Li[c], Yiguo Wang[d]

[a] Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions, Ministry of Education, Henan University, Kaifeng, Henan 475001, China.

[b] College of Geography and Environmental Science, Henan University, Kaifeng, Henan 475001, China.

[c] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei 430079, China.

[d] Nansen Environmental and Remote Sensing Center, Bergen 5004, Norway.

1. About

This document describes a basic application of the KADF method to create an improved surfaced air temperature (Ta). KADF integrates three categories of prior knowledge concerning Ta: spatial autocorrelation, temporal autocorrelation, and temporal heterogeneity in the relationship between Ta and predictive variables. This tailored strategy enables the model to more efficiently explore the intricate spatiotemporal relationships between grounded Ta observations and satellite-derived auxiliary variables, culminating in accurate Ta estimates through the deep fusion of these datasets. We use as case study Henan (Chinese mainland), as an example of how to generate this product at daily temporal scale and at 1 km spatial resolution, from January to December 2010. This example requires the following data from the user: 1) time series of daily mean temperature (T_{mean}) observations, 2) the spatial coordinates of the rain gauges, 3) the seamless, global MODIS-like daily LST dataset (Zhang et al., 2021), generated from the MYD11A1 and MOD11A1, 4) the MODIS MOD13A2 product's Enhanced Vegetation Index (EVI) products, and 4) the Shuttle Radar Topography Mission (SRTM-v4) digital elevation model (DEM).

2. Input data for training

The input data required for the KADF model are listed as follows: DLST, NLST, EVI, DEM, disM, and TMean. They are provided in the .mat format. The data format for DLST, NLST, EVI, and TMean are shown in Fig.1. The first row contains the EIDs of meteorological stations. Each subsequent row represents a day's data, and each column represents the data time series for a specific station location.

58017	58111	54817	54900	54901	54902	54903	58001	58004	58005	58006	58007	58008	58100	58101	58104
12.7000	12	7.4000	7.3000	9.3000	11	7.9000	12.1000	13	10.6000	9.1000	12.1000	12.9000	14.2000	8.1000	9
9.3000	8.9000	3	1.8000	4.2000	5.1000	3	5.2000	4.6000	5.4000	6	5.9000	6.2000	8.5000	5.3000	7.5000
9.2000	10.3000	-1.9000	1.2000	5.8000	5.6000	0.3000	6	4.6000	5.6000	6.6000	4.6000	6.4000	8.7000	6	7.1000
4	3.9000	1.3000	1.3000	2.8000	4.4000	1.4000	0.7000	0.4000	1.4000	1.6000	1.9000	1.8000	5.3000	1.6000	3.4000
3.5000	1.9000	1.6000	1.5000	2.4000	4.4000	0.7000	-0.1000	-0.6000	0.3000	1.7000	-0.8000	0.5000	3.6000	0.6000	0.5000
5.5000	5.5000	1.4000	1.1000	2.7000	4.7000	0.2000	5	4.7000	4.2000	2.9000	3.3000	5	7	3	5.1000
4.4000	4.4000	-0.9000	0	0.6000	3.8000	-0.8000	2.9000	2.3000	3	2.1000	0.6000	3.5000	5.7000	1.7000	3.3000
6.4000	6	-2	-0.5000	1	2.6000	-1.6000	2.8000	2.3000	3.5000	3.6000	2.4000	3.6000	8.7000	3.7000	7.6000
10	9.2000	1.6000	0.6000	2.5000	3.2000	1.3000	6.6000	5.8000	7.6000	7.4000	6.6000	7.5000	10.3000	7	7.7000
8.3000	7.3000	2.9000	2.8000	3.1000	4.5000	3.2000	5.7000	5.5000	4.8000	5.5000	4.6000	6.1000	8.8000	6	7.1000
3.1000	2	1.4000	1.9000	5.8000	6.2000	-0.1000	1.5000	1.9000	2.9000	1.5000	-2.2000	2.4000	0.9000	-1.6000	-1.5000
7.2000	8.1000	2.4000	2.4000	2.8000	5.2000	2	5	1.3000	2.9000	7.4000	6.5000	5.3000	7.4000	5.1000	8.6000
4	4.6000	4	4.2000	5.3000	6.1000	4.4000	4.2000	4.2000	5.2000	4.6000	3.7000	5	5.4000	4.4000	6.1000
12.8000	12.4000	9.8000	10.9000	12.9000	13.7000	9.6000	13.8000	12.6000	13.2000	13	10.6000	12.5000	13	11.1000	12.4000

Fig.1 Data format for DLST, NLST, EVI, and TMean.

The DEM file (DEM.mat) contains EID, Lat, Lon, and elevation, descripting the spatial information of meteorological stations. The data format for the file DEM.mat is as follows:

58017	34.2500	116.1300	44
58111	33.9670	116.4500	35
54817	35.9830	115.8700	43
54900	35.7000	115.0200	54
54901	36.0830	115.1700	48
54902	35.9500	115.1500	50
54903	35.8500	115.4800	46
58001	34.4330	115.1000	57
58004	34.6500	115.1500	63
58005	34.4330	115.5300	50
58006	34.3830	115.8800	48
58007	34.0670	115.3000	50
58008	34.4670	115.3300	56
58100	33.6500	115.1700	41
58101	33.8670	115.4800	42
58104	33.4000	115.0700	44

Fig.2 Data format for the file DEM.mat.

The file disM.mat is a distance matrix for meteorological stations, calculating the distance from each meteorological station to the others. It is a diagonal matrix, represented as follows:

0	0.4272	1.7524	1.8261	2.0692	1.9622	1.7270	1.0461	1.0585	0.6273	0.2832	0.8499	0.8289	1.1321	0.7
0.4272	0	2.0978	2.2468	2.4730	2.3711	2.1182	1.4282	1.4685	1.0313	0.7057	1.1543	1.2265	1.3187	0.9
1.7524	2.0978	0	0.8959	0.7071	0.7208	0.4121	1.7307	1.5150	1.5869	1.6000	1.9990	1.6093	2.4358	2.1
1.8261	2.2468	0.8959	0	0.4113	0.2818	0.4838	1.2695	1.0580	1.3658	1.5729	1.6568	1.2714	2.0555	1.8
2.0692	2.4730	0.7071	0.4113	0	0.1345	0.3878	1.6515	1.4331	1.6888	1.8423	2.0202	1.6239	2.4330	2.2
1.9622	2.3711	0.7208	0.2818	0.1345	0	0.3448	1.5178	1.3000	1.5639	1.7287	1.8890	1.4939	2.3001	2.1
1.7270	2.1182	0.4121	0.4838	0.3878	0.3448	0	1.4671	1.2445	1.4179	1.5206	1.7921	1.3911	2.2217	1.9
1.0461	1.4282	1.7307	1.2695	1.6515	1.5178	1.4671	0	0.2227	0.4300	0.7816	0.4171	0.2325	0.7861	0.6
1.0585	1.4685	1.5150	1.0580	1.4331	1.3000	1.2445	0.2227	0	0.4376	0.7773	0.6020	0.2567	1.0002	0.8
0.6273	1.0313	1.5869	1.3658	1.6888	1.5639	1.4179	0.4300	0.4376	0	0.3536	0.4323	0.2029	0.8618	0.5
0.2832	0.7057	1.6000	1.5729	1.8423	1.7287	1.5206	0.7816	0.7773	0.3536	0	0.6605	0.5564	1.0205	0.6
0.8499	1.1543	1.9990	1.6568	2.0202	1.8890	1.7921	0.4171	0.6020	0.4323	0.6605	0	0.4011	0.4368	0.2
0.8289	1.2265	1.6093	1.2714	1.6239	1.4939	1.3911	0.2325	0.2567	0.2029	0.5564	0.4011	0	0.8325	0.6
1.1321	1.3187	2.4358	2.0555	2.4330	2.3001	2.2217	0.7861	1.0002	0.8618	1.0205	0.4368	0.8325	0	0.3
0.7544	0.9751	2.1516	1.8898	2.2376	2.1090	1.9830	0.6817	0.8497	0.5682	0.6529	0.2691	0.6185	0.3784	

Fig. 3 Data format for the file disM.mat.

3. Input data for mapping

The input data for mapping is contained within four folders as well as the TMean.mat file. The LOC-Mapping folder includes three files: Lat.mat, Lon.mat, and DEM.mat, which respectively represent the longitude, latitude, and elevation of each estimated location in the study area. The LST-Mapping and EVI-Mapping folders contain the daily values of LST and EVI at each location in the study area. The UT-Mapping folder contains the Mmatrix.mat file, which are the indexes of each estimated location to the four nearest Ta observations.

4. Runing KADF

4.1 Run program

Run the KADF.py file to perform model training, testing, and spatial mapping of T_{mean} in the study area.

4.2 Model evaluation

The result folder contains the observed and estimated values for independent samples. The validation result for the example of T_{mean} estimation in Henan Province in 2010 is as follows:

Table 1. Overall estimation performance

	R^2	MAE（℃）	RMSE（℃）
2010	0.998	0.52	0.68

4.3 Ta mapping results

The results of the temperature mapping can be found in the Mapping folder. The files are saved as integers and should be multiplied by 0.01 to convert to the actual temperature values (℃). The average temperature map for Henan Province on January 6, 2010, is shown in Figure 3.

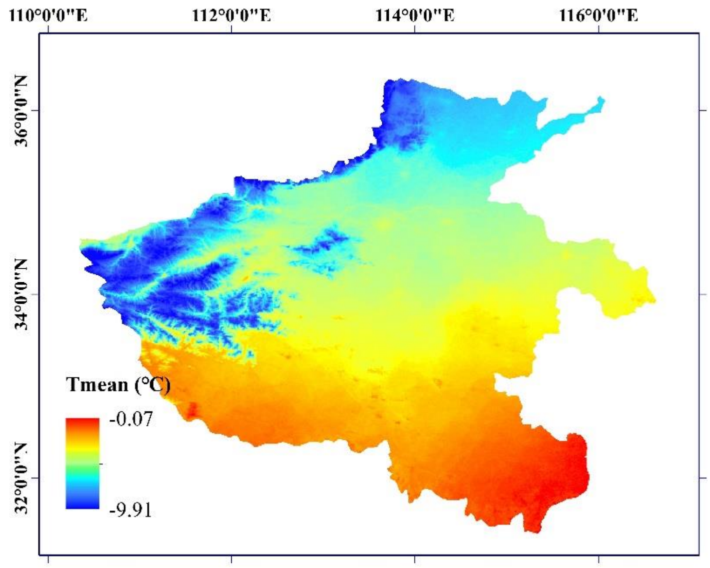

Fig. 4. Spatial distribution of T_{mean} in Henan Province on January 6, 2010.