A Robust Hybrid Deep Learning Model for Spatiotemporal Image Fusion

Zijun Yang, Chunyuan Diao and Bo Li

Supplementary Material

1. Generation of simulation data

The CDL data of the main study area in 2017 are downloaded, re-projected, and co-registered with the Landsat and MODIS images. Figure S1 illustrates how the simulation data are generated using CDL and time-series MODIS imagery.

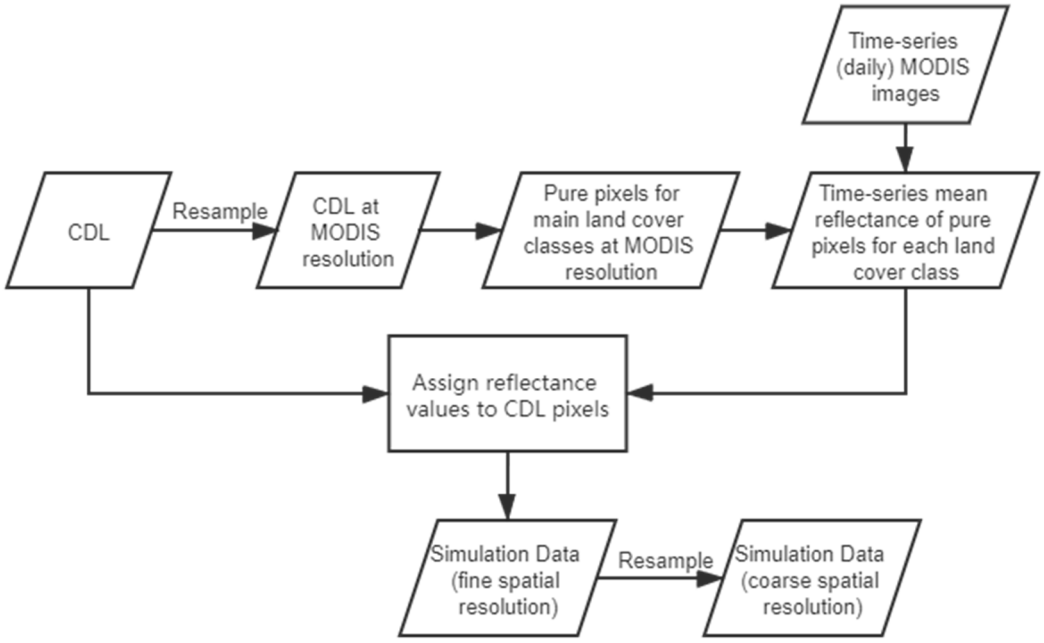

Figure S1. The process of the generation of simulation dataset using CDL and time-series MODIS imagery.

Five major land cover classes are identified in our main study site, namely corn, soybeans, forests, built-up areas, and water bodies. The CDL data in 2017 is first resampled to the MODIS resolution, and the fractions of each land cover class are calculated. For each land cover class, only the pure pixels with corresponding fractions being 100% are selected. The pure-pixel mask for each land cover class will be applied to the daily time-series MODIS images, and daily mean spectral reflectance values for those five land cover classes are calculated from the MODIS data in 2017. Those daily mean spectral reflectance values are further smoothed using the Gaussian filter to reduce the atmospheric and sensor outlying effects.

To generate fine-resolution simulated Landsat images, CDL data with 30-meter spatial resolution are used as the template. Pixels that fall into the five major land covers in the CDL data are retained, while the pixels that do not belong to the five major classes are randomly assigned as corn or soybeans. The pixel values of those five classes in the simulation data are then assigned on a daily basis according to the smoothed mean spectral reflectance values of the classes of the corresponding MODIS data in 2017. Consistent with the spatial resolution of Landsat data, the simulation data are, thus, called simulated Landsat images, with the temporal resolution being one day. The simulated Landsat images are then resampled to the MODIS resolution to generate the corresponding simulated MODIS images. On each date, the simulated MODIS and Landsat images constitute the MODIS-Landsat image pairs for spatiotemporal image fusion. The simulated data will

preserve the spatial distribution of major land cover classes as well as simplify the simulation process. The simulation dataset provides densely time-series images and will facilitate flexible assessments of fusion models' performances under various circumstances of phenological changes. The flexibility of the simulation process is also valuable in that it can be extended and applied to different regions of interests, with a better control over not only the phenological changes but also the land cover composition, landscape configuration, etc.

The simulated images contain six bands: blue, green, red, NIR, SWIR1, and SWIR2. The simulation dataset mimics temporal phenological change patterns in reflectance and the spatial configurations of a multitude of major land cover classes. The spatial extent of the simulation data in this study is the same as that of the satellite data. A segment of simulated images is presented in standard false color composite, alongside the same part of the real MODIS image and the corresponding CDL data (Figure S2).

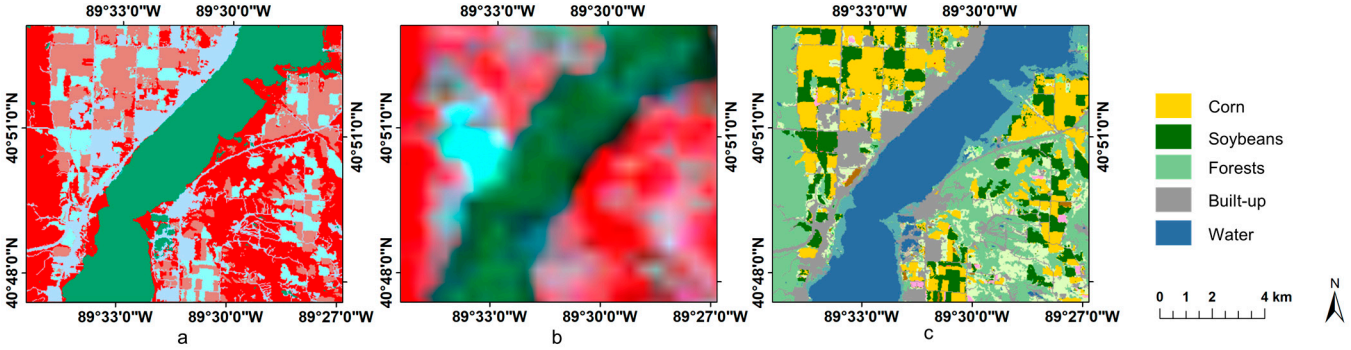

Figure S2. (**a**) A part of the simulated Landsat image on DOY 175 in false color composite; (**b**) the same part of the real MODIS image on DOY 175 in false color composite; (**c**) the same part of CDL data; legend is for the major land cover classes in CDL data. The simulated Landsat image has very similar tones as the real MODIS. The spatial patterns in the simulated Landsat image resemble the patterns in CDL data.

2. Supplementary information about the study sites

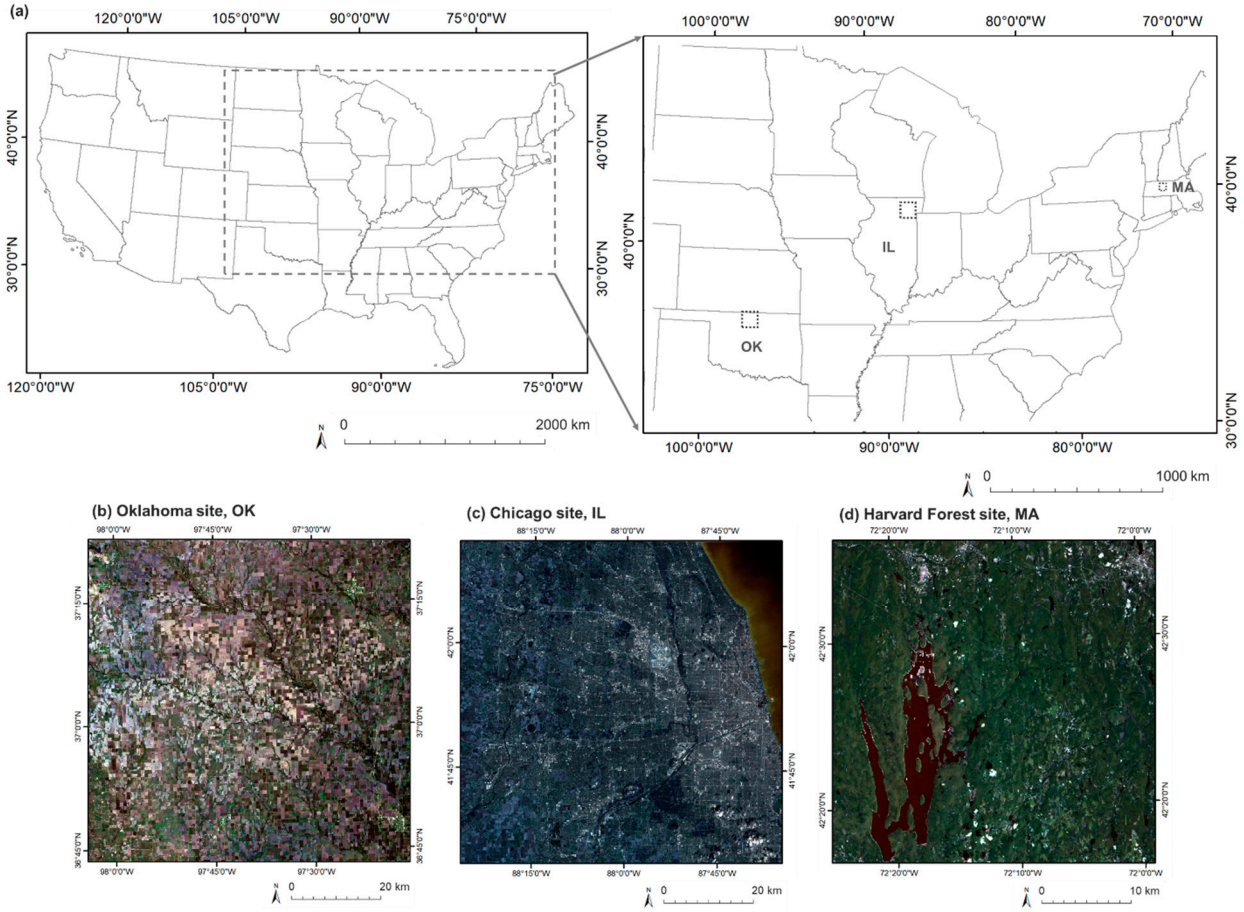

Figure S3. (**a**) The geographic locations of the three additional test sites; (**b**) A Landsat image of the Oklahoma site; (**c**) A Landsat image of the Chicago site; (**d**) A Landsat image of the Harvard Forest site.

Table S1. Land cover composition of the main study site. Data are obtained from 2017 CDL. More than 70% of the main study site are crop fields of corn/soybeans.

Land Cover	Percentage
Corn	37.01%
Soybeans	35.16%
Deciduous Forest	6.78%
Grass/Pasture	5.64%
Developed/Low Intensity	5.29%
Developed/Open Space	4.04%
Developed/Medium Intensity	1.96%
Open Water	1.77%

Note: Land cover types that occupy less than 1% are left out.

Table S2. Land cover composition of the Oklahoma test site. Data are obtained from 2017 CDL. Approximately 70% of the Oklahoma site are crop fields (in italics).

Land Cover	Percentage
Winter Wheat	37.25%
Pasture	20.44%
Soybeans	13.45%
Double Cropping: Winter Wheat/Soybeans	6.14%
Sorghum	3.77%
Cotton	3.48%
Developed/Open Space	3.38%
Corn	2.95%

Deciduous Forest	1.87%
Other Hay/Non Alfalfa	1.77%
Woody Wetlands	1.33%

Note: Land cover types that occupy less than 1% are left out.

Table S3. Land cover composition of the Chicago test site. Data are obtained from 2017 CDL. Over 75% of the Chicago site are developed areas.

Land Cover	Percentage
Developed/Low Intensity	33.58%
Developed/Medium Intensity	19.72%
Developed/Open Space	12.22%
Developed/High Intensity	9.72%
Deciduous Forest	6.47%
Corn	4.37%
Grass/Pasture	3.97%
Soybeans	3.96%
Woody Wetlands	2.38%
Open Water	1.91%

Note: Land cover types that occupy less than 1% are left out.

Table S4. Land cover composition of the Harvard Forest test site. Data are obtained from 2017 CDL. Forests take up more than 70% of the site.

Land Cover	Percentage
Deciduous Forest	46.69%
Evergreen Forest	17.44%
Mixed Forest	9.15%
Open Water	8.99%
Woody Wetlands	4.97%
Developed/Open Space	4.63%
Other Hay/Non Alfalfa	3.47%
Developed/Low Intensity	1.86%

Note: Land cover types that occupy less than 1% are left out.

In terms of the determination of scenarios of phenological changes, the crop pixels are used for the Oklahoma site; the pixels of built-up areas with low density, open space, and deciduous forests are used for the Chicago site; the forest pixels are used for the Harvard Forest site. All images used in this study were acquired in 2017. Specifically, the image acquisition dates (in DOY) for the Oklahoma site include 051, 069, 133, 149, and 229; acquisition dates (in DOY) for the Chicago site include 034, 066, 114, 178, and 258; acquisition dates (in DOY) for the Harvard Forest site include 108, 140, 268, and 300.

3. Supplementary results

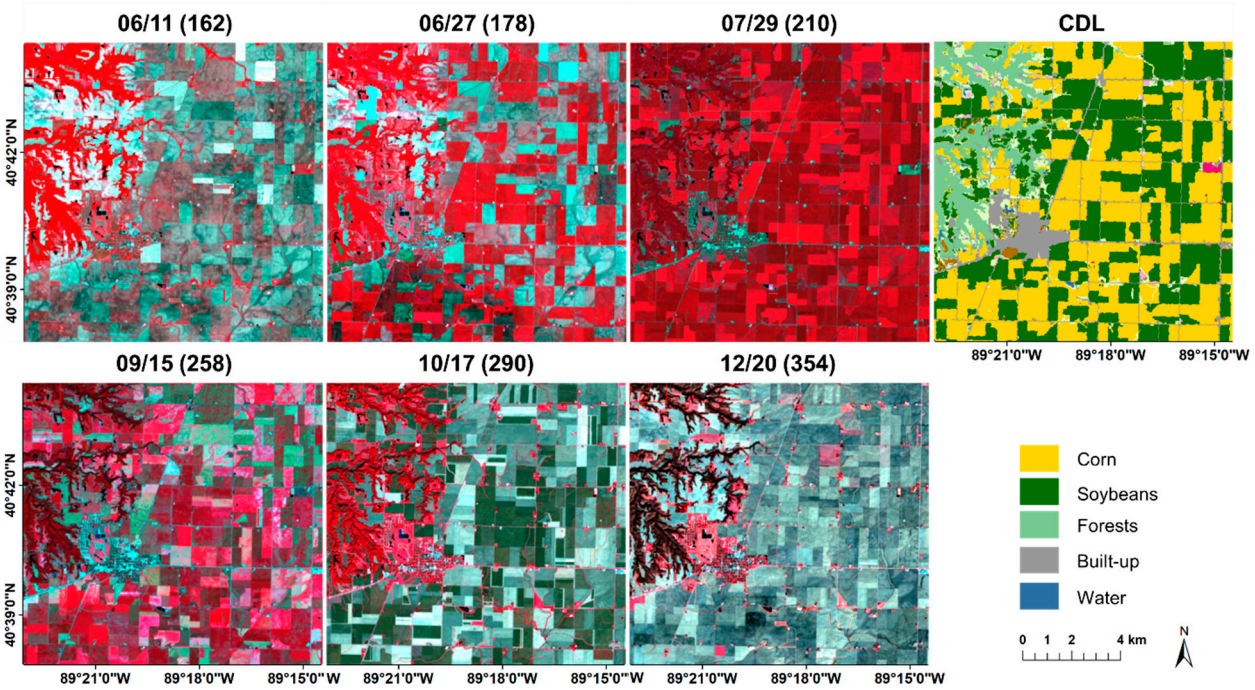

Figure S4. Segments of the six Landsat images (on DOY 162, 178, 210, 258, 290, and 354) in false color composite alongside the CDL data of the corresponding area. Most of the Landsat images have distinct tones, except for images on DOY 162 and 178. Image on DOY 162 has similar crop field patterns as image on DOY 178 but with more pinkish tones.

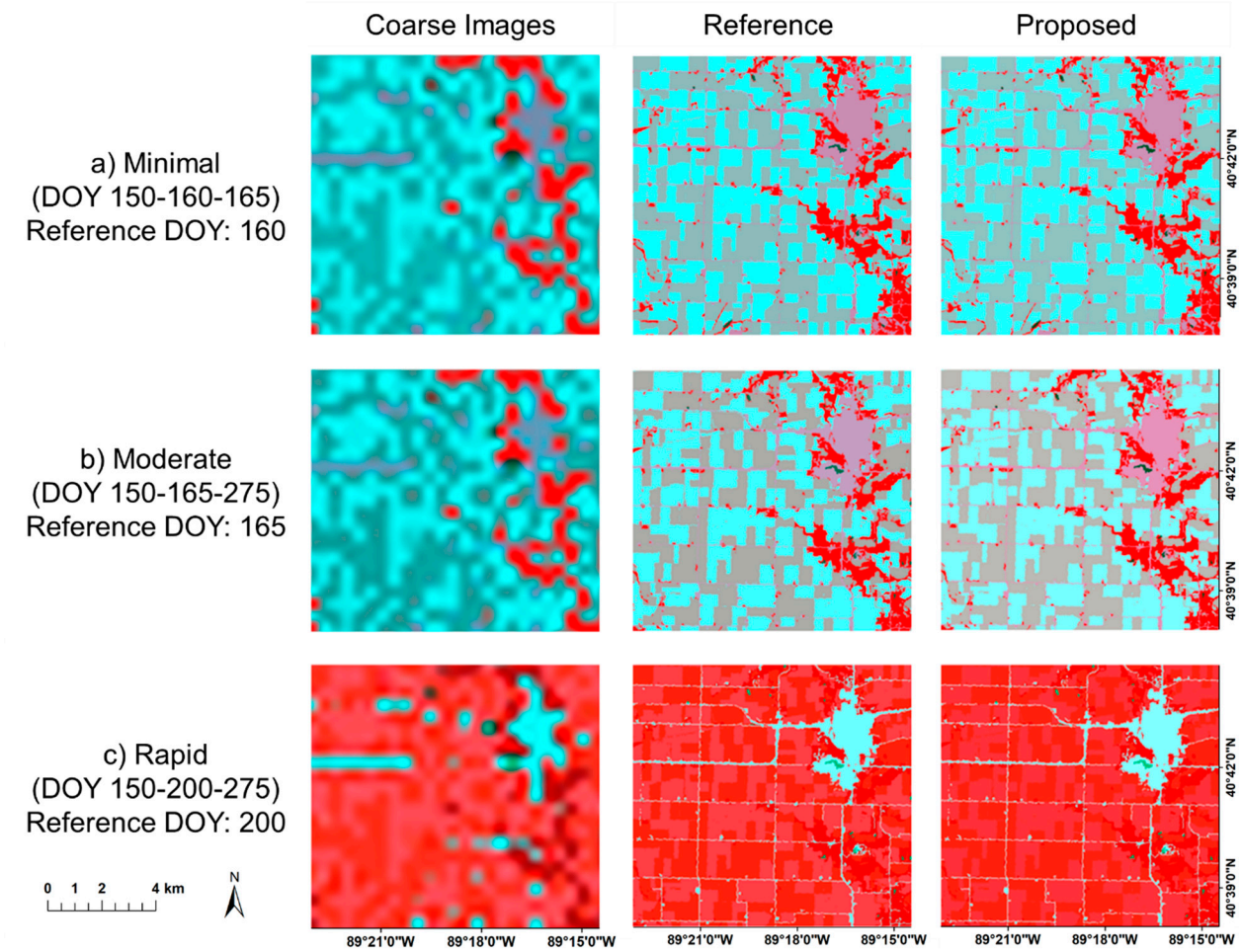

Figure S5. Example segments of predicted images generated by the hybrid deep learning model using simulation data. Row (**a**): the predicted image on DOY 160 with two bracketing dates on DOY 150 and 165, as an example of "minimal" scenario, Row (**b**): the predicted image on DOY 165 with two bracketing dates on DOY 150 and 275, as an example of "moderate" scenario. Row (**c**): the predicted image on DOY 200 with two bracketing dates on DOY 150 and 275, as an example of "rapid" scenario. The coarse images and reference images on these dates are presented in the left two columns.

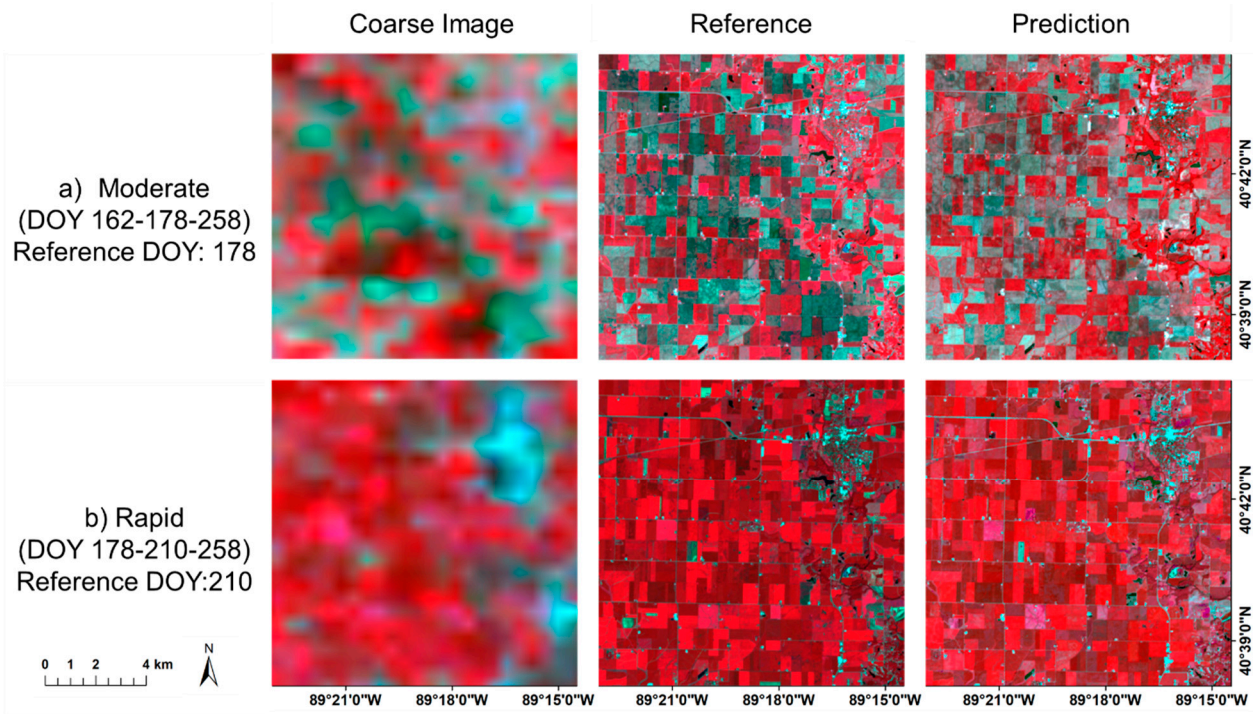

Figure S6. Example segments of predicted images generated by the hybrid deep learning model using satellite data. Row (**a**): the predicted image on DOY 178 with two bracketing dates on DOY 162 and 258, as an example of "moderate" scenario, Row (**b**): the predicted image on DOY 210 with two bracketing dates on DOY 178 and 258, as an example of "rapid" scenario. The coarse images and reference images on these dates are presented in the left two columns.

Table S5. Mean SSIM values for experiments using simulation data and real satellite data in the main study site and the three additional test sites.

	Simulation Data				**Main Site**			
	STARFM	**ESTARFM**	**STFDCNN**	**Hybrid**	**STARFM**	**FSDAF**	**STFDCNN**	**Hybrid**
Blue	0.925	0.950	0.878	**0.985**	0.634	0.714	0.699	**0.723**
Green	0.914	0.946	0.880	**0.983**	0.685	0.709	0.703	**0.712**
Red	0.929	0.920	0.832	**0.981**	0.644	0.674	0.652	**0.689**
NIR	0.934	0.918	0.824	**0.989**	0.634	0.696	0.655	**0.706**
SWIR1	0.941	0.944	0.881	**0.990**	0.560	0.647	0.648	**0.697**
SWIR2	0.885	0.892	0.791	**0.985**	0.518	0.612	0.590	**0.687**

	Oklahoma Site				Chicago Site			
	STARFM	FSDAF	STFDCNN	Hybrid	STARFM	FSDAF	STFDCNN	Hybrid
Blue	0.667	0.718	0.711	**0.734**	0.652	**0.669**	0.660	0.622
Green	0.693	0.737	0.703	**0.743**	0.673	0.660	0.671	**0.682**
Red	0.718	0.781	0.766	**0.789**	0.705	**0.761**	0.710	0.745
NIR	0.513	0.621	0.551	**0.623**	0.724	0.835	0.768	**0.842**
SWIR1	0.732	**0.798**	0.760	0.788	0.715	0.804	0.585	**0.810**
SWIR2	0.696	**0.774**	0.753	0.765	0.699	0.805	0.747	**0.815**

	Harvard Forest Site			
	STARFM	FSDAF	STFDCNN	Hybrid
Blue	0.711	**0.713**	**0.713**	0.708
Green	0.800	**0.802**	0.800	0.799
Red	0.746	0.744	0.740	**0.749**

NIR	**0.902**	0.901	0.899	0.896
SWIR1	**0.890**	0.887	0.881	0.886
SWIR2	0.808	0.807	0.803	**0.812**

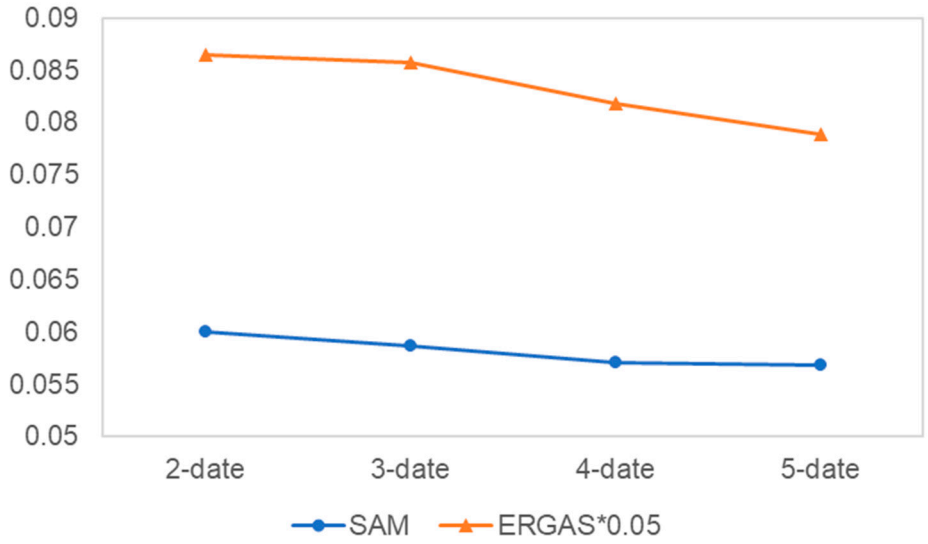

Figure S7. Average SAM and ERGAS values for the predictions of Landsat images on DOY 162, 178, 210, 258, 290, and 354 by varying lengths of temporal sequences. For instance, "2-date" means using the two image pairs acquired on the nearest two dates around each prediction date.