Estimating fine resolution shortwave broadband albedo of croplands from Harmonized Landsat and Sentinel-2 data

Jie Dai,[a] Jiquan Chen,[a,b,c] Cheyenne Lei,[d] Grant Falvo,[e] G Philip Robertson, [a,c,f]

[a] *DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, Michigan*

[b] *Department of Geography, Environment and Spatial Sciences, Michigan State University, East Lansing, Michigan*

[c] *Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, Michigan*

[d] *Department of Geographical and Sustainability Sciences, University of Iowa, Iowa City, Iowa*

[e] *Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona*

[f] *W. K. Kellogg Biological Station, Michigan State University, Hickory Corners, Michigan*

Corresponding author: Jie Dai, daijie2@msu.edu

ABSTRACT

Altered surface albedo due to land cover conversions and management is a significant driver of global climate change. Albedo can be directly measured at ground stations, and remote sensing data can be used to scale-up albedo values to regional and global levels. Some previous studies have retrieved fine resolution (10-30 m) instantaneous albedo and coarse resolution (500-1000 m) daily mean albedo from remote sensing data, but they all required the input of MODIS albedo information at 500 m resolution, and none have assembled both instantaneous and daily albedo based exclusively on fine resolution satellite data. To address this issue, we compiled 387 instantaneous and 346 daily albedo records using field net radiometer measurements from the bioenergy croplands at the W. K. Kellogg Biological Station in southwest Michigan. We then connected these albedo records with a suite of variables derived from Harmonized Landsat and Sentinel-2 data through two machine learning algorithms (random forest regression and extreme gradient boosting) to retrieve clear-sky instantaneous and daily shortwave broadband albedo. The performance statistics indicate reasonable accuracy of model results (RMSE around or below 0.03 except for snow-covered surfaces), suggesting that the retrieval of both instantaneous and daily albedo based exclusively on fine resolution satellite data is promising. To facilitate the use of fine resolution albedo products at the global level, future efforts need to include more albedo records of diverse surface cover types, as well as to accurately model daily albedo for cloudy days to address the "clear-sky bias".

KEYWORDS: albedo, bioenergy crop, Harmonized Landsat and Sentinel-2, remote sensing, direct estimation, random forest, XGBoost

1. Introduction

Albedo-induced radiative forcing ($RF_{\Delta\alpha}$) is a significant component of global climate change, which has been recognized in the Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) since 2001 (IPCC 2001). A growing number of studies on $RF_{\Delta\alpha}$ in terrestrial ecosystems have demonstrated the global significance of land use-induced changes in surface albedo (Carrer et al. 2018; McGregor et al. 2024; Wu et al. 2024a; Zhu et al. 2024) and that the cooling effect of albedo change has been largely overlooked (Chen et al. 2024). In southwest Michigan, the CO_2-equivalent climate mitigation provided by agriculture-dominated landscapes is 52% stronger than that provided by the virgin forest-dominated landscapes of the pre-European era, due to elevated land surface albedo (Sciusco et al. 2020). In the same area, converting from brome grassland to bioenergy crops of corn, switchgrass or restored prairie has resulted in higher annual albedo and cooling of local climate (Abraha et al. 2021; Lei et al. 2023, 2024). Accurate quantification of surface albedo is a prerequisite for estimating local to global $RF_{\Delta\alpha}$ and the consequent global warming impact ($GWI_{\Delta\alpha}$) and is thus crucial to developing earth system models and future IPCC assessments.

In situ surface albedo usually is calculated from measurements by net radiometers, and remote sensing data has been used to extrapolate albedo values across space at local, regional and global scales (Schaaf et al. 2002; Ouyang et al. 2022; Wu et al. 2024b). Processed from the data collected by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor onboard the NASA Terra and Aqua satellites, the MODIS Bidirectional Reflectance Distribution Function (BRDF) and Albedo Parameters is one of the most used albedo products for regional and global purposes (Schaaf et al. 2002). This dataset provides daily coverage of global directional-hemispherical (black-sky) and bi-hemispherical (white-sky) albedos at 500 m spatial resolution, as well as corresponding BRDF parameters to calculate albedo at any given solar and viewing angle (Schaaf et al. 2002; Dahlin et al. 2020). However, depending on the scale of the investigation, albedo products with finer spatial resolutions might be more desirable, such as for understanding management impacts of forest patches and agricultural fields for precision farming, as well as for looking at urban infrastructure (Barnes and Roy 2008). To address this need, finer resolution (10-30 m) albedo products have been developed by integrating Landsat or Sentinel-2 spectral reflectance with MODIS BRDF and albedo information (Shuai et al. 2011; Lin et al. 2022).

The *in situ* albedo calculated from ground measurements, blue-sky albedo, is a combination of black-sky and white-sky albedos, whereas the observation from satellites represents the instantaneous bidirectional reflectance of the surface target (Liang 2000; Nagol et al. 2015). Multi-angle observations and surface anisotropy information are needed to estimate the former from the latter (Lucht et al. 2000). Two major categories of approaches have been developed to retrieve surface albedo from Landsat or Sentinel-2 data: “MODIS-concurrent” and “direct estimation” (Liang 2003; Shuai et al. 2011; Lin et al. 2022). The “MODIS-concurrent” approach links the target Landsat/Sentinel-2 pixel to a nearby homogenous MODIS pixel. It assumes that the target Landsat/Sentinel-2 and the corresponding MODIS pixels have the same type of landscape or homogeneous land surface within their focal areas, so that the anisotropy information needed to calculate the blue-sky albedo can be shared. The “direct estimation” approach builds up a pre-defined BRDF library and directly links remote sensing data with *in situ* albedo measurements. Both methods have been applied in several case studies with comparable accuracies (RMSE around 0.03; Shuai et al. 2014; He et al. 2018; Erb et al. 2022).

The albedo values retrieved from remote sensing data are usually instantaneous, corresponding to conditions within the data acquisition time window. From the perspective of surface energy budgets, temporal mean albedo (daily, monthly, or annual) is more useful than instantaneous albedo, although the two can be closely related (Wang et al. 2015). Instantaneous albedo for a specific surface area varies during daytime, mostly due to the constantly changing solar azimuth and zenith angles (Chen et al. 2024). Under clear-sky conditions, this diurnal variation typically exhibits a U-shaped pattern, with lowest albedo values observed at solar noon and highest values at sunrise or sunset. Furthermore, this U-shaped pattern is usually asymmetric, primarily due to varying atmospheric conditions (Han et al. 2024). Because of the diurnal variations of surface albedo, using the local noon albedo value as a surrogate for daily mean albedo leads to an underestimation of daily albedo and overestimation of daily shortwave net radiation (Grant et al. 2000; Wang et al. 2015; Chen et al. 2024).

In recent decades, several attempts have been made to retrieve fine resolution (10-30 m) instantaneous albedo from remote sensing data, but these endeavors employed the input of MODIS albedo information at 500 m resolution (Liang 2000; Shuai et al. 2011; Li et al. 2018). In addition, a few studies have estimated daily mean albedo at coarse spatial resolution (500-1000 m) from MODIS or Visible Infrared Imaging Radiometer Suite (VIIRS) data (Liang et al. 2005; Wang et al. 2015, 2017). Nevertheless, the retrieval of both instantaneous and daily mean albedo based exclusively on fine resolution satellite data is lacking. To this end, we leveraged remote sensing data collected by the Landsat/Sentinel-2 sensors and direct field measurements to explore the potential of fine resolution instantaneous and daily shortwave broadband albedo estimation. Specifically, we compiled albedo records from field measurements at seven bioenergy croplands in southwest Michigan between 2014 and 2024. We then incorporated a suite of training variables from remote sensing as well as ancillary data to connect them with albedo records through two machine learning methods. Our objective is to test if clear-sky surface instantaneous and daily mean albedo can be estimated exclusively from fine spatial resolution (30 m) satellite imagery.

2. Methods

a. Study sites

This study was conducted at seven Great Lakes Bioenergy Research Center (GLBRC) experimental scale-up fields during 2014-2024 at the Lux Arbor Reserve and Marshall Farm of the W. K. Kellogg Biological Station (KBS) in southwest Michigan (Fig. 1; Table 1; (Zenone et al. 2011; Zou et al. 2025). KBS has a humid continental climate with a mean annual air temperature of 9.2 °C and average annual precipitation of 926 mm (Hsieh et al. 2024; Robertson et al. 2024). The three agricultural fields (AGR) at Lux Arbor were managed in a tilled corn-soybean-wheat rotation for decades before 2009, after which they were planted with no-till biofuel crops of continuous corn (*Zea mays*; AGR_C), switchgrass (*Panicum virgatum*; AGR_SW) and restored mixed prairie species (AGR_PR). The four fields at Marshall Farm were enrolled in the U.S. Department of Agriculture (USDA) Conservation Reserve Program (CRP) and planted to smooth brome grass (*Bromus inermis* Leyss) in 1987; three were converted to no-till soybean in 2009 (Abraha et al. 2021). In 2010 the three soybean fields were planted to the same biofuel crops as those at Lux Arbor (i.e., CRP_C, CRP_SW, and CRP_PR), while the fourth field remained as smooth brome grass to serve as a reference site (CRP_REF). Corn at both sites was planted annually in May and harvested in October or November, while perennial crops of switchgrass and mixed prairie were fertilized in the spring after the winter thaw and harvested annually in November or December (Abraha et al. 2021). The reference field (CRP_REF) has received no agronomic management since its establishment.

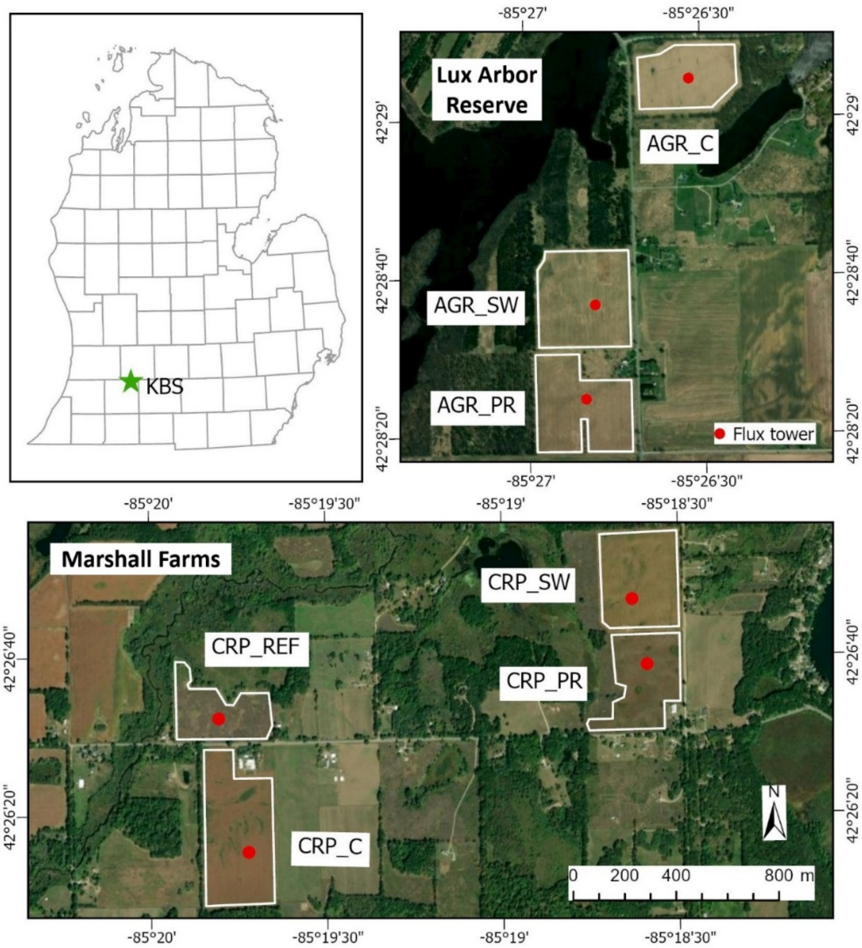

Fig. 1. Locations of the seven GLBRC scale-up experimental fields at the Kellogg Biological Station (KBS) Long Term Ecological Research site in southwest Michigan. Examined fields are delineated and annotated by white polygons. Flux tower locations (red dots) are also labeled. Basemap source: ESRI, Maxar, Earthstar Geographics, and the GIS User Community.

ID	Field Name	Vegetation	Latitude	Longitude
1	CRP_SW	Switchgrass	42.4464	-85.3105
2	CRP_PR	Prairie	42.4441	-85.3098
3	CRP_C	Corn	42.4376	-85.3287
4	AGR_C	Corn	42.4847	-85.4422
5	AGR_SW	Switchgrass	42.4768	-85.4468
6	AGR_PR	Prairie	42.4735	-85.4473
7	CRP_REF	Reference	42.4423	-85.3301

Table 1. Name and vegetation type of the seven GLBRC scale-up fields, as well as the latitudes and longitudes of the flux towers.

b. Data

1) Ground Measurements

At each field we deployed a four-component net radiometer (CNR1, Kipp & Zonen, Delft, The Netherlands) to measure surface downward (incident) and upward (reflected) shortwave radiation (Abraha et al. 2021). Each net radiometer was mounted on the south side of an eddy covariance flux tower at ~1.5 m above the canopy. The sensor mounting arm was adjusted horizontally and vertically until the sensor head was level. The expected accuracy for measured daily total radiation was ±10%. Both longwave and shortwave radiation data was recorded every half hour with a Campbell CR5000 datalogger (CR5000, Campbell Scientific Inc. Logan, UT, USA). Radiation records between one hour after sunrise and one hour before sunset were used for further analyses (Abraha et al. 2021). Surface albedo was first computed as the ratio of the half-hour reflected to incident shortwave radiation. Daily mean albedos were calculated as the ratio of the total reflected to total incident shortwave radiation within the day (Wang et al. 2015).

At Marshall Farm the radiometers were further extended into the fields from October 2013 to September 2018. For all other time periods at Marshall Farm and for the whole time at Lux Arbor, the radiometers' downward facing field of view was close to the flux tower. This was not an issue for fields with homogenous land cover, but it was problematic for the corn fields, where agronomic activities avoided the immediate vicinity of the tower. To address this issue, all radiation records from the field AGR_C and records after September 2018 from the field CRP_C were not included in our analyses.

2) Harmonized Landsat and Sentinel-2 Data

We used the Harmonized Landsat and Sentinel-2 (HLS) data (version 2.0) to model the albedo of the study fields. The HLS project distributes consistent surface reflectance data by integrating Landsat 8/9 and Sentinel-2 products (Claverie et al. 2018). These combined measurements enable improved frequency of global data acquisition (2-3 days) at 30 m spatial resolution. Depending on the availability of the ground measurements, we compiled HLS data between 2014 and 2024 and used the image quality information to select low aerosol cloud-free pixels (Fmask = 64 or 80). The seasonal range of the identified images spanned all months from Jan 18 to Dec 19. We then selected twelve images to construct the annual snow-free surface reflectance dynamics of the study sites. One image was selected for each month between March and December, and one additional image was selected for May and June (each) when the seven fields sprouted and closed their canopies at different rates. For each image, we calculated the mean spectra of pixels within each field as the field-level surface reflectance.

Based on the phenological phases of the study fields, we divided the surface conditions into five periods: green-up, peak, senescence, harvested and snow (except for CRP_REF with no harvested phase). In HLS data, snow-covered pixels were much brighter than the others and were identified through visual interpretation. For snow-free pixels, with the help of *in situ* agronomic activity logs, we used Normalized Difference Vegetation Index (NDVI) to distinguish between green-up, peak and senescence/harvested conditions (Dai et al. 2024; Jamalinia et al. 2024). A NDVI value of 0.7 or higher indicates domination by green vegetation, corresponding to peak conditions. For images collected between April and June, pixels with NDVI values of 0.5-0.7 were classified as green-up. All remaining pixels were classified as senescence or harvested based on the Normalized Harvest Phenology Index (NHPI; Liu et al. 2025)), with NHPI values of close to 0 for senescence and close to 1 for harvested.

c. Instantaneous and daily albedo

Under clear-sky conditions, the instantaneous albedo curve shows somewhat symmetrical U-shaped patterns of diurnal variability, with higher albedos in the morning or late afternoon, and lower albedo values around noon (Chen et al. 2024). We examined whether any morning half-hourly albedo could be used as a surrogate for the daily mean value and compared the ground measured half-hourly albedo with the daily mean albedo to identify which half-hourly values were closest to the daily mean values. Since the local acquisition times of Landsat 8/9 and Sentinel-2 imagery fall between 10:20 h and 10:40 h, we examined half-hourly albedos within four time slots: 9:30-10:00 h, 10:00-10:30 h, 10:30-11:00 h, and 11:00-11:30 h. We applied a linear regression between daily mean and the half-hourly albedos with four statistical metrics for evaluations: the root mean squared error (RMSE), mean bias error (MBE), mean absolute error (MAE) and coefficient of determination (R^2). The detailed algorithms for calculating these metrics are provided in Appendix B. A good fit will produce MBE, MAE and RMSE values close to 0, with a R^2 value close to 1. Finally, we connected the corresponding half-hourly albedo with the HLS data acquisition as the instantaneous albedo.

d. Estimating albedo from HLS data

We identified target HLS pixels for the flux tower locations and assumed homogenous cover types and conditions within each study field (except for AGR_C records after September 2018 and all CRP_C records, which were not included in further analyses). We averaged the information of the target and its eight neighboring pixels to reduce potential adjacency effects and sub-pixel heterogeneity. We then adopted two machine learning methods to estimate the instantaneous and daily mean albedo from HLS data: random forest regression (RFR; Breiman 2001) and extreme gradient boosting (XGBoost; Chen and Guestrin 2016). We first conducted model selection to identify the optimal algorithm-input variable combination, upon which we applied the official model training and testing. We also implemented additional model evaluations to test the robustness of the modeling results.

1) MODELING METHODS

Random forest is an ensemble machine learning algorithm that can be adopted for both classification and regression purposes. It builds multiple independent decision trees and merges their results to improve accuracy and reduce overfitting. Due to its robustness, ability to handle nonlinear relationships and suitability for high-dimensional data, RFR has been widely used to predict continuous variables from remote sensing data (Pal 2005; Belgiu and Drăguţ 2016; Hutengs and Vohland 2016; Obata et al. 2021). Like random forest, XGBoost is a tree-based method that can be used for both classification and regression tasks. But instead of building trees independently, XGBoost implements a gradient boosting framework and builds trees sequentially, with each tree learning from the residuals of the previous ones. Due to its high accuracy, speed, flexibility and ability to handle complex nonlinear relationships, a growing number of studies have applied XGBoost to build predictive models in remote sensing applications (Ghatkar et al. 2019; Wei et al. 2019; Huber et al. 2022).

2) INPUT VARIABLES

Three sets of variables were included as potential model inputs (Table 2). First, the HLS data included surface reflectance values of six spectral bands, namely Blue (B; 490 nm), Green (G; 560 nm), Red (R; 665 nm), Near-Infrared (NIR; 865 nm), and two Shortwave Infrared bands (SWIR1 and SWIR2; 1610 nm and 2200 nm). In addition, four angles that describe the sun-viewing geometry information of the image pixel were also selected, including the Solar Azimuth Angle (SAA), Solar Zenith Angle (SZA), Viewing Azimuth Angle (VAA) and

Viewing Zenith Angle (VZA). This first set of variables was denoted as “Basic” inputs. Second, a suite of spectral indices, highlighted by the U.S. Geological Survey (USGS) website (https://www.usgs.gov/landsat-missions/landsat-surface-reflectance-derived-spectral-indices; Table A1), were calculated from the HLS surface reflectance data and denoted as “Index” inputs. These indices can be used to quantify surface biophysical conditions, such as vegetation greenness, water content and snow coverage. They include the Normalized Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil Adjusted Vegetation Index (SAVI), Modified Soil Adjusted Vegetation Index (MSAVI), Normalized Difference Moisture Index (NDMI), Normalized Burn Ratio (NBR), Normalized Burn Ratio 2 (NBR2), and Normalized Difference Snow Index (NDSI). The third set of variables included the satellite image acquisition date (day of year, DOY), and the ID (1-7) of the seven scale-up fields (i.e., reflecting potential differences among the fields). This set was denoted as “Ancillary” inputs.

Input variable	**Description**
Basic	B, G, R, NIR, SWIR1, SWIR2, SAA, SZA, VAA, VZA
Index	NDVI, EVI, SAVI, MSAVI, NDMI, NBR, NBR2, NDSI
Ancillary	Field ID, Day of year
Pooled	All of the above

Table 2. Three sets of input variables were incorporated for estimating albedo. Basic variables included the spectral reflectance (B, G, R, NIR, SWIR1 and SWIR2; central wavelength at 490 nm, 560 nm, 665 nm, 865 nm, 1610 nm and 2200 nm) and angular information (SAA, SZA, VAA, VZA) directly retrieved from the HLS data. Index variables included the spectral indices calculated from the HLS data (Table A1). Ancillary variables included the field ID and HLS data acquisition day of the year. Pooled variables were all three sets of variables combined.

We linked ground measured albedo with the input variables described above. Based on the availability of ground measurements and HLS data, a total of 387 data records were identified for instantaneous albedo, with each record representing one albedo value and its corresponding input variables. Furthermore, we investigated the daily albedo curves and identified 346 clear-sky daily mean albedo records. Note that when a Landsat imagery acquisition date coincided with a Sentinel-2 imagery acquisition date, one albedo value might be linked with two different sets of Basic and Index input variables, and they were treated as two separate records.

3) MODEL SELECTION

We compiled the 387 instantaneous and 346 daily mean albedo records and conducted model selection to identify the optimal algorithm and input variable combination. RFR was implemented through the Python *scikit-learn* package, whereas XGBoost was applied using the *xgboost* package in Python. We tested four different scenarios of input variables: Basic, Basic + Index, Basic + Ancillary, and Pooled (Table 2). For each algorithm and input scenario, we conducted hyperparameter optimization with a five-fold cross validation method (Li et al. 2023). To address the cover type imbalance in training data, we assigned corn/snow and reference records with higher sample weights compared to switchgrass/prairie (4:2:1, roughly inverse the ratio of their relative sample sizes) during model training. While the differences among different algorithms and input variable combinations were not prominent, the optimal model was identified as RFR-Basic+Indices for instantaneous albedo, and RFR-Basic for daily mean albedo (Table A2). The technical details of model selection are provided in Appendix C.

4) MODEL TRAINING, TESTING AND EVALUATION

Model training, validation and testing were conducted for both instantaneous and daily mean albedo datasets. For each dataset, we divided all records into calibration and testing data, where training and validation were based upon the calibration data, and resultant optimal model and hyperparameters were applied to the testing data. Specifically, we first randomly selected 25% of the records as testing data and assigned the remaining 75% as the calibration data. Next, we applied five-fold cross validation to the calibration data, where the data were split into five groups, with four of them assigned to training and the remaining one to validation. This training and validation process was repeated five times, with each group assigned as validation dataset once. Then the optimal model calibrated from the cross validation was applied to the set-aside testing data. The above calibration-testing split and training-validation-testing procedures were repeated 50 times. To assess the predicting power of the trained models on testing data, we applied a linear regression between ground measured albedo and model predicted values for evaluation by four metrics: RMSE, MBE, MAE and R^2. These performance statistics were averaged from the 50 iterations and calculated for all testing data, as well as grouped by crop types, and phenological periods. We also conducted residual diagnostics (Breusch-Pagan Test) against predicted albedo values with NDVI, SZA and DOY to check for potential heteroscedasticity and geometry dependence.

To test the robustness of the model, we propagated the radiometer measurement uncertainty (10% for daily totals) to prediction uncertainty through Monte Carlo simulations. For both incident and reflected radiation, we randomly generated radiation values using normal distribution with the mean as the measured value and standard deviation as 10% of the mean value. We then calculated the simulated instantaneous and daily albedos and used them for model training and testing. The simulation was operated 50 times and we calculated the average and 95% confidence interval of the performance statistics. Apart from the Monte Carlo simulations, additional model evaluation was incorporated in both spatial and temporal domains through leave-one-site-out and leave-one-year-out cross validation. In leave-one-site-out cross validation, the data from each site was assigned as validation dataset once, so that performance statistics were averaged from six iterations (all seven sites but AGR_C). In leave-one-year-out cross validation, the data from each year was assigned as validation dataset once, and the performance statistics were averaged from 11 iterations (2014-2024).

3. Results & Discussion

a. Field surface reflectance and albedo

Twelve HLS images were selected and used to construct the annual snow-free surface reflectance dynamics of the study fields, and the spectral information captured by the HLS data reflected the phenological dynamics (Fig. 2). In winter months, no green vegetation signals could be identified in the spectra of any field. Based on the ratio of the two shortwave infrared bands, these fields were dominated by non-photosynthetic vegetation (e.g., crop residue or senesced grass), with limited areas of exposed soil (Dai et al. 2018). The perennial fields with switchgrass, restored prairie, and reference brome grass sprouted in late April and closed their canopies in May at varying rates, and by early June they were dominated by green vegetation (Abraha et al. 2021). In contrast, corn was planted annually in mid or late May, and the corn fields remained mostly exposed until mid or late June (Lei et al. 2024). During this time the spectra of the two corn fields did not present significant signals of green vegetation. All seven fields were dominated by green vegetation throughout July and August. The reference field was the first to senesce in September, followed by the other fields. Corn was harvested in October with some stover residues (up to 12 inches) on the ground, whereas the perennial crops were harvested in November with stem stalks mown to approximately 10 cm high (Abraha et al. 2021). With the same type of surface material (non-photosynthetic vegetation), the spectral reflectance and albedo values of the restored prairie and switchgrass fields increased after harvesting due to changes in plant physiology (i.e., senesced vegetation to crop residue).

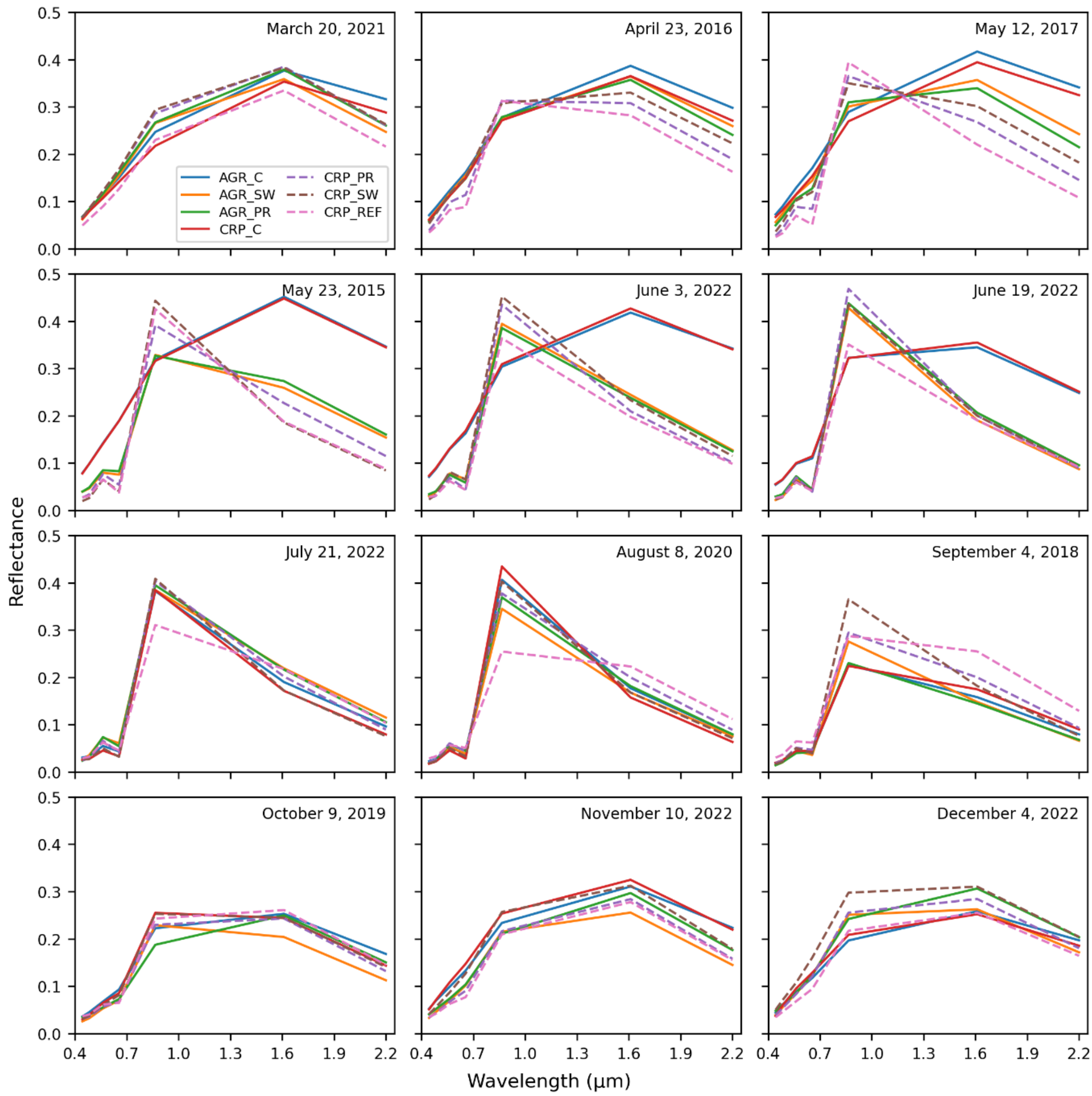

Fig. 2. Changes in field-level mean surface reflectance with wavelength for the seven fields calculated from twelve HLS images. Each panel is plotted from left-right, top-bottom by month of the year starting in March and ending in December to demonstrate the differences in snow-free surface reflectance. The acquisition date (month-day-year) of each image is labeled at the top right of the subplot.

Among the 387 instantaneous albedo records, 366 of the albedo values were unique, including 26 snow-covered and 340 snow-free records. Within the 340 snow-free records, there were 111 switchgrass, 137 prairie, 29 corn and 63 reference. Within the 346 daily mean albedo records, 327 of the albedo values were unique, which including 23 snow-covered and 304 snow-free records. Within the 304 snow-free records, there were 97 switchgrass, 127 prairie, 25 corn and 55 reference. In terms of surface phenological conditions, the dataset included roughly the same amount of peak, senescence and harvested records but fewer green-up and snow ones (Table 3).

Instantaneous albedo					
Type	Green-up	Peak	Senescence	Harvested	Total
Corn	0	6	3	20	29
Switchgrass	7	47	16	41	111
Prairie	7	48	37	45	137
Reference	2	16	45	0	63
Total	16	117	101	106	340
Daily mean albedo					
Type	Green-up	Peak	Senescence	Harvested	Total
Corn	0	5	1	19	25
Switchgrass	7	37	15	38	97
Prairie	7	43	34	43	127
Reference	2	14	39	0	55
Total	16	99	89	100	304

Table 3. The number of unique snow-free albedo values in each surface phenological condition for every vegetation type.

The instantaneous albedo values ranged from 0.1223 for CRP_REF on Aug 20, 2023, to 0.9037 for snow-covered CRP_SW on Feb 21, 2021, whereas the daily mean values ranged from 0.1272 for CRP_REF on Aug 20, 2023, to 0.8466 for snow-covered CRP_C on Dec 19, 2016 (Fig. 3). As expected, the albedo of snow-covered field was much higher than that of the other cover types.

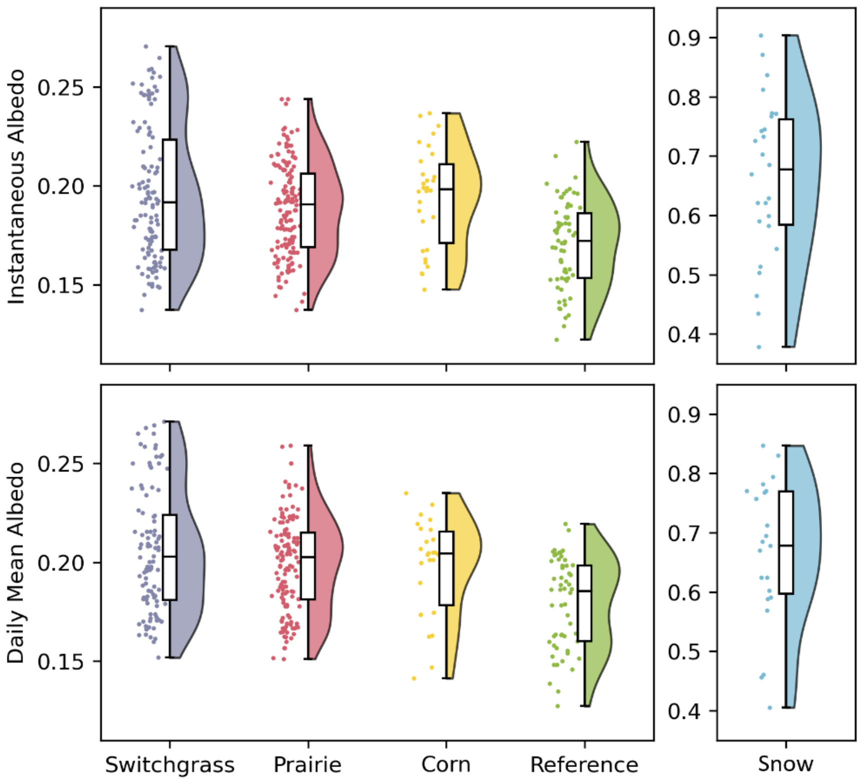

Fig. 3. Distribution of instantaneous and daily mean albedo values at four bioenergy crop sites. For each violin plot, the center line represents the median, the edges show the first quartile (Q1) and the third quartile (Q3) of the data, and the whiskers extend to 1.5 times the interquartile range. Two panels on the left column correspond to snow-free records, whereas the two panels on the right column represent the distributions during snow-covered periods.

Comparing the daily mean albedo with half-hourly values, the former is generally lower than the latter in the early morning and higher at noon. For green-up and peak records, the daily mean albedo was closest to the half-hourly albedo between 9:30 h and 10:00 h (RMSE = 0.0097 and 0.0092; Table 4). For senescence and harvested periods, the daily mean albedo was closest to the half-hourly albedo between 10:00 h and 10:30 h (RMSE = 0.0092 and 0.0119). For fully or partially snow-covered records, the daily mean albedo was closest to the half-hourly albedo between 11:00 h and 11:30 h (RMSE = 0.0265). So, no single half-hourly albedo was consistently closest to the daily mean value across all phenological periods. In addition to RMSE, we also listed the other metrics for reference.

Phenology	Time	RMSE	MBE	MAE	R^2
Green-up (n = 16)	9:30 ~ 10:00	0.0097	-0.0059	0.0069	0.8141
	10:00 ~ 10:30	0.0132	-0.0120	0.0120	0.9125

	10:30 ~ 11:00	0.0178	-0.0169	0.0169	0.9023
	11:00 ~ 11:30	0.0222	-0.0211	0.0211	0.8577
Peak (n = 99)	9:30 ~ 10:00	0.0092	0.0000	0.0073	0.8792
	10:00 ~ 10:30	0.0115	-0.0083	0.0100	0.9061
	10:30 ~ 11:00	0.0170	-0.0152	0.0156	0.9113
	11:00 ~ 11:30	0.0220	-0.0206	0.0207	0.9087
Senescence (n = 89)	9:30 ~ 10:00	0.0101	0.0025	0.0078	0.7486
	10:00 ~ 10:30	0.0092	-0.0057	0.0072	0.8565
	10:30 ~ 11:00	0.0128	-0.0114	0.0115	0.9137
	11:00 ~ 11:30	0.0164	-0.0154	0.0154	0.9214
Harvested (n = 100)	9:30 ~ 10:00	0.0168	0.0047	0.0114	0.6895
	10:00 ~ 10:30	0.0119	-0.0056	0.0089	0.8153
	10:30 ~ 11:00	0.0136	-0.0104	0.0112	0.8688
	11:00 ~ 11:30	0.0158	-0.0136	0.0137	0.8882
Snow (n = 23)	9:30 ~ 10:00	0.0543	-0.0045	0.0426	0.8382
	10:00 ~ 10:30	0.0562	-0.0172	0.0421	0.8583
	10:30 ~ 11:00	0.0312	-0.0015	0.0251	0.9352
	11:00 ~ 11:30	0.0265	0.0028	0.0220	0.9510

Table 4. Comparisons between ground measured morning half-hourly and daily mean albedo for different surface phenological conditions. Four statistic metrics were listed, including the root mean squared error (RMSE), mean bias error (MBE), mean absolute error (MAE) and coefficient of determination (R^2). The formula for calculating the metrics can be found in Appendix B. Similarities were determined by the RMSE values, and the other metrics were also listed for reference.

b. Estimating albedo from HLS data

We report on the performance statistics of using models developed from the calibration data to predict the testing data. For both instantaneous and daily albedo, the trained models performed well with reasonable precision and accuracy (Table 5). We calculated the performance statistics for all testing records, as well as for different cover types and surface phenological conditions. In most cases, the bias fell well within the recommended accuracy requirement of 0.02-0.05 (Sellers et al. 1995; Franch et al. 2014). The only exception was for snow-covered records, in which case the RMSE values were 0.0811 and 0.0769 for instantaneous and daily albedo, respectively. We applied the Breusch-Pagan Test for the residuals against predicted albedo, NDVI, SZA and DOY values and the p-values were >0.1. We also examined the residuals against the above variables, and all appeared randomly scattered around zero. No significant evidence of heteroscedasticity was detected.

Instantaneous albedo				
	RMSE	MBE	MAE	R^2
Overall	0.0318 (±0.0062)	-0.0001 (±0.0028)	0.0209 (±0.0076)	0.9466 (±0.0213)
Corn	0.0225 (±0.0055)	0.0016 (±0.0043)	0.0195 (±0.0058)	0.5636 (±0.0734)
Switchgrass	0.0264 (±0.0079)	-0.0010 (±0.0027)	0.0213 (±0.0051)	0.7043 (±0.0477)
Prairie	0.0144 (±0.0046)	0.0004 (±0.0039)	0.0120(±0.0033)	0.8291(±0.0526)
Reference	0.0166 (±0.0062)	0.0029 (±0.0051)	0.0130(±0.0034)	0.8588(±0.0832)
Snow	0.0811 (±0.0177)	-0.0062 (±0.0094)	0.0760(±0.0138)	0.8207(±0.0341)
Green-up	0.0149 (±0.0043)	0.0017 (±0.0021)	0.0067(±0.0027)	0.7134(±0.0584)
Peak	0.0199 (±0.0075)	-0.0007 (±0.0028)	0.0221(±0.0034)	0.7364(±0.0315)
Senescence	0.0128 (±0.0031)	0.0011 (±0.0017)	0.0127(±0.0029)	0.6711(±0.0429)
Harvested	0.0188 (±0.0049)	0.0008 (±0.0016)	0.0153(±0.0042)	0.6393(±0.0324)
Daily mean albedo				
	RMSE	MBE	MAE	R^2
Overall	0.0243 (±0.0074)	0.0006 (±0.0022)	0.0166(±0.0042)	0.9615(±0.0128)

Corn	0.0113 (±0.0067)	-0.0030 (±0.0014)	0.0079(±0.0056)	0.5946(±0.0413)
Switchgrass	0.0242 (±0.0049)	0.0026 (±0.0039)	0.0092(±0.0031)	0.6746(±0.0624)
Prairie	0.0064 (±0.0037)	-0.0009 (±0.0026)	0.0050(±0.0029)	0.8751(±0.0205)
Reference	0.0163 (±0.0052)	0.0098 (±0.0031)	0.0119(±0.0044)	0.9130(±0.0245)
Snow	0.0769 (±0.0168)	-0.0169 (±0.0045)	0.0585(±0.0157)	0.8875(±0.0439)
Green-up	0.0084 (±0.0055)	0.0049 (±0.0019)	0.0099(±0.0037)	0.6873(±0.0676)
Peak	0.0128 (±0.0028)	0.0014 (±0.0037)	0.0097(±0.0046)	0.6167(±0.0785)
Senescence	0.0126 (±0.0031)	0.0009 (±0.0034)	0.0053(±0.0031)	0.7096(±0.0533)
Harvested	0.0159 (±0.0045)	0.0006 (±0.0025)	0.0128(±0.0027)	0.7651(±0.0426)

Table 5. The performance statistics (RMSE, MBE, MAE and R^2) of the trained RFR-Pooled model predicting the testing data. Here we present the mean and 95% confidence interval of the 50 Monte Carlo simulations. Statistics were calculated for all records as well as for different cover types and surface phenological conditions. Only the bias of "Snow" records exceeded the recommended accuracy requirement of 0.02-0.05.

In a separate test where snow-covered records were removed from the model training and testing, the RMSE was further reduced to 0.0183 for instantaneous albedo and 0.0166 for daily mean albedo, respectively. The higher RMSE of snow records can be potentially attributed to the inadequate training data for snow albedo estimation, especially partially snow-covered vegetation (He et al. 2018). During winter months, the study area was usually cloudy with few clear days. Consequently, limited HLS data from this time of the year could be used. In addition, for partially snow-covered fields, the image-tower representation may not be ideal due to uneven snow conditions within the tower and image focal areas. Further efforts are thus needed to improve the albedo estimation over partial snow-covered surfaces by incorporating more ground measurements and corresponding HLS data.

Additional model evaluation was conducted through leave-one-site-out and leave-one-year-out cross validations. Here we report the average test statistics with 95% confidence interval (Table 6). All testing RMSE values fall between 0.02 and 0.04. For both instantaneous and daily mean albedo, the RMSE in the leave-one-site-out cross validation was higher than that in leave-one-year-out cross validation, indicating that the spatial variations of the dataset were more prominent than the temporal ones. The other metrics are also listed for reference.

	Leave Out	RMSE	MAE	R^2
Instantaneous Albedo	Site	0.0356 (±0.0088)	0.0243 (±0.0061)	0.9106 (±0.0298)
	Year	0.0305 (±0.0091)	0.0186 (±0.0039)	0.9066 (±0.0754)
Daily Albedo	Site	0.0308 (±0.0102)	0.0229 (±0.0094)	0.9401 (±0.0168)
	Year	0.0245 (±0.0095)	0.0167 (±0.0048)	0.9146 (±0.0967)

Table 6. The average performance statistics (RMSE, MAE and R^2) of leave-one-site-out and leave-one-year-out cross validations. 95% confidence intervals were labeled in parentheses.

c. "MODIS-concurrent" vs. "direct estimation" approaches

Compared to the "MODIS-concurrent" approach (Shuai et al. 2011), the "direct estimation" approach utilizes the spectral and angular information to estimate albedo without the input of concurrent coarse-resolution surface anisotropy or ancillary data (Liang 2000; He et al. 2018). Previous "direct estimation" methods require a pre-defined BRDF database, usually derived from MODIS data and physically based radiative transfer modeling, to generate regression coefficients and calculate surface albedo (Liang et al. 1999; He et al. 2018). Here we applied a completely empirical approach and incorporated multi-year (2014-2024) observation in model calibration. In addition, the HLS data contains both Landsat 8/9 and Sentinel-2 records, thus the harmonized and remodeled surface reflectance data integrates multi-angle observations from different sensors (Claverie et al. 2018). In this way, surface anisotropy information can be incorporated into the model development, further improving the estimation accuracy.

d. Albedo and cloud conditions

On a clear (cloud-free) day, the ground measured solar irradiance and half-hourly albedo values show somewhat symmetrical U-shaped patterns of diurnal variability, with lower radiation and higher albedos in the morning or late afternoon, and higher radiation and lower albedo values around noon (Fig. 4). However, varying cloudiness conditions may result in different diurnal albedo changes and daily mean albedo values. If clouds are sparse and the sky is mostly sunny, half-hourly and daily mean albedo values will be almost identical to those on clear days, whereas on an overcast day, albedo values will be significantly lower than on clear days. On average, cloudy-sky daily mean albedo is lower than that on clear days for land surfaces. Accurately estimating cloudy daily mean albedo from HLS data remains challenging and requires further investigation.

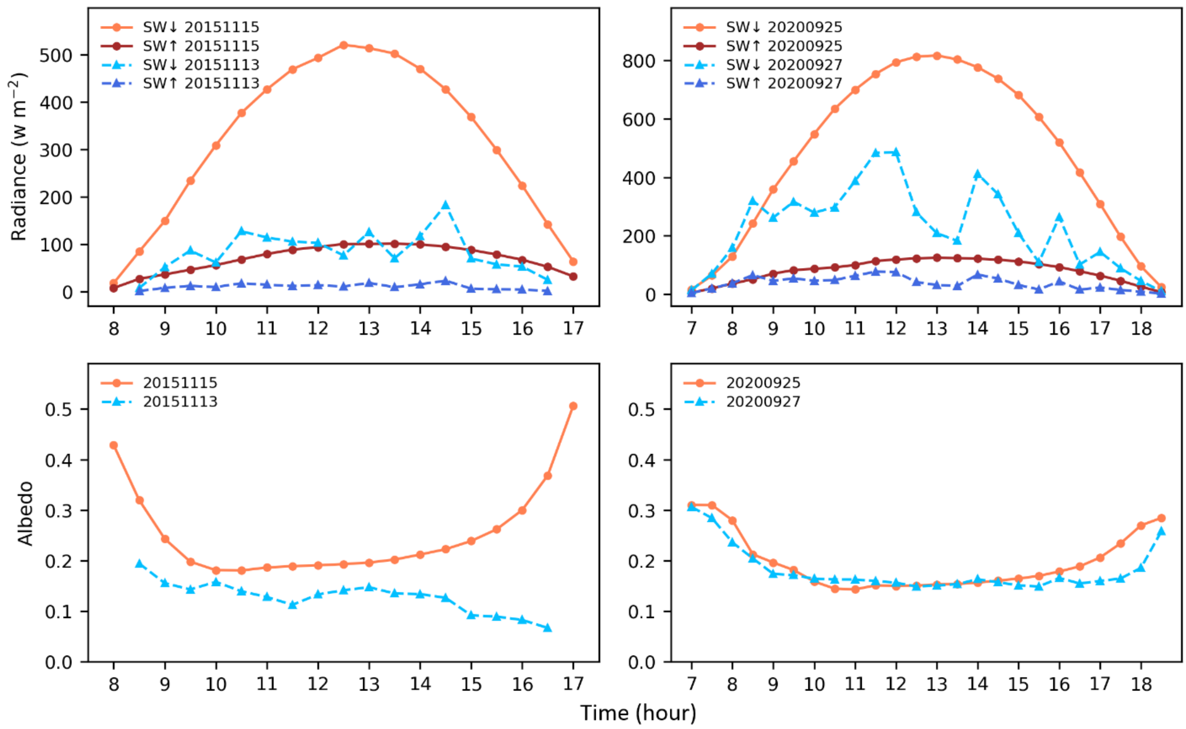

Fig. 4. Shortwave solar radiation and half-hourly albedo on clear and cloudy days. The comparison between clear (20200925) and partially cloudy (20200927) conditions are shown in the left panel, whereas the right panel shows the comparison between clear (20151115) and overcast (20151113) conditions. On clear and partially cloudy days, irradiation (SW↓) and reflection (SW↑) present close to symmetrical patterns.

e. Remaining challenges and future directions

It should be noted that the model developed and tested in this study is location dependent, suitable for estimating albedo of the study fields, and can be applied to land surface with similar cover types. However, to facilitate albedo estimation at the global scale, a general model developed upon comprehensive investigation of diverse locations and land cover/condition types will be required (Dai et al. 2023). In addition, training data representing various BRDF conditions will also be needed to account for the anisotropy information of land surfaces. Traditional Landsat and Sentinel-2 sensors were designed for observations with similar revisit local time and solar-viewing geometry. The integrated and remodeled HLS data can provide multi-angular observations which can be valuable for high-resolution land surface albedo retrieval.

Although our model estimated clear-sky daily mean albedo with reasonable accuracy (RMSE less than 0.03), the approach is empirical to retrieve daily value from an instantaneous satellite observation between 10:20 h and 10:40 h. A hybrid method integrating physical models with empirical estimations may potentially increase the transferability of model results. With the increased acquisition frequency of the HLS data, near same-day multi-angle observation may be applicable. Future endeavors may integrate ground measurements with the RossThick-LiSparse-Reciprocal model to produce BRDF parameters for the HLS data so that daily albedo can be calculated as the integral of instantaneous albedo in the temporal domain.

Furthermore, when utilizing remote sensing data to scale-up flux tower observations, it is vital to address the potential "clear-sky bias" issue (Zhang et al. 2020). Even with the surface reflectance data from the HLS project and potential 2-3 days temporal resolution (Claverie et al. 2018), large gaps may arise due to cloudy conditions between neighboring remote sensing days, especially for regions with high cloud cover. Besides, albedo modeled from satellite instantaneous observations correspond to clear-sky conditions, which may not be accurate on cloudy days. Atmospheric conditions, including cloud cover and optical thickness, are needed to model cloudy-sky daily mean albedo from clear-sky daily albedo. These issues need to be addressed to facilitate high resolution albedo products at the global scale (Radeloff et al. 2024).

4. Conclusions

Leveraging remote sensing data collected by Landsat and Sentinel-2 sensors (HLS data) and field measurements of albedo, we estimated clear-sky fine resolution (30 m) instantaneous and daily shortwave broadband albedo for seven bioenergy crop fields at the W. K. Kellogg Biological Station in southwest Michigan. We first compared the ground measured half-hourly albedo with the daily mean values. Depending on surface phenological period, the daily mean albedo appears similar to the half-hourly values between 9:30 h and 10:00 h (green-up and peak), 10:00 h and 10:30 h (senescence and harvested), and 11:00 h and 11:30 h (snow). Next, we linked the ground measured albedo values to the HLS data, with each record including an albedo value and a suite of input variables. A total of 387 instantaneous albedo records were identified, and the albedo values ranged from 0.1223 to 0.9037. Similarly, 346 daily mean albedo records were identified, and the values range from 0.1272 to 0.8466. We incorporated two machine learning algorithms (RFR and XGBoost) and identified the optimal model-input variable combination through hyperparameter optimization. We then developed the estimation model from the training data and used it to predict the testing data. The model performed well for most crop types and phenological periods (RMSE around 0.03), except for snow-covered period. It should be noted that our study only provided a local proof-of-concept case for the temperate croplands at KBS. To facilitate the use of fine resolution albedo products at the regional and global level, future efforts need to include surfaces with diverse land cover types, as well as more albedo records of snow-covered surfaces. Additionally, accurately model albedo for cloudy days is vital for addressing the "clear-sky bias" issue.

Acknowledgments

This material is based upon work at the Great Lakes Bioenergy Research Center supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award Number DE-SC0018409. Additional support was provided by the NASA LCLUC Program (80NSSC20K0410) and the National Science Foundation Long-term Ecological Research Program (DEB 2224712) at the Kellogg Biological Station, and by the Michigan State University AgBioResearch. The authors would like to thank Kristine Blakeslee for editing the manuscript, as well as the three anonymous reviewers and the handling editors for their constructive comments for revising the manuscript.

Data Availability Statement.

The HLS data is freely available in Google Earth Engine. The net radiometer data that support the findings of this study can be requested at https://data.sustainability.glbrc.org/. The code used in this study can be found at Zenodo (doi.org/10.5281/zenodo.18332628).

APPENDICES

Appendix A. The spectral indices highlighted by USGS

Index	**Algorithm**
Normalized Difference Vegetation Index (NDVI)	(NIR-R)/(NIR+R)
Enhanced Vegetation Index (EVI)	2.5((NIR-R)/(NIR+6R-7.5B+1))
Soil Adjusted Vegetation Index (SAVI)	1.5((NIR-R)/(NIR+R+0.5))
Modified Soil Adjusted Vegetation Index (MSAVI)	(2NIR+1-sqrt((2NIR+1)^2-8(NIR-R)))/2

Normalized Difference Moisture Index (NDMI)	(NIR-SWIR1)/(NIR+SWIR1)
Normalized Burn Ratio (NBR)	(NIR-SWIR2)/(NIR+SWIR2)
Normalized Burn Ratio 2 (NBR2)	(SWIR1-SWIR2)/(SWIR1+SWIR2)
Normalized Difference Snow Index (NDSI)	(G-SWIR1)/(G+SWIR1)

Table A1. Spectral indices incorporated as training variables. B, G, R, NIR, SWIR1 and SWIR2 refer to the surface reflectance value of relevant bands in HLS data.

Appendix B. Model performance statistics

We used four performance statistics, including the mean bias error (MBE), mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2), to examine the fitness between model predicted and ground measured albedo. They were calculated using the following equations:

$$\boldsymbol{MBE} = \frac{\sum_{i=1}^{n}(\boldsymbol{P_i} - \boldsymbol{O_i})}{\boldsymbol{n}}$$

$$\boldsymbol{MAE} = \frac{\sum_{i=1}^{n}|\boldsymbol{P_i} - \boldsymbol{O_i}|}{\boldsymbol{n}}$$

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\boldsymbol{P_i} - \boldsymbol{O_i})^2}{\boldsymbol{n}}}$$

$$\boldsymbol{R^2} = \boldsymbol{1} - \frac{\sum_{i=1}^{n}(\boldsymbol{P_i} - \boldsymbol{O_i})^2}{\sum_{i=1}^{n}(\boldsymbol{P_i} - \overline{\boldsymbol{p}})^2}$$

where $\boldsymbol{P_i}$ is the predicted value, $\boldsymbol{O_i}$ is the observed value, $\overline{\boldsymbol{p}}$ is the mean of predicted values, and n is the number of observations. While the first three metrics depict the deviation of the modeled values to the measured values, R^2 is the proportion of the variation in the measured values that is predictable from the modeled values. A good fit will have MBE, MAE and RMSE values close to 0 and R^2 close to 1.

Appendix C. Model selection procedures and results

The model selection and hyperparameter optimization was conducted through five-fold cross validation, implemented with the RandomizedSearchCV function (n_iter = 100) in Python.

In random forest regression, the hyperparameters involved and their search space were: 1) n_estimators: 100, 200, 500, 1000; 2) max_depth: None, 10, 20; 3) min_samples_split: 2, 5, 10; 4) min_samples_leaf: 1, 2, 4; and 5) max_features: auto, log2, sqrt.

In XGBoost, the hyperparameters involved and their search space were: 1) n_estimators: 100, 500, 1000; 2) max_depth: None, 5, 10, 20; 3) learning_rate: 0.01, 0.1, 0.2, 0.3; 4) subsample: 0.8, 1.0; 5) colsample_bytree: 0.6, 0.8, 1.0; 6) min_child_weight: 1, 3; 7) reg_alpha: 0, 0.1, 0.5; and 8) reg_lambda: 1, 1.5, 2.

For each algorithm-input variable combination, we report the performance statistic (RMSE; Table A2). While the differences among combinations were not prominent, the optimal model was identified as RFR-Basic+Indices for instantaneous albedo, and RFR-Basic+Ancillary for daily mean albedo.

Input variable	**Instantaneous**		**Daily**	
	RFR	**XGBoost**	**RFR**	**XGBoost**
Basic	0.03508	0.03964	0.02872	0.02975
Basic + Indices	0.03500	0.03946	0.02967	0.03003
Basic + Ancillary	0.03509	0.03631	0.02925	0.02965
Pooled	0.03505	0.03761	0.02944	0.02983

Table A2. The performance statistic (RMSE) for each algorithm-input variable combination. The optimal models (i.e., RFR-Basic+Indices for instantaneous albedo and RFR-Basic for daily albedo) were identified based on the smallest RMSE.

REFERENCES

Abraha, M., J. Chen, S. K. Hamilton, P. Sciusco, C. Lei, G. Shirkey, J. Yuan, and G. P. Robertson, 2021: Albedo-induced global warming impact of Conservation Reserve Program grasslands converted to annual and perennial bioenergy crops. *Environ. Res. Lett.*, **16**, 084059, https://doi.org/10.1088/1748-9326/ac1815.

Barnes, C. A., and D. P. Roy, 2008: Radiative forcing over the conterminous United States due to contemporary land cover land use albedo change. *Geophysical Research Letters*, **35**, 2008GL033567, https://doi.org/10.1029/2008GL033567.

Belgiu, M., and L. Drăguţ, 2016: Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, **114**, 24–31, https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Breiman, L., 2001: Random Forests. *Machine Learning*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Carrer, D., S. Moparthy, G. Lellouch, X. Ceamanos, F. Pinault, S. C. Freitas, and I. F. Trigo, 2018: Land Surface Albedo Derived on a Ten Daily Basis from Meteosat Second Generation Observations: The NRT and Climate Data Record Collections from the EUMETSAT LSA SAF. *Remote Sensing*, **10**, 1262, https://doi.org/10.3390/rs10081262.

Chen, J., C. Lei, H. Chu, X. Li, M. Torn, Y.-P. Wang, P. Sciusco, and G. P. Robertson, 2024: Overlooked cooling effects of albedo in terrestrial ecosystems. *Environ. Res. Lett.*, **19**, 093001, https://doi.org/10.1088/1748-9326/ad661d.

Chen, T., and C. Guestrin, 2016: XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *KDD '16*, New York, NY, USA, Association for Computing Machinery, 785–794, https://doi.org/10.1145/2939672.2939785.

Claverie, M., J. Ju, J. G. Masek, J. L. Dungan, E. F. Vermote, J.-C. Roger, S. V. Skakun, and C. Justice, 2018: The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, **219**, 145–161, https://doi.org/10.1016/j.rse.2018.09.002.

Dahlin, K. M., D. Akanga, D. L. Lombardozzi, D. E. Reed, G. Shirkey, C. Lei, M. Abraha, and J. Chen, 2020: Challenging a Global Land Surface Model in a Local Socio-Environmental System. *Land*, **9**, 398, https://doi.org/10.3390/land9100398.

Dai, J., D. Roberts, P. Dennison, and D. Stow, 2018: Spectral-radiometric differentiation of non-photosynthetic vegetation and soil within Landsat and Sentinel 2 wavebands. *Remote Sensing Letters*, **9**, 733–742, https://doi.org/10.1080/2150704X.2018.1470697.

Dai, J., E. Jamalinia, N. R. Vaughn, R. E. Martin, M. König, K. L. Hondula, J. Calhoun, J. Heckler, and G. P. Asner, 2023: A general methodology for the quantification of crop canopy nitrogen across diverse species using airborne imaging spectroscopy. *Remote Sensing of Environment*, **298**, 113836, https://doi.org/10.1016/j.rse.2023.113836.

Dai., J., M. König, E. Jamalinia, K. L. Hondula, N. R. Vaughn, J. Heckler, and G. P. Asner, 2024: Canopy-Level Spectral Variation and Classification of Diverse Crop Species with Fine Spatial Resolution Imaging Spectroscopy. *Remote Sensing*, **16**, 1447.

Erb, A. M., Z. Li, Q. Sun, I. Paynter, Z. Wang, and C. Schaaf, 2022: Evaluation of the Landsat-8 Albedo Product across the Circumpolar Domain. *Remote Sensing*, **14**, 5320, https://doi.org/10.3390/rs14215320.

Franch, B., E. F. Vermote, and M. Claverie, 2014: Intercomparison of Landsat albedo retrieval techniques and evaluation against in situ measurements across the US SURFRAD network. *Remote Sensing of Environment*, **152**, 627–637, https://doi.org/10.1016/j.rse.2014.07.019.

Ghatkar, J. G., Singh ,Rakesh Kumar, and P. and Shanmugam, 2019: Classification of algal bloom species from remote sensing data using an extreme gradient boosted decision tree model. *International Journal of Remote Sensing*, **40**, 9412–9438, https://doi.org/10.1080/01431161.2019.1633696.

Grant, I. F., A. J. Prata, and R. P. Cechet, 2000: The Impact of the Diurnal Variation of Albedo on the Remote Sensing of the Daily Mean Albedo of Grassland. *Journal of Applied Meteorology*.

Han, Y., and Coauthors, 2024: Asymmetry in the Diurnal Variation of Land Surface Albedo and Its Impacts on Daily Mean Albedo Estimation. *Journal of Geophysical Research: Atmospheres*, **129**, e2023JD039728, https://doi.org/10.1029/2023JD039728.

He, T., S. Liang, D. Wang, Y. Cao, F. Gao, Y. Yu, and M. Feng, 2018: Evaluating land surface albedo estimation from Landsat MSS, TM, ETM +, and OLI data based on the unified direct estimation approach. *Remote Sensing of Environment*, **204**, 181–196, https://doi.org/10.1016/j.rse.2017.10.031.

Hsieh, H.-Y., S. Bohm, and G. P. Robertson, 2024: Kellogg biological station climate trends.

Huber, F., A. Yushchenko, B. Stratmann, and V. Steinhage, 2022: Extreme Gradient Boosting for yield estimation compared with Deep Learning approaches. *Computers and Electronics in Agriculture*, **202**, 107346, https://doi.org/10.1016/j.compag.2022.107346.

Hutengs, C., and M. Vohland, 2016: Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment*, **178**, 127–141, https://doi.org/10.1016/j.rse.2016.03.006.

IPCC, 2001: Climate Change 2001: The Scientific Basis | Climatology and climate change. *Cambridge University Press*. Accessed December 5, 2024, https://www.cambridge.org/sk/academic/subjects/earth-and-environmental-science/climatology-and-climate-change/climate-change-2001-scientific-basis-contribution-working-group-i-third-assessment-report-intergovernmental-panel-climate-change, https://www.cambridge.org/sk/academic/subjects/earth-and-environmental-science/climatology-and-climate-change.

Jamalinia, E., J. Dai, N. Vaughn, R. E. Martin, K. Hondula, M. Koenig, J. Heckler, and G. P. Asner, 2024: Crop canopy nitrogen estimation from mixed pixels in agricultural lands using imaging spectroscopy. *Remote Sensing* **16**, 1382.

Lei, C., J. Chen, and G. P. Robertson, 2023: Climate cooling benefits of cellulosic bioenergy crops from elevated albedo. *GCB Bioenergy*, **15**, 1373–1386, https://doi.org/10.1111/gcbb.13098.

Lei, C., J. Chen, I. Ibáñez, P. Sciusco, G. Shirkey, M. Lei, P. Reich, and G. P. Robertson, 2024: Albedo of crops as a nature-based climate solution to global warming. *Environ. Res. Lett.*, **19**, 084032, https://doi.org/10.1088/1748-9326/ad5fa2.

Li, X., H. Jia, and L. Wang, 2023: Remote Sensing Monitoring of Drought in Southwest China Using Random Forest and eXtreme Gradient Boosting Methods. *Remote Sensing*, **15**, 4840, https://doi.org/10.3390/rs15194840.

Li, Z., A. Erb, Q. Sun, Y. Liu, Y. Shuai, Z. Wang, P. Boucher, and C. Schaaf, 2018: Preliminary assessment of 20-m surface albedo retrievals from sentinel-2A surface reflectance and MODIS/VIIRS surface anisotropy measures. *Remote Sensing of Environment*, **217**, 352–365, https://doi.org/10.1016/j.rse.2018.08.025.

Liang, S., 2000: Narrowband to broadband conversions of land surface albedo I Algorithms. *Remote Sensing of Environment*.

Liang, S., 2003: A direct algorithm for estimating land surface broadband albedos from MODIS imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 136–145, https://doi.org/10.1109/TGRS.2002.807751.

Liang, S., A. H. Strahler, and C. Walthall, 1999: Retrieval of Land Surface Albedo from Satellite Observations: A Simulation Study. *Journal of Applied Meteorology and Climatology*.

Liang, S., J. Stroeve, and J. E. Box, 2005: Mapping daily snow/ice shortwave broadband albedo from Moderate Resolution Imaging Spectroradiometer (MODIS): The improved direct retrieval algorithm and validation with Greenland in situ measurement. *Journal of Geophysical Research: Atmospheres*, **110**, https://doi.org/10.1029/2004JD005493.

Lin, X., and Coauthors, 2022: Estimating 10-m land surface albedo from Sentinel-2 satellite observations using a direct estimation approach with Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, **194**, 1–20, https://doi.org/10.1016/j.isprsjprs.2022.09.016.

Liu, Y., C. Diao, Z. Yang, W. Mei, and T. Guo, 2025: A novel Normalized Harvest Phenology Index (NHPI) for corn and soybean harvesting date detection using Landsat and Sentinel-2 imagery on Google Earth Engine. *Remote Sensing of Environment*, **331**, 115016, https://doi.org/10.1016/j.rse.2025.115016.

Lucht, W., C. B. Schaaf, and A. H. Strahler, 2000: An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 977–998, https://doi.org/10.1109/36.841980.

McGregor, S., J. P. G. M. Cromsigt, M. Te Beest, J. Chen, D. P. Roy, H.-J. Hawkins, and G. I. H. Kerley, 2024: Grassland albedo as a nature-based climate prospect: the role of growth form and grazing. *Environ. Res. Lett.*, **19**, 124004, https://doi.org/10.1088/1748-9326/ad8765.

Nagol, J. R., J. O. Sexton, D.-H. Kim, A. Anand, D. Morton, E. Vermote, and J. R. Townshend, 2015: Bidirectional effects in Landsat reflectance estimates: Is there a problem to solve? *ISPRS Journal of Photogrammetry and Remote Sensing*, **103**, 129–135, https://doi.org/10.1016/j.isprsjprs.2014.09.006.

Obata, S., C. J. Cieszewski, R. C. Lowe, and P. Bettinger, 2021: Random Forest Regression Model for Estimation of the Growing Stock Volumes in Georgia, USA, Using

Dense Landsat Time Series and FIA Dataset. *Remote Sensing*, **13**, 218, https://doi.org/10.3390/rs13020218.

Ouyang, Z., and Coauthors, 2022: Albedo changes caused by future urbanization contribute to global warming. *Nat Commun*, **13**, 3800, https://doi.org/10.1038/s41467-022-31558-z.

Pal, M., 2005: Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, **26**, 217–222, https://doi.org/10.1080/01431160412331269698.

Radeloff, V. C., and Coauthors, 2024: Need and vision for global medium-resolution Landsat and Sentinel-2 data products. *Remote Sensing of Environment*, **300**, 113918, https://doi.org/10.1016/j.rse.2023.113918.

Robertson, G. P., and Coauthors, 2024: The LTAR Cropland Common Experiment at the Kellogg Biological Station. *J of Env Quality*, **53**, 893–903, https://doi.org/10.1002/jeq2.20638.

Schaaf, C. B., and Coauthors, 2002: First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*, **83**, 135–148, https://doi.org/10.1016/S0034-4257(02)00091-3.

Sciusco, P., and Coauthors, 2020: Spatiotemporal variations of albedo in managed agricultural landscapes: inferences to global warming impacts (GWI). *Landscape Ecol*, **35**, 1385–1402, https://doi.org/10.1007/s10980-020-01022-8.

Sellers, P. J., and Coauthors, 1995: Remote sensing of the land surface for studies of global change: Models — algorithms — experiments. *Remote Sensing of Environment*, **51**, 3–26, https://doi.org/10.1016/0034-4257(94)00061-Q.

Shuai, Y., J. G. Masek, F. Gao, and C. B. Schaaf, 2011: An algorithm for the retrieval of 30-m snow-free albedo from Landsat surface reflectance and MODIS BRDF. *Remote Sensing of Environment*, **115**, 2204–2216, https://doi.org/10.1016/j.rse.2011.04.019.

Shuai, Y., J. G. Masek, F. Gao, and T. He, 2014: An approach for the long-term 30-m land surface snow-free albedo retrieval from historic Landsat surface reflectance and MODIS-based a priori anisotropy knowledge. *Remote Sensing of Environment*, **152**, 467–479, https://doi.org/10.1016/j.rse.2014.07.009.

Wang, D., S. Liang, T. He, Y. Yu, C. Schaaf, and Z. Wang, 2015: Estimating daily mean land surface albedo from MODIS data. *JGR Atmospheres*, **120**, 4825–4841, https://doi.org/10.1002/2015JD023178.

Wang, D., S. Liang, Y. Zhou, T. He, and Y. Yu, 2017: A New Method for Retrieving Daily Land Surface Albedo From VIIRS Data. *IEEE Transactions on Geoscience and Remote Sensing*, **55**, 1765–1775, https://doi.org/10.1109/TGRS.2016.2632624.

Wei, L., Z. Yuan, Y. Zhong, L. Yang, X. Hu, and Y. Zhang, 2019: An Improved Gradient Boosting Regression Tree Estimation Model for Soil Heavy Metal (Arsenic) Pollution Monitoring Using Hyperspectral Remote Sensing. *Applied Sciences*, **9**, 1943, https://doi.org/10.3390/app9091943.

Wu, L., and Coauthors, 2024a: Global warming impacts of carbon dioxide, methane, and albedo in an island forest nature reserve. *Environ. Res. Lett.*, **19**, 114085, https://doi.org/10.1088/1748-9326/ad86d1.

Wu, S., and Coauthors, 2024b: Satellite observations reveal a decreasing albedo trend of global cities over the past 35 years. *Remote Sensing of Environment*, **303**, 114003, https://doi.org/10.1016/j.rse.2024.114003.

Zenone, T., and Coauthors, 2011: CO2 fluxes of transitional bioenergy crops: effect of land conversion during the first year of cultivation: CO2 FLUXES OF TRANSITIONAL BIOENERGY CROPS. *GCB Bioenergy*, **3**, 401–412, https://doi.org/10.1111/j.1757-1707.2011.01098.x.

Zhang, Z., Y. Zhang, Y. Zhang, and J. M. Chen, 2020: Correcting Clear-Sky Bias in Gross Primary Production Modeling From Satellite Solar-Induced Chlorophyll Fluorescence Data. *Journal of Geophysical Research: Biogeosciences*, **125**, e2020JG005822, https://doi.org/10.1029/2020JG005822.

Zhu, Q., J. Chen, L. Wu, Y. Huang, C. Shao, G. Dong, Z. Xu, and X. Li, 2024: Changes in albedo and its radiative forcing of grasslands in East Asia drylands. *Ecol Process*, **13**, 17, https://doi.org/10.1186/s13717-024-00493-w.

Zou, H., and Coauthors, 2025: Contributions of vegetation heterogeneity within tower footprint to CO2 flux estimations through graph neural network modeling. *Remote Sensing of Environment*, **329**, 114952, https://doi.org/10.1016/j.rse.2025.114952.