Title: Development of an Open-Source Web-Based Tool for Alfalfa Yield and Quality Prediction

PI, Co-PIs, Authors:

PI: Zhou Zhang

Address: 460 Henry Mall, Madison, WI 53705

Phone: 608-265-2575

Email: zzhang347@wisc.edu

Co-PI: Matthew Digman

Address: 460 Henry Mall, Madison, WI 53705

Phone: 608-890-1941

Email: digman@wisc.edu

Abstract:

Alfalfa is one of the most cultivated perennial legume crops as feedstock for animals. Efficiently estimating alfalfa yield and quality traits before harvesting is critical for the decision-making process regarding precision management activities and harvesting time to ensure high profitability. Satellite-based radar is a powerful tool for crop monitoring because it provides high-quality data regardless of weather conditions. Therefore, this study investigated the potential of satellite radar features in estimating alfalfa yield and quality. Alfalfa yield and quality traits, including dry matter yield (DMY), crude protein (CP), neutral detergent fiber (NDF), NDF digestibility (NDFD), and acid detergent fiber (ADF), were collected over 16 alfalfa fields from 2016 to 2021, leading to 126 samples in total. Environmental factors and High-resolution radar backscattering coefficients (σ^o_{VV} and σ^o_{VH}) of all the fields across all growing seasons were collected from Daymet V4 product and Sentinel-1 Ground Range Detected products in C-band. Four commonly used machine learning models were established to estimate alfalfa traits using the radar backscattering coefficients with different time features. Results showed that the Extreme Gradient Boosting model consistently performed the best for all alfalfa traits. The DMY estimation highly agreed with the measured values with an average R^2 of 0.64 and RMSE of 0.71 tons/ha. The best result for estimating CP was with an average R^2 of 0.72 and RMSE of 1.58%. The performance in estimating alfalfa fiber indicators (i.e., ADF, NDF, and NDFD) was achieved with the highest average R^2 of 0.54, 0.64, and 0.60, respectively. Machine learning models were integrated into an open-source website built on Flask, HTML and JavaScript, allowing users to display existing yield maps. In addition to this, users can also customize the date and area, and the backend of the website performs online data search and yield prediction and display the results.

1. Introduction:

Alfalfa is a valuable and intensively produced perennial forage crop in the United States, and precise management to achieve forage yield and quality goals is critical to optimize profitability. Pre-harvest insight into yield and quality can enable real-time management responses to fluctuating market conditions or feed needs. However, the acquisition of timely information at the field scale is limited using traditional measurements such as destructive sampling and assessment of plant maturity. Although on-harvester yield mapping systems exist, the state of technology in yield mapping of alfalfa results in yield maps that are generated after harvest. While on-harvester prediction exists for silage making, dry hay production quality is expensive in terms of labor to obtain samples and laboratory fees. Consequently, producers are not able to use this data to make timely management decisions. Therefore, there is a critical need to develop an efficient and cost-effective tool to estimate alfalfa yield and quality at the field-scale.

The advances of remote sensing techniques can provide a non-destructive and efficient way to monitor crop growth, and thus have great potential for crop yield and nutritive quality analysis. In this context, publicly available satellite imagery offers unparalleled opportunities for land surface monitoring and characterization over space and time domains, with far-reaching socio-economic benefits [1]. Satellite remote sensing has been widely used in agriculture, including weed–crop discrimination [2]; monitoring crop growth and health status [3]; and yield mapping and prediction [4]. Although demonstrated by previous studies with promising performance, success in crop monitoring on particular regions could be limited by solely relying on satellite optical data because areas at high latitudes on earth are severely affected due to cloud, fog, and evaporation, resulting in up to 80% of them missing for those areas[5,6]. Instead, radar-based satellites could be a solution to the areas as they function by transmitting and receiving electromagnetic waves in microwave ranges, which are nearly unaffected by clouds due to their long wavelength[7]. Synthetic aperture radar (SAR) satellites, a form of radar satellites, collect data with a finer spatial resolution by utilizing the synthetic aperture antenna for capturing data from the target. Sentinel-1 is a combination of two SAR satellites deployed by the European Space Agency in 2014 and 2015, providing C-band SAR data with a spatial resolution of 5 meters in range and 20 meters azimuth in Interferometric Wide Swath mode [8]. Sentinel-1 provides free and high-resolution SAR backscattering coefficients which have been recently used for a variety of crop monitoring purposes, including aboveground biomass (AGB), height, and leaf area index in large pasture fields [9], quantity and quality of semi-natural grassland forage[10], and biophysical parameters of wheat, soybeans, and canola [11]. This study investigated the potential of satellite radar data, particularly Sentinel-1 SAR data, and environmental factors during the alfalfa regrowth in estimating their DMY and quality traits.

Simple linear regression models have been developed in remote sensing of crop yield and nutritive quality. However, remote sensing data are complex and the relationship between the yield/quality and the imagery can be non-linear. More complex machine learning models are needed. Through training and updating the parameters, the machine learning model gradually approaches the true numerical relationship between input features and the target yield and quality.

Given this background, we developed an open-source, web-based tool that employs the use of satellite remote sensing and machine learning to provide prediction of alfalfa yield and quality for precision management.

2. Materials and Methods:

2.1. Study Area and ground data sampling

The study was conducted from 2013 to 2021 at the Arlington Agricultural Research Station (AARS, latitude of 43° 16’ N to 43° 20’ N and longitude of 89° 19’ W to 89° 24’ W) and Marshfield Agricultural Research Station (MARS, latitude of 44° 43’ N to 44° 48’ N and longitude of 90° 04’ W to 90° 08’ W) at the University of Wisconsin-Madison. From 2016 to 2021, ground data (i.e., alfalfa DMY, CP, ADF, NDF, and NDFD) were collected in seven sites at AARS and nine sites at MARS. The geographic location of the research stations and distribution of the sampling sites are shown in Figure 1. Alfalfa in each site was sampled three to four times in each growing season starting from the first year to three years after seeding. The detailed sampling times for all the sites are listed in Table 1. The first cut (sampling) of individual sites in a year usually happened at the end of May, there was around one-month interval between two cuts. Therefore, the second, third, and fourth cuts happened at the end of June, July, and August, respectively. In total, 126 samples were collected for measuring alfalfa DMY and quality traits, featured by the sites, years, and cutting orders.

The total wet yield (kg) of each experimental site was obtained during the harvest by weighing the empty and full trucks carrying the harvested forage with an on-farm drive-over scale [12]. The total DMY (kg) was then obtained by multiplying the total wet yield by a dry/wet ratio obtained using two dried samples which were randomly selected from each harvest. Then, the DMY (kg/ha) was the total DMY divided by each field size. The two selected samples were also used for quality measurement using NIR analysis. Results from the two forage samples were averaged for each harvest [12].

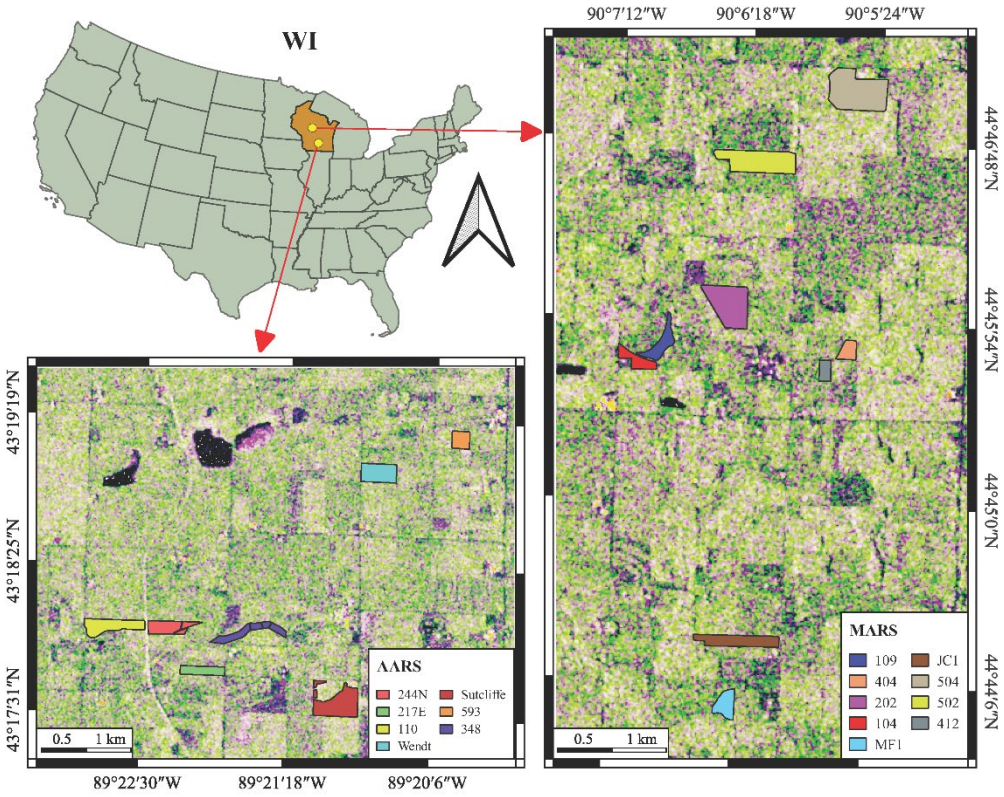

Figure 1. The two study areas and the distribution of their sampling sites. The RGB images were created using backscattering coefficients derived from Sentinel-1 GRDH data on July 24, 2020, where blue: σ^o_{VV}, green: σ^o_{VH}, and Red: $\sigma^o_{VH}/\sigma^o_{VV}$.

Table 1. Sampling sites and timeline

Site ID	Year seeded	Alfalfa variety	Sampling years (Number of samples)
AARS-593	2013	Pioneer 55V50	2016 (4), 2017 (4)
AARS-348	2014	Mixed RR varieties	2016 (4), 2017 (4), 2018 (4)
AARS-217E	2015	Pioneer 55V50/ Dairyland HF3400	2016 (4), 2017 (4), 2018 (4)
AARS-244N	2016	HVXRR 4.0 Brand	2017 (4), 2018 (4), 2019 (4)
AARS-Wendt	2017	Pioneer 55VR08	2018 (4), 2019 (4), 2020 (4)
AARS-110	2018	Jung 4R418	2019 (4), 2020 (4), 2021 (4)
AARS-Sutcliffe	2019	HybriForce 3431	2020 (4), 2021 (4)
MARS-109	2014	Dairyland Magnum 7/wet	2016 (3), 2017 (4)
MARS-JC1	2015	Dairyland Magnum 7/wet	2016 (3), 2017 (3)
MARS-404	2016	Dairyland Hybriforce 3405	2017(3),2018 (4)
MARS-504	2016	Dairyland 2420/wet	2017 (3), 2018 (4)
MARS-MF1	2017	Dairyland 3420/wet	2018 (4), 2019 (3)
MARS-202	2018	Dairyland 3420/wet	2019 (4), 2020 (4)
MARS-502	2018	Dairyland 3420/wet	2019 (4)
MARS-412	2016	Pioneer 55V50	2016 (1)
MARS-104	2019	Dairyland 3420/wet	2020 (3)

Major environmental factors during alfalfa regrowth were collected from Daymet V4 [13], including the daylight duration (period of days during which the sun is above a hypothetical flat horizon), precipitation, incident shortwave radiation flux density or radiation for short (taken as an average over the daylight period of the day), and vapor pressure (VP, daily average partial pressure of water vapor). Growing degree days (GDD) were calculated to assess crop development over the regrowth period, using Eq. 1. We used the cumulative values of each environmental factor over the regrowth before each cut (around one month period) as input variables for further analysis. For the first cut in a year, we calculated the cumulative values from 30 days prior to the cut to the day before the cut, and for the second to fourth cut, we calculated the cumulative values from the day after the previous cut to the day before the current cut.

$$GDD = \frac{T_{a\_max}+T_{a\_min}}{2} - T_{a\_base} \quad (1)$$

where, T_{a_max} and T_{a_max} is the maximum and minimum air temperature in a day. T_{a_base} was set to 5 °C.

2.2. Satellite Image Acquisition and Processing

Sentinel-1 imagery over all the experimental sites from 2016 to 2021 was acquired and processed for estimating alfalfa yield and quality traits. The two polar-orbiting satellites (Sentinel-1 A and B) have provided imagery with a nominal six-day temporal resolution, leading to a total of 53 imaging days during the sampling timeline. The SAR data were collected from the high-resolution Ground Range Detected (GRD) products of Sentinel-1 in interferometric wide swath (IW) mode.

As shown in Figure 2, the GRD images were downloaded and preprocessed using Google Earth Engine (GEE) cloud platform, which have been pre-processed using the Sentinel-1 toolbox provided by the European Space Agency. To derive GRD product, there were five steps: 1) applying orbit file, 2) GRDH border noise removal, 3) thermal noise removal, 4) radiometric calibration computing backscatter coefficients using sensor calibration parameters in the GRD metadata, and 5) terrain correction. Note

that GEE provides SAR data at spatial resolution of 10-meters that was resampled to 30-meters with nearest approach in this study. Backscattering coefficients VV (σ^o_{VV}) and VH (σ^o_{VH}) along with the CR ($\sigma^o_{VV}/\sigma^o_{VH}$), DPSVIm ($(\sigma^o_{VV} * \sigma^o_{VV} + \sigma^o_{VH} * \sigma^o_{VV})/\sqrt{2}$) [14],Pol ($(\sigma^o_{VH} - \sigma^o_{VV})/(\sigma^o_{VH} + \sigma^o_{VV})$) [15], RVIm ($4 * \sigma^o_{VH}/(\sigma^o_{VH} + \sigma^o_{VV})$) [16] and the VH-VV ($\sigma^o_{VH} - \sigma^o_{VV}$) were used as input features to estimate the alfalfa DMY and quality traits. To evaluate performance of radar features collected at different alfalfa growth stages, those from single-time (i.e., the closest day to the ground sampling date) and dual-time (the day before the closest day and the closest day) were separately used to train the machine learning models for estimating each of the alfalfa traits. Each sample was featured with a Gap variable, indicating the day's difference between when the sample was cut and imaged (the satellite acquisition dates). To increase the sample size for machine learning modeling, we resampled the dataset by stacking samples with early features and late features and removed those that were imaged too long ago from the sampling days (i.e., Gap > 12 days), leading to a dataset with 216 samples training and validating the machine learning models.

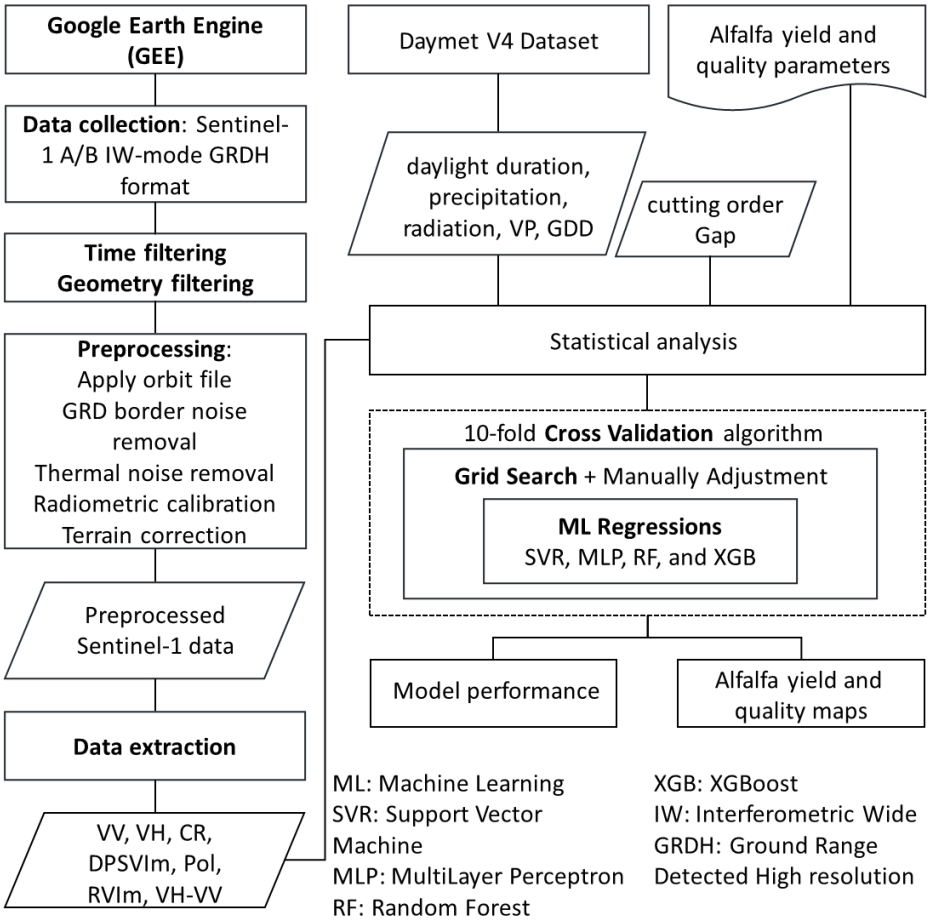

Figure 2. Workflow of objective 1

2.3. Machine Learning Models and Performance Evaluation

To evaluate the performance of machine learning models in estimating alfalfa DMY and quality traits, four models that have been widely used in estimating crop traits with remote sensing data, namely Support Vector Regression (SVR), Multilayer Perceptron (MLP), Random Forest (RF), and Extreme Gradient Boosting (XGB), were developed and validated using the dataset. Optimal hyperparameters in each model were found for each alfalfa trait using a grid search algorithm. The search range and steps for all the hyperparameters are listed in Table. 2 as well as their optimized values. The model performance was evaluated using the ten-fold cross-validation (CV) which was repeated for each alfalfa trait for 20 times. In each of the repetitions, the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were calculated between the estimated and measured values. Averages of the three metrics over the 20 CV repetitions were reported.

Table 2. Grid search specifications and optimized values for hyperparameters of the machine learning models.

Model	Hyperparameter	Searching Range	Optimized values*				
			DMY	ADF	NDF	NDFD	CP
SVR	c [1]	0.01, 0.1, 0.5,1, 5,10, 100	10(5)	100(100)	100(100)	100(10)	5(100)
	g [2]	0.01, 0.1, 0.5,1, 5,10, 100	1(0.01)	0.5(0.01)	1(0.01)	0.1(5)	5(0.01)
MLP	act [4]	'identity', 'logistic', 'tanh', 'relu'	'relu' ('relu')	'tanh' ('logistic')	'tanh' ('relu')	'tanh' ('tanh')	'tanh' ('logistic')
	solver [5]	'adam', 'sgd'	'adam'('adam')	'sgd'('sgd')	'sgd'('sgd')	'sgd'('sgd')	'sgd'('sgd')
	hl [6]	10,15,20,25,30,35,40	30(15)	15(35)	30(40)	30(40)	40(35)
	bs [7]	20,24,32,40	20(20)	20(32)	24(20)	20(40)	32(32)
	rs [8]	0,1,2,3	0(3)	3(2)	0(1)	3(3)	1(1)
RF	m-feature [9]	'sqrt', 'log2', 0.3,0.5,1	0.5(0.5)	0.5(0.5)	0.5(1)	0.5(1)	0.5(0.5)
	n-est [10]	30, 40, 50, 80,100	100(40)	30(40)	40(100)	100(40)	50(80)
	m-sl [11]	2,4,6,8,10	2(6)	2(2)	2(2)	2(2)	2(8)
	rs	0,1,2,3	1(1)	3(2)	0(3)	0(1)	3(3)
XGB	m-depth [12]	3,4,5,6,7,8	5(3)	6(3)	3(3)	7(3)	6(3)
	n-est	30,40,50,80,100	100(40)	100(100)	100(40)	100(50)	100(40)
	lr [13]	0.01,0.05,0.1	0.1(0.1)	0.1(0.05)	0.1(0.1)	0.1(0.1)	0.1(0.1)
	tree_method	'exact','approx','hist'	'approx' ('exact')	'approx' ('exact')	'exact' ('exact')	'approx' ('exact')	'exact' ('exact')
	alpha [14]	0.4,0.5,0.8,1	0.4(0.4)	0.5(1)	0.4(1)	0.8(1)	1 (0.8)
	lambda [15]	0.6,0.8,1,1.2	0.6(0.8)	0.8(0.6)	0.6(0.6)	0.8(0.6)	0.6(1.2)

SVR parameters: [1] squared L2 penalty and [2] gamma in the radial basis function kernel; MLP parameters: [4] activation function, [5] the solver for weight optimization, [6] hidden layers, [7] batch size, and [8] random state; RF parameters: [9] the number of features for the best split, [10] number of trees, and [11] minimum number of samples required to be at a leaf node; XGB parameters: [12] maximum depth, [13] learning rate, [14] L1 regularization term on weights, and [15] L2 regularization.

*Values in parentheses are obtained by using features other than environmental factors as model inputs

2.4. Python-based web tool

To improve data accessibility and promote the utilization of our yield prediction model, we have designed and implemented a user-friendly web-based interface for visualizing the results. we chose to employ PythonAnywhere as our online development environment and hosting service, allowing us to seamlessly deploy and run our application on the web.

For the front-end of our interface, we crafted an interactive web page using a combination of HTML and JavaScript. This front-end component is responsible for presenting the results and enabling user interactions, ensuring a smooth and intuitive experience for our users. On the backend side, we harnessed the power of Flask, a lightweight micro web framework for Python, to build the server and application logic. Flask provides us with the necessary tools for handling HTTP requests, routing, and

data processing. This choice of backend technology not only streamlines the development process but also ensures the scalability and performance of our web application. All the data generated and utilized by our yield prediction model is stored in the file system provided by PythonAnywhere, making it readily available for processing and presentation within the web interface.

By combining these technologies and services, we have created an accessible and user-friendly platform for visualizing the results of our yield prediction model. Users can effortlessly access and interact with the data, ultimately enhancing the utility and adoption of our model.

3. Project Objectives and Corresponding Results:

Objective 1: Develop machine learning models to estimate in-season alfalfa yield and quality using multi-source data and validate the results on Wisconsin Alfalfa Yield and Persistence database.

Result: Four machine learning models were established to estimate alfalfa traits using the radar backscattering coefficients. The DMY estimates given by the MLP model highly agreed with the measured values (R^2 = 0.64 and *RMSE* = 0.71 tons/ha). Among all the traits, the CP was estimated with the highest accuracy with R^2 = 0.72 and *RMSE* = 1.58% by the XGB model. Compared to the DMY and CP, the estimates for three fiber indicators had lower accuracies with the highest average R^2 of 0.54, 0.64, and 0.60, respectively.

Objective 2: Create an open-source web-based tool to provide end-users with in-season alfalfa yield and quality maps to assess their spatial variability for precision management decisions.

Result: The publicly accessible website address is https://alfalfabse.pythonanywhere.com/ . Aside from merely viewing the yield prediction map for the study area, users also have the option to customize the area-of-interest (AOI) and date. The website has the capability to search for data, execute the prediction algorithm, and subsequently display the yield prediction map.

4. Results and Discussion:

4.1. Performance for alfalfa DMY and quality traits

Figure 3 shows scatterplots between the alfalfa measured and estimated values given by the optimal trained models. The DMY estimates given by the XGB models highly agreed with the measured values with an average R^2 of 0.64 and *RMSE* of 0.71 tons/ha over the 20 repeated CV. The model performances benefited from the relationships between either DMY or backscattering coefficients with vegetation water content (VWC). DMY, though measured as the weights of water-free substances in vegetation samples, has a high correlation with the VWC. Zhou et al in [17] and Bhattacharya et al. [18] in showed a strong linear relationship (r = 0.96) between DMY and VWC. Logically, vegetation with more water has a higher capacity to sustain the water, supported by a larger volume of dry matter. On the other hand, radar backscattering coefficients from vegetation are highly related to its VWC because water absorbs more energy than other substances in vegetation [19,20]. Consequently, vegetation with higher VWC tends to reflect fewer radar signals than those with lower VWC.

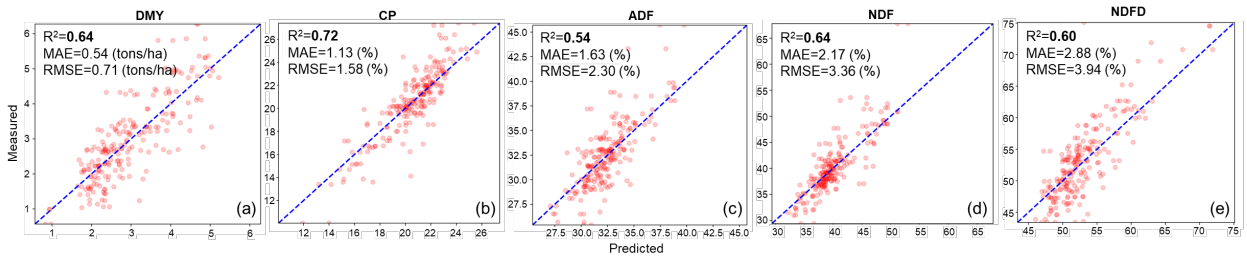

Figure 3. Scatterplots of the estimated and measured alfalfa yield and quality traits, where a - e are for DMY, CP, ADF, NDF, and NDFD, respectively. Estimates of DMY and quality traits were given by the XGB. Estimated values in the plots for individual samples and evaluation metrics were averaged over the 20 times cross validation process.

Our results are competitive compared to previous studies using satellite optical data. [21] reported an R^2 of 0.92, RMSE of 1.1 tons/ha, and MAE of 0.8 tons/ha using the GPR model with the selected time-series Landsat 8 and PROBA-V features. Using single-time satellite SAR data requires less data processing and computational efforts, but in the meanwhile achieving lower average errors. The best performance in estimating alfalfa DMY in published literature was with an R^2 = 0.94 and RMSE = 0.25 t/ha, reported by [22], which, however, requires high-resolution images from UAV as well as satellite imagery from three platforms (i.e., Sentinel-2, Landsat-8, and MODIS). Overall, the performance of this study is promising for practical use in estimating alfalfa DMY.

For estimating CP, the best result with an average R^2 of 0.72 and RMSE of 1.58% was given by the XGB model (Figure 3b). The CP concentration in alfalfa is related to chlorophyll concentration that is closely dependent on its water and nitrogen status [10]. Therefore, it could be reflected by the backscattering coefficients. The CP estimates agreed the most with the measured values in our studies, demonstrating its utility for practical use [10,23,24]. The performance in estimating alfalfa fiber indicators (i.e., ADF, NDF, and NDFD) is shown in Figures 3 c-e, with the highest average R^2 of 0.54, 0.64, and 0.60, respectively. The best estimates for the three fiber indicators were all given by the XGB models. The relationships between backscattering coefficients and plant density account for the models' performance. Plant density primarily affects the structure of the plant population. Increasing plant density causes more competition among individuals for nutrients, water, and light, and consequently lowers the quality by increasing fiber contents in the plant. It was reported that the fiber contents (ADF and NDF) increased with higher plant densities, thus resulting in decreasing quality [25–28]. In addition to VWC, radar backscattering coefficients from vegetation are affected by its geometric factors including plant density, pattern, and height. Dense and high objects increase both times and energies of radar signals being reflected to the satellite receiver, accordingly higher backscattering coefficients [29–35].

4.2. Performance between estimates with and without environmental factors

Table 3 shows the comparison of the performance between two groups of inputs, i.e., with and without the environmental factors, in different machine learning models. The best result for each feature is bolded. When without the environmental factors, the best performance in DMY estimation was obtained by the RF and MLP with an R^2 of 0.47 and RMSE of 0.87 tons/ha. The averaged errors were still competitive with previous studies. However, the performance for the fiber indicators and CP was not

ideal only with the SAR features with the best R^2 all lower than 0.2. The performance was significantly higher when the environmental factors were added, especially for the quality traits.

The lower effects of the satellite data could be because we did not account for the separate impacts of alfalfa status and soil water content on the radar backscatter. Radar signals interact with the vegetation and soil in alfalfa fields at the same time. Given the same vegetation conditions, the radar signals experience less attenuation (weakening) in dry soil and thus lead to higher backscatter values. And the backscatter values at earlier alfalfa growth stages tended to be affected by soil conditions as less ground was covered when alfalfa was short and thin. Neglecting to distinguish the effects of soil water content from alfalfa traits in the radar backscatter data likely contributed to the errors in our estimates. Going forward, we recognize the importance of distinguishing these backscatter effects as a crucial aspect of our future research directions.

Table 3. Performance of the four machine learning models in estimating alfalfa yield and quality traits.

Input	Alfalfa traits	SVR			MLP			RF			XGB		
		R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
With Environmental factors	DMY	0.53	0.82	0.64	0.53	0.81	0.65	0.59	0.77	0.59	**0.64**	**0.71**	**0.54**
	CP	0.23	2.61	1.83	0.24	2.60	1.92	0.59	1.91	1.39	**0.72**	**1.58**	**1.13**
	ADF	0.27	2.90	2.05	0.26	2.91	2.11	0.35	2.73	1.97	**0.54**	**2.30**	**1.63**
	NDF	0.23	4.91	3.30	0.24	4.88	3.51	0.49	3.98	2.70	**0.64**	**3.36**	**2.17**
	NDFD	0.23	5.44	4.12	0.27	5.30	4.07	0.40	4.82	3.57	**0.60**	**3.94**	**2.88**
Without Environmental factors	DMY	0.40	0.92	0.74	**0.47**	**0.87**	0.71	**0.47**	**0.87**	**0.69**	0.45	0.88	0.70
	CP	**0.19**	**2.69**	**2.03**	0.18	2.70	2.05	0.13	2.77	2.07	0.11	2.82	2.11
	ADF	**0.15**	**3.12**	**2.32**	**0.15**	**3.12**	2.34	0.07	3.27	2.48	0.07	3.26	2.41
	NDF	0.01	5.56	**3.92**	**0.04**	**5.47**	4.04	-0.00	5.60	4.11	-0.10	5.85	4.14
	NDFD	0.12	5.82	4.32	0.12	5.81	4.39	**0.18**	**5.62**	**4.31**	0.14	5.75	4.34

4.3. Visualization of the estimated alfalfa DMY and quality traits

The estimated maps of an example field are shown in Figure 5 to visualize in-field variations in the alfalfa DMY, and quality traits estimated by the XGB model with both environmental factors and SAR features. Each column in Figure 5 shows maps before each cut in the season of 2020. In general, the average values of each map corresponded well with their field-level ground data. The estimated maps also demonstrated the spatial variations of each field. Given complicated interactions of weather, in-field microenvironments, and management activities, alfalfa growth status in fields could be affected as a whole or in specific areas. By having the spatial variation maps, further investigations could be conducted and decisions regarding field management practices, such as time, amount, and location for precision irrigation, fertilization, and alfalfa cutting windows could be made on time to maximize production potential, quality, and profitability.

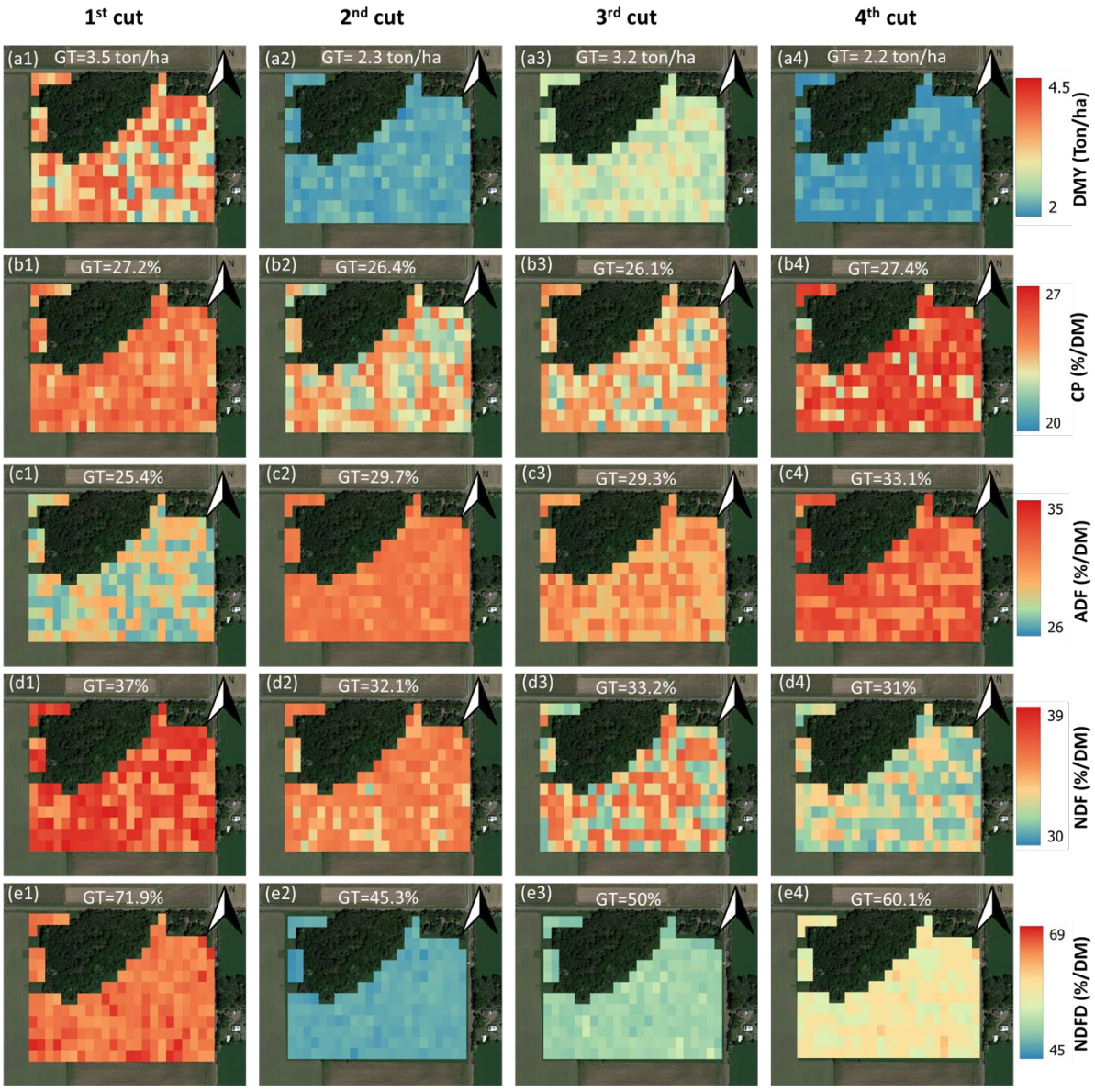

Figure 5. Alfalfa yield and quality maps of field Sutcliffe generated using the XGB model. a, b, c, d, and e are maps for DMY, CP, ADF, NDF, and NDFD, respectively. Each column was from each of the four cuts in 2020 with the leftmost column from the first cut and the rightmost one from the fourth cut. GT: Ground Truth.

4.4. Web tool

The website has two sections, the map section, and the operation section, as shown in Figure 6. The map section uses Google satellite map as its base map. Field name and other information are displayed in separate windows on the base map. Colors in the colormap on top of the map show the values of estimated yield of the study area with red indicating the maximum and green indicating the minimum

values. The information window appears when users click on a field and contains the relevant data of each field.

The operation section is below the map section, providing two rows of operational options. In the first row, there are two location buttons, i.e., Location A and Location B. They can be used to switch the map centers between two areas of interest. For example, Location A is centered in the AARS, and B is centered in the MARS. As we have saved several yield estimation results of the fields in these two areas, the ‘Select All’, ‘Unselect All’ and the year buttons in the Filters options are used to select which year's yield results to display. The year buttons are automatically generated based on the years of the saved results. In the second row, the ‘Help Menu’ button is used to display some explanations about the web tool, including usage instructions, button descriptions, data sources, and model introductions, etc. The ‘Change Data Date’ button is used to set the time range for satellite image search, while the ‘Reset Data Date’ button is used to reset the time range to the previous month from the current time. Once the dates are changed, all generated Yield images will be updated. The ‘Define Area of Interest’ button is used to define the study area, and the ‘Reset Area of Interest’ button is used to clear the defined area. The ‘Toggle Field Yield’ button is used to show and hide the yield results.

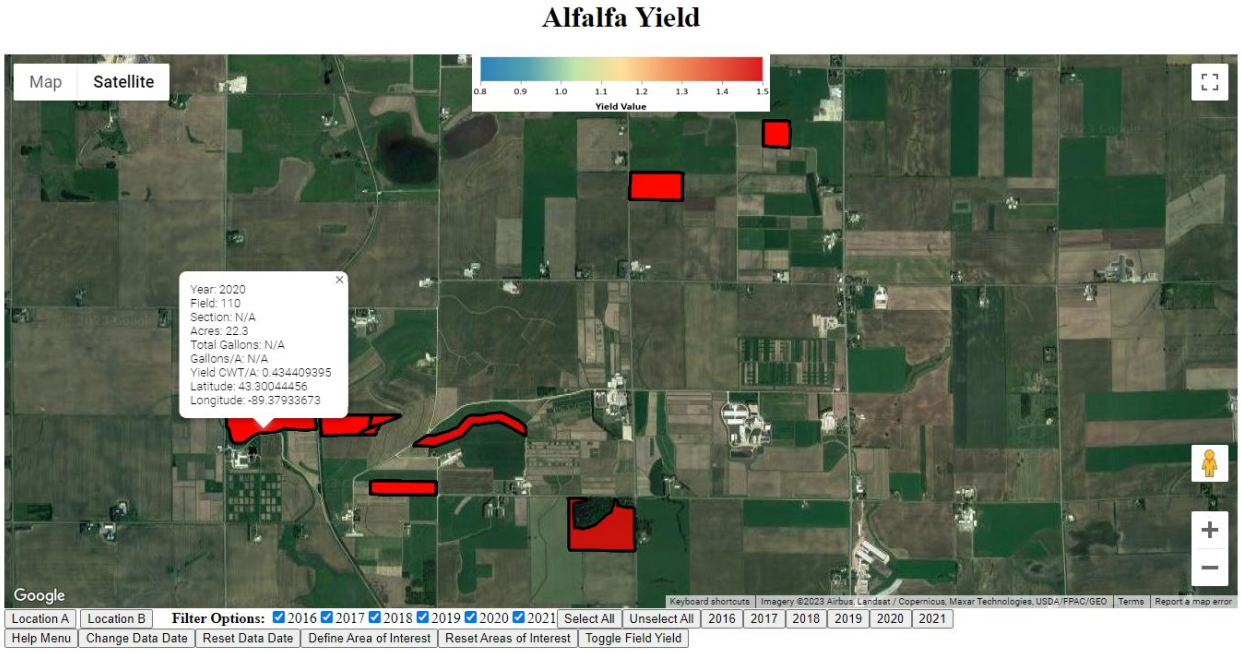

Figure 6. Interface of the website.

To view existing yield estimation results in AARS or MARS, users need to first select a location (Location A or Location B), then check one or more years (2016-2021), and finally click the ‘Toggle Field Yield’ button, as shown in Figure 7.

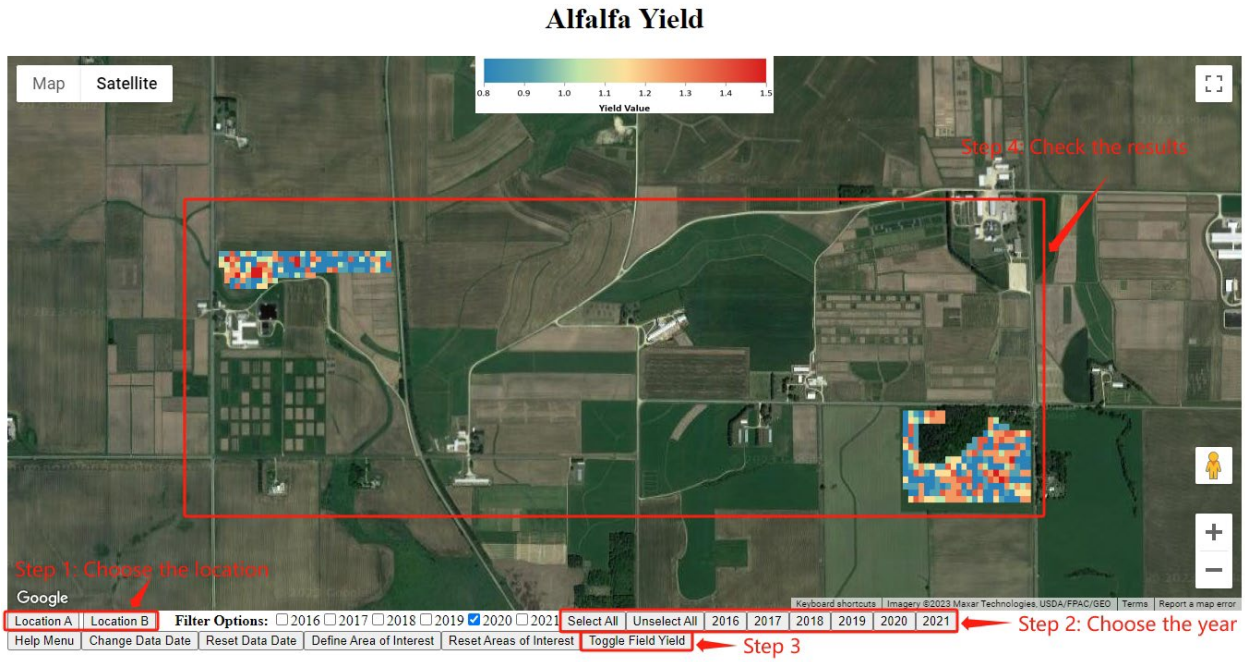

Figure 7. Yield map of fields in location A.

The website tool also allows users to define a new AOI and obtain estimated yield of the new area over a certain time period. To do so, users need to first set the search time using the 'Change Data Date' button, then define the area using the 'Define Area of Interest' button, and then wait for several seconds for the corresponding area to display the yield results. An example of yield estimation map near the Arlington station is shown in Figure 8.

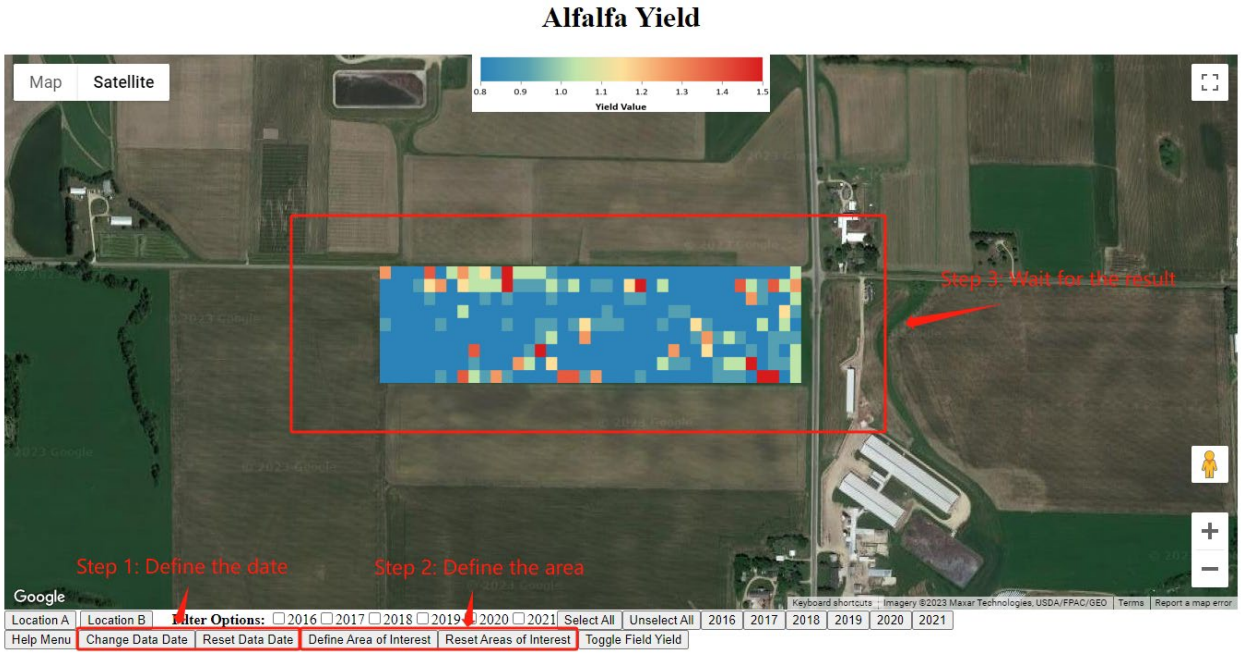

Figure 8. Yield map of a newly defined area.

Acknowledgements: Funding for this study was provided by the U.S. Alfalfa Farmer Research Initiative of the National Alfalfa & Forage Alliance. Alfalfa yield and quality records were provided by Wisconsin Alfalfa Yield and Persistence (WAYP) Program.

References:

1. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; C.e., W.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; et al. Landsat-8: Science and Product Vision for Terrestrial Global Change Research. *Remote Sens. Environ.* **2014**, *145*, 154–172, doi:10.1016/j.rse.2014.02.001.
2. Bisht, P.; Kumar, P.; Yadav, M.; Rawat, J.S.; Sharma, M.P.; Hooda, R.S. Spatial Dynamics for Relative Contribution of Cropping Pattern Analysis on Environment by Integrating Remote Sensing and GIS. *Int. J. Plant Prod.* **2014**.
3. Liang, L.; Qin, Z.; Zhao, S.; Di, L.; Zhang, C.; Deng, M.; Lin, H.; Zhang, L.; Wang, L.; Liu, Z. Estimating Crop Chlorophyll Content with Hyperspectral Vegetation Indices and the Hybrid Inversion Method. *Int. J. Remote Sens.* **2016**, *37*, 2923–2949, doi:10.1080/01431161.2016.1186850.
4. Lobell, D.B. The Use of Satellite Data for Crop Yield Gap Analysis. *Field Crops Res.* **2013**, *143*, 56–64, doi:10.1016/j.fcr.2012.08.008.
5. Sarafanov, M.; Kazakov, E.; Nikitin, N.O.; Kalyuzhnaya, A.V. A Machine Learning Approach for Remote Sensing Data Gap-Filling with Open-Source Implementation: An Example Regarding Land Surface Temperature, Surface Albedo and NDVI. *Remote Sens.* **2020**, *12*, 3865, doi:10.3390/rs12233865.
6. He, M.; Hu, Y.; Chen, N.; Wang, D.; Huang, J.; Stamnes, K. High Cloud Coverage over Melted Areas Dominates the Impact of Clouds on the Albedo Feedback in the Arctic. *Sci. Rep.* **2019**, *9*, 9529, doi:10.1038/s41598-019-44155-w.
7. Ulaby, F.T.; Long, D.G. *Microwave Radar and Radiometric Remote Sensing*; The University of Michigan Press: Ann Arbor, 2014; ISBN 978-0-472-11935-6.
8. Potin, P.; Rosich, B.; Grimont, P.; Miranda, N.; Shurmer, I.; O'Connell, A.; Torres, R.; Krassenburg, M. Sentinel-1 Mission Status. In Proceedings of the Proceedings of EUSAR 2016: 11th European conference on synthetic aperture radar; VDE, 2016; pp. 1–6.
9. Crabbe, R.A.; Lamb, D.W.; Edwards, C.; Andersson, K.; Schneider, D. A Preliminary Investigation of the Potential of Sentinel-1 Radar to Estimate Pasture Biomass in a Grazed Pasture Landscape. *Remote Sens.* **2019**, *11*, 872.
10. Raab, C.; Riesch, F.; Tonn, B.; Barrett, B.; Meißner, M.; Balkenhol, N.; Isselstein, J. Target-Oriented Habitat and Wildlife Management: Estimating Forage Quantity and Quality of Semi-Natural Grasslands with Sentinel-1 and Sentinel-2 Data. *Remote Sens. Ecol. Conserv.* **2020**, *6*, 381–398, doi:10.1002/rse2.149.
11. Bahrami, H.; Homayouni, S.; McNairn, H.; Hosseini, M.; Mahdianpari, M. Regional Crop Characterization Using Multi-Temporal Optical and Synthetic Aperture Radar Earth Observations Data. *Can. J. Remote Sens.* **2022**, *48*, 258–277.
12. Bertram, M.; Cavadini, J. *Wisconsin Alfalfa Yield and Persistence (Wayp) Program 2020 Summary Report*; University of Wisconsin-Madison, 2020;
13. Thornton, M.M.; Shrestha, R.; Wei, Y.; Thornton, P.E.; Kao, S.-C.; Wilson, B.E. Daymet: Daily Surface Weather Data on a 1-Km Grid for North America, Version 4 R1 2022.

14. dos Santos, E.P.; Da Silva, D.D.; do Amaral, C.H. Vegetation Cover Monitoring in Tropical Regions Using SAR-C Dual-Polarization Index: Seasonal and Spatial Influences. *Int. J. Remote Sens.* **2021**, *42*, 7581–7609, doi:10.1080/01431161.2021.1959955.
15. Hird, J.N.; DeLancey, E.R.; McDermid, G.J.; Kariyeva, J. Google Earth Engine, Open-Access Satellite Data, and Machine Learning in Support of Large-Area Probabilistic Wetland Mapping. *Remote Sens.* **2017**, *9*, 1315, doi:10.3390/rs9121315.
16. Nasirzadehdizaji, R.; Balik Sanli, F.; Abdikan, S.; Cakir, Z.; Sekertekin, A.; Ustuner, M. Sensitivity Analysis of Multi-Temporal Sentinel-1 SAR Parameters to Crop Height and Canopy Coverage. *Appl. Sci.* **2019**, *9*, 655, doi:10.3390/app9040655.
17. Zhou, H.; Zhou, G.; Song, X.; He, Q. Dynamic Characteristics of Canopy and Vegetation Water Content during an Entire Maize Growing Season in Relation to Spectral-Based Indices. *Remote Sens.* **2022**, *14*, 584, doi:10.3390/rs14030584.
18. Bhattacharya, A. Dry Matter Production, Partitioning, and Seed Yield Under Soil Water Deficit: A Review. In *Soil Water Deficit and Physiological Issues in Plants*; Bhattacharya, A., Ed.; Springer: Singapore, 2021; pp. 585–702 ISBN 978-981-336-276-5.
19. Baghdadi, N.; El Hajj, M.; Zribi, M.; Bousbih, S. Calibration of the Water Cloud Model at C-Band for Winter Crop Fields and Grasslands. *Remote Sens.* **2017**, *9*, 969, doi:10.3390/rs9090969.
20. Ranjbar, S.; Akhoondzadeh, M.; Brisco, B.; Amani, M.; Hosseini, M. Soil Moisture Change Monitoring from C and L-Band SAR Interferometric Phase Observations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7179–7197, doi:10.1109/JSTARS.2021.3096063.
21. Azadbakht, M.; Ashourloo, D.; Aghighi, H.; Homayouni, S.; Shahrabi, H.S.; Matkan, A.; Radiom, S. Alfalfa Yield Estimation Based on Time Series of Landsat 8 and PROBA-V Images: An Investigation of Machine Learning Techniques and Spectral-Temporal Features. *Remote Sens. Appl. Soc. Environ.* **2022**, *25*, 100657, doi:10.1016/j.rsase.2021.100657.
22. Sadenova, M.A.; Beisekenov, N.A.; Apshikur, B.; Khrapov, S.S.; Kapasov, A.K.; Mamysheva, A.M.; Klemeš, J.J. Modelling of Alfalfa Yield Forecasting Based on Earth Remote Sensing (ERS) Data and Remote Sensing Methods. *Chem. Eng. Trans.* **2022**, *94*, 697–702, doi:10.3303/CET2294116.
23. Dvorak, J.S.; Pampolini, L.F.; Jackson, J.J.; Seyyedhasani, H.; Sama, M.P.; Goff, B. Predicting Quality and Yield of Growing Alfalfa from a UAV. *Trans. ASABE* **2021**, *64*, 63–72, doi:10.13031/trans.13769.
24. Bahrami, H.; Homayouni, S.; Safari, A.; Mirzaei, S.; Mahdianpari, M.; Reisi-Gahrouei, O. Deep Learning-Based Estimation of Crop Biophysical Parameters Using Multi-Source and Multi-Temporal Remote Sensing Observations. *Agronomy* **2021**, *11*, 1363, doi:10.3390/agronomy11071363.
25. Sanderson, M.A.; Jones, R.M.; Read, J.C.; Lippke, H. Digestibility and Lignocellulose Composition of Forage Corn Morphological Components. *J. Prod. Agric.* **1995**, *8*, 169–174, doi:10.2134/jpa1995.0169.
26. Graybill, J.S.; Cox, W.J.; Otis, D.J. Yield and Quality of Forage Maize as Influenced by Hybrid, Planting Date, and Plant Density. *Agron. J.* **1991**, *83*, 559–564, doi:10.2134/agronj1991.00021962008300030008x.
27. Phipps, R.H.; Weller, R.F. The Development of Plant Components and Their Effects on the Composition of Fresh and Ensiled Forage Maize: 1. The Accumulation of Dry Matter, Chemical Composition and Nutritive Value of Fresh Maize. *J. Agric. Sci.* **1979**, *92*, 471–483, doi:10.1017/S0021859600063012.
28. Widdicombe, W.D.; Thelen, K.D. Row Width and Plant Density Effect on Corn Forage Hybrids. *Agron. J.* **2002**, *94*, 326–330, doi:10.2134/agronj2002.3260.
29. Morain, S.A.; Simonett, D.S. Vegetation Analysis with Radar Imagery. In Proceedings of the SYMP. ON REMOTE SENSING OF THE ENVIRON.; 1966.
30. Chiu, T.; Sarabandi, K. Electromagnetic Scattering from Short Branching Vegetation. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 911–925, doi:10.1109/36.841974.

31. Wu, L.; Moore, R.K.; Zoughi, R. Sources of Scattering from Vegetation Canopies at 10 GHz. *IEEE Trans. Geosci. Remote Sens.* **1985**, *GE-23*, 737–745, doi:10.1109/TGRS.1985.289392.
32. Patel, P.; Srivastava, H.S.; Panigrahy, S.; Parihar, J.S. Comparative Evaluation of the Sensitivity of Multi-polarized Multi-frequency SAR Backscatter to Plant Density. *Int. J. Remote Sens.* **2006**, *27*, 293–305, doi:10.1080/01431160500214050.
33. Evans, D.L.; Farr, T.G.; van Zyl, J.J.; Zebker, H.A. Radar Polarimetry: Analysis Tools and Applications. *IEEE Trans. Geosci. Remote Sens.* **1988**, *26*, 774–789, doi:10.1109/36.7709.
34. PREVOT, L.; DECHAMBRE, M.; TACONET, O.; VIDAL-MADJAR, D.; NORMAND, M.; GALLEJ, S. Estimating the Characteristics of Vegetation Canopies with Airborne Radar Measurements. *Int. J. Remote Sens.* **1993**, *14*, 2803–2818, doi:10.1080/01431169308904310.
35. Ulaby, F. Radar Response to Vegetation. *IEEE Trans. Antennas Propag.* **1975**, *23*, 36–45, doi:10.1109/TAP.1975.1140999.

Keywords:

Alfalfa Yield and quality estimation, Satellite radar imagery, Machine learning, open-source web tool