ESTIMATING CANOLA'S BIOPHYSICAL PARAMETERS FROM TEMPORAL, SPECTRAL, AND POLARIMETRIC IMAGERY USING MACHINE LEARNING APPROACHES

O. Reisi Gahrouei[1, *], S. Homayouni[2], A. Safari[1]

[1] Dept. of Geomatics Engineering, College of Eng., U. of Tehran, Iran, {omid.reisi.gahroi, asafari}@ut.ac.ir

[2] Centre Eau Terre Environnement, Institut National de la Recherche Scientifique, Québec, Canada, Saeid_Homayouni@ete.inrs.ca

KEYWORDS: Satellite Earth Observations, Crop parameters, Support Vector Regression (SVR), Artificial Neural Network (ANN)

Abstract

The objective of this study was to investigate the application of multi-temporal optical and polarimetric synthetic aperture radar (PolSAR) Earth observations for crop characterization. Crop dry biomass, Leaf Area Index (LAI), and Plant Water Content (PWC) were estimated and assessed using Machin learning approaches. An accurate estimation of crop parameters provides essential information to increased food production and plays a crucial role in the management of agricultural lands. Multispectral and PolSAR data provide valuable observations of spectral and structural properties which are essential for crops parameter modelling. The Earth observations used in this paper were collected by RapidEye satellites and Unmanned Aerial Vehicle Synthetic Aperture Radar (UAVSAR) system in the summer of 2012, over an agriculture area in Winnipeg, Manitoba, Canada. The RapidEye vegetation indices (VIs) and UAVSAR polarimetric parameters were used as inputs in artificial neural network (ANN) and support vector regression (SVR) models for canola biophysical parameters estimation. The best models were provided by SVR for canola. Also combining optical VIs and polarimetric features appeared as a powerful tool for crop parameters estimation in agricultural lands.

1. INTRODUCTION

Crop forecasting helps to predict crop yields and production before harvesting, which is an essential factor for national food security, food trade, market management and evaluation of supply and demand (J. Liu, Pattey, and Jégo 2012; Wiseman et al. 2014; Yue et al. 2017). Information on crop yield can be used to support decisions made by national and regional agencies regarding the allocation of agricultural products and commercial institutions for investment in various sectors (Hoogenboom 2000). Dry biomass and leaf area index (LAI) are essential indicators for crop production and has a direct relationship with it. Also, plant water content (PWC), the total amount of water in stems and leaves, is one of the most critical parameters for the successful retrieval of soil moisture content from active and passive microwave remote sensing (Jackson et al., 1982, 2004).

Satellite and aerial observations were utilized with different targets in agricultural in recent decades. Various sensors have unique advantage and cons, thus using suitable data is very important for modelling and mapping in agricultural lands. Multispectral, SAR and hyperspectral data were used in many studies about agricultural such as biophysical parameters estimation, crop mapping, crop phenology estimation, etc. (Kross et al. 2015; Li et al. 2019; J. Liu et al. 2010). Wiseman et al. (2014) assessed RADARSAT-2 C-band polarimetric SAR for agricultural production monitoring.

They compared 21 polarimetric parameters with dry biomass of canola, corn, soybean, and winter wheat, then used linear and logarithmic regression to assess SAR response to crop biomass.

Kross et al. (2015) reported on the correlation between nine vegetation indices extracted from multi-temporal RapidEye images, and biomass and LAI for corn and soybean crops in Eastern Canada. They modelled the relationship between indices and crop parameters with linear and exponential regressions and showed that the best estimation accuracy for biomass and LAI could be obtained for corn.

In the past decade, remote sensing sensors are dramatically improved in different aspects. The advantages of the new remote sensing systems include better temporal resolutions (e.g., daily to every few days), higher spatial resolutions (e.g., better than 5 m), and augmented spectral information (e.g., visible, near-infrared, and red-edge bands) provide compelling information and tools for various thematic applications (Reisi-gahrouei et al. 2019). Besides, a variety of sensors, such as multispectral, hyperspectral, SAR, etc., bring about the facility of data integration and joint analysis.

SAR sensors propagate pulses of microwave energy and as such, are not affected by atmospheric and illumination conditions, yet are sensitive to biophysical properties of land cover surfaces(Li et al. 2019). On the other hand, multispectral optical sensors provide

spectral information about vegetation for detecting and modelling of crops characteristics. Sensors with the low spatial resolution are more suited for monitoring of agricultural fields at national and global scales. Nevertheless, high temporal and spatial resolution satellite sensors such as RADARSAT-2, RapidEye, Sentinel-2A/2B, and PlanetScope are gathering growing interest from agronomists in their efforts to better understand field-scale variability of crop conditions in space and time.

Furthermore, mathematical algorithms are updating and using a suitable method for modelling is a vital factor in estimating plants' parameters in agricultural lands. In recent years, machine learning and artificial intelligence have found a proponent role in modelling and classification, and these methods bring about high accuracy of parametres estimation in agricultural lands.

Canola produces a healthy oil and a highly nutritious animal feed, better than any other oilseed crop. Also, canola is being tested as alternative oil for petroleum products like motor fuels (called biofuels) and motor oils. Canola is one of the leading agricultural products in Canada, and this country has the world's second-largest grower of canola. Therefore accurate estimation of its parameters helps to manage and increase yield (McNairn et al. 2018; Wiseman et al. 2014).

The first goal of this study was to estimate canola's biophysical parameters by RapidEye vegetation indexes (VIs) and UAVSAR polarimetric features. Also, investigation ability to support vector regression and artificial neural network in biophysical parameters are modelling in agricultural lands.

2.1. Study area

The study site is an agricultural area in Winnipeg a city of Manitoba, Canada entered at (49°36N 97°35′W). Being located very far inland, the local climate is exceptionally seasonal even by Canadian standards with average January lows of around −21 °C (−6 °F) and average July highs of 26 °C (79 °F).
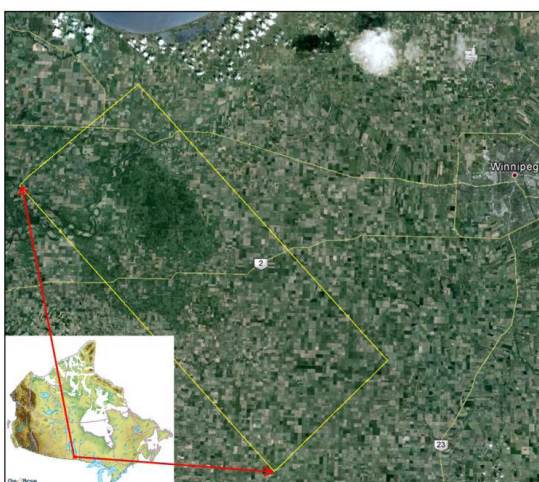

Figure 1. Location of study site Winnipeg in southern Manitoba, Canada (Geobase.ca © and Google Earth ©)

3. PROPOSED METHOD

The satellite time series is composed of 7 RapidEye, and UAVSAR images and images were collected at the same time from 17 Jun to 17 Jul in 2012. A total of 11 VIs and 14 polarimetric feature were extracted from RapidEye and UAVSAR data and were used as input in ANN and SVR for modelling, and soil moisture active passive validation experiment camping (SMAPVEX2012) were used as in-situ Observations. All RapidEye images were atmospherically corrected using PCI Geomatica's ATCOR2 (v10.3) and then orthorectified in PCI Geomatica's OrthoEngine (v10.3) using the rational functions. Additional GCPs were manually collected from the regional roads network maps from the National Topographic Database of Canada. VIs were used in this study consist of NDVI, GNDVI, SR, NDVI-Red Edge, SR-Red Edge, MTVI2, EVI, CL-Green, CL-Red Edge, MTCI, and SAVI. The UAVSAR data were Calibrating by NASA JPL and ortho-rectification(Hosseini and McNairn 2017). For speckle reduction, a 3×3 boxcar filtered was applied. Several radar parameters were generated from these matrices including linear backscatter intensities (i.e. HH, HV, and VV), Cloude–Pottier (Entropy (H), Anisotropy (A) and Alpha Angele (α)), Freeman-Durden (Surface, Double, and Volume Scattering), Yamaguchi decompositions (Surface, Double, Volume, and Helix Scattering), and Radar Vegetation Index (RVI)(Yamaguchi et al. 2005; Lee and Pottier 2009). In this study, we used a feedforward multilayer perceptron, which consists of one input layer, hidden layers, and one output layer. The ANN was trained with the Levenberg-Marquardt (1963) algorithm, and nonlinear sigmoid functions (Eq. 2) were used as the cost function MLP used to describe any general feedforward network. It utilizes a nonlinear activation function that lets it classify data that are not linearly separable (Li et al. 2012). Every node in a layer connects to each node in the following layer making the network fully connected. Minimizing the output errors on a particular set of training data by adjusting the network weights, w is the aim of neural network training. The mathematical representation of an artificial neuron, (n_1) at a sample (i_1), and its cost function are given by Eq.1. There are many choices for designing a neural network (NN). Nevertheless, a poor selection of any of the corresponding parameters can affect the efficiency of the NN. There have been some attempts to generate formulas or rules for designing the structure of a NN. Eq. (3) was developed to create a range of the desired number of hidden neurons, i.e., h, based on the number of inputs and outputs of a NN (Chen et al. 2008):

$$a_1(i_1) = \sum_{j=1}^{n} w_{j1}x_j + b_1(i_1) \quad (1)$$

Where, $x_1,..., x_n$ are inputs, $w_1,..., w_n$ are the associated connection weights and b_1 is the bias value.

$$K(x, \acute{x}) = exp(-\frac{\|x-\acute{x}\|^2}{2\sigma^2}) \quad (2)$$

$$h = \sqrt{n + m} + (1\sim10) \quad (3)$$

where n is the number of inputs, m is the number of outputs.

The Support Vector Regression (SVR) is a powerful tool for modelling in remote sensing. It works by the same principles as the SVM for classification, with only a few minor differences. First, because the output is a real number, it becomes challenging to predict the information at hand, which has infinite possibilities

(Cristianini and Shawe-Taylor 2000). For nonlinear SVR, some kernel functions have been found to provide functional generalization capabilities. These include linear, polynomial, radial basis function (RBF) and sigmoid functions (Javed et al. 2009). According to different studies, RBF is more common and efficient than other kernels in biomass and LAI modelling(Ali et al. 2017; Wang et al. 2016). Radial basis kernel parameters were tuned using the Grid search method (C. Liu et al. 2014). Grid search methodically builds and evaluates a model for each combination of algorithm parameters specified in a grid. Also, a cross-validation method was implemented to set the number of neurons in the hidden layer. Moreover, the range of the desired number of hidden neurons was calculated from 6 to 15, by the Eq. 3 as n = 25, i.e., number of VIs, m = 1, i.e., number of unknown variables which is either dry biomass, LAI, and WC.

$$y_1 = f\big(a_1(i_1)\big) = \frac{1}{1+e^{-a_1(i_1)}} \quad (4)$$

Three statistical criteria were used to evaluate the performance of each model - Root Mean Square Error (RMSE), the coefficient of determination (R^2) and Mean Absolute Error (MAE). RMSE and MAE characterize the mean differences between measured and estimated variables: MAE is less sensitive to extreme data values (Willmott and Matsuura 2005)

$$R^2 = 1 - \frac{SS_E}{SS_T} \quad (5)$$

$$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2 \quad , \quad SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

4. RESULTS AND DISCUSSION

After pre-processing the RapidEye and UAVSAR data, the in situ samples, VIs, and polarimetric features were extracted and considered as inputs to the ANN and SVR. These samples were used for both calibration and validation of the ANN and SVR models. From all the sampling observations, 60% were selected randomly and used for calibration, with the remaining 40% used for validation. The models had been obtained by calibration data, and then that was evaluated by remaining samples (validation), and consequently, the accuracies were determined using validation data. In this process, the values of all training data (VIs and polarimetric features) were normalized between 0 and 1 as ANN inputs due to the nature of the nonlinear sigmoid function. Hyper-parameters in the case of SVR and the number of neurons in the hidden layer of ANN have been obtained by grid search, and cross-validation are reported in Table 1. Canola has broad leaves with a very characteristic change in the canopy over the growing season. Upon emergence, the crop develops a thick rosette of leaves close to the ground before sending up a flowering stalk. After flowering and repining, canola usually is faced with chlorophyll reduction. According to the sensitivity of VIs to chlorophyll content, this can reduce the accuracy of biomass and LAI estimation. As such, aside

from genetics, chlorophyll is significantly affected by the environment(Ward et al. 1995) and is highly weather dependent. Therefore in this study have been used different Vis and polarimetric parameters with different properties for improving modeling. Results showed that the SVR modelled LAI, biomass and PWC are more accurate than ANN for canola. VIs modelled crop parameters based crop spectral properties moreover, the radar signal is sensitive to crop structure in different growth stages. Because canola crop is a broadleaf plant with a very distinctive change in canopy structure through the growing season, consequently using temporal, spectral, and polarimetric observations and machine learning simultaneously were beneficial and reliable. The accuracies and scatterplots of canola parameters modelling by ANN and SVR are reported in Table 2 and Fig 2.

Table 1. Parameters used in ANN and SVR

Model	Parameter	C	Gamma	Epsilon
SVR	Biomass	71	0.089	0.01
	LAI	67	0.092	0.09
	WC	75	0.09	0.01
	Parameter	Number of hidden neurons		
ANN	Biomass	8		
	LAI	11		
	WC	7		

5. CONCLUSIONS

In this study, we investigated the potential of combing several multi-temporal vegetation indices and polarimetric features to estimate crop parameters (dry biomass, LAI, and PCW) using machine learning approaches for canola. 14 polarimetric feature and 11 vegetation indexes were used as inputs in support vector regression and neural natural network models for canola's biomass, leaf area index (LAI) and water content (WC) modelling. The results showed that modelling was by SVR more accurately than ANN. Also combining VIs and polarimetric features can use as a powerful tool for crop parameters estimation in agricultural lands. Canola had a strong correlation with volume scattering features because it is normally planted more compact than other crop and combination of structure descriptors (polarimetric features) and spectral descriptors (VIs) bring about sutable estimation in this study. SVR's coefficients of correlation (R^2) were found as 0.82, 0.87, and 0.81 for dry biomass, LAI, and PWC modelling respectively. The SRV's RMSEs were respectively 44.33 g.m^{-2}, 0.75 m^2 m^{-2}, and 2.33 % for dry biomass, LAI and PWC modelling.

Table 1. Correlation statistics of measured and estimated total dry biomass, LAI, and PWC using ANN and SVR

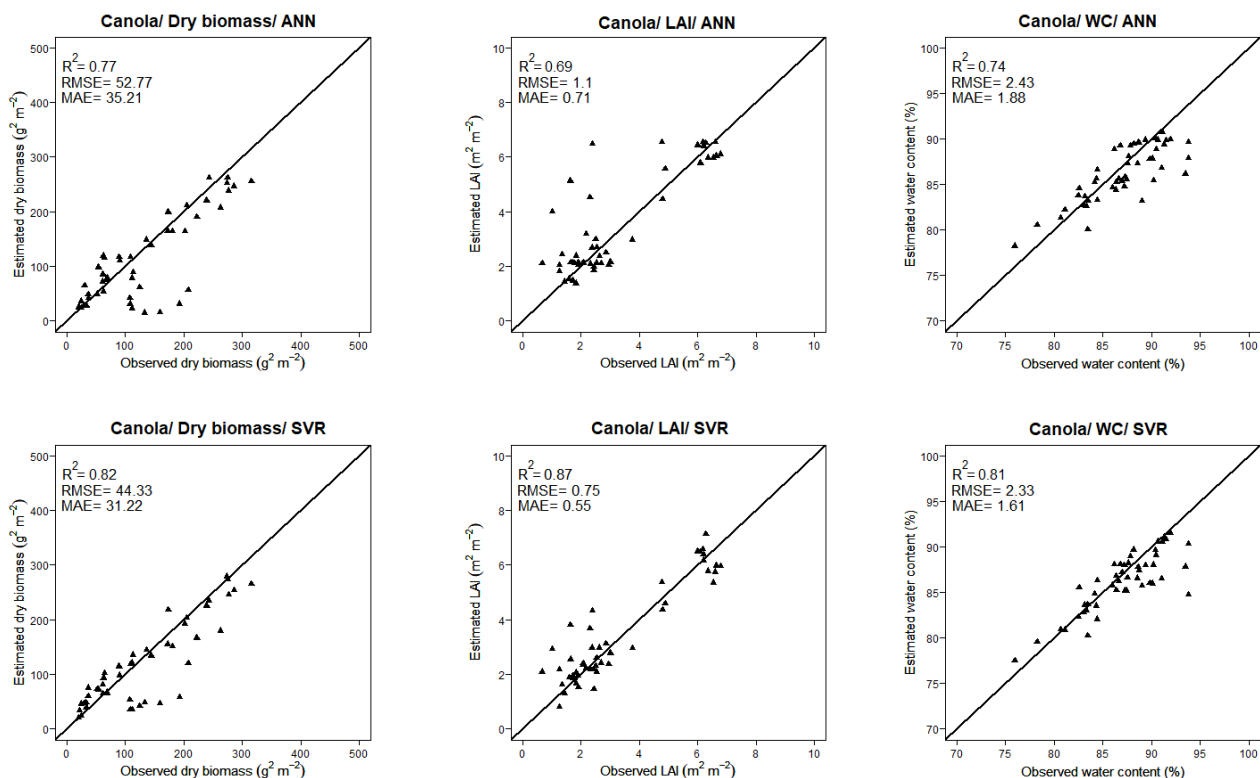| Parameter	Model	Validation			Calibration		
		R^2	RMSE	MAE	R^2	RMSE	MAE
Dry biomass (g/m²)	SVR	0.82	44.33	31.22	0.95	15.75	10.48
	ANN	0.77	52.77	35.21	0.89	18.20	13.60
LAI (m²/m²)	SVR	0.87	0.75	0.55	0.92	0.25	0.20
	ANN	0.69	1.10	0.71	0.87	0.47	0.35
PWC (%)	SVR	0.81	2.33	1.61	0.92	0.61	0.47
	ANN	0.74	2.43	1.88	0.81	1.61	1.28

Figure 2. One-to-one relationships between predicted and measured crop parameters; modelling by ANN and SVR.

REFERENCES

Ali, Iftikhar, Fiona Cawkwell, Edward Dwyer, and Stuart Green. 2017. "Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data-A Machine Learning Approach." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10 (7): 3254–64. https://doi.org/10.1109/JSTARS.2016.2561618.

Cristianini, Nello., and John. Shawe-Taylor. 2000. *An Introduction to Support Vector Machines : And Other Kernel-Based Learning Methods*. Cambridge University Press.

Hoogenboom, Gerrit. 2000. "Contribution of Agrometeorology to the Simulation of Crop Production and Its Applications." *Agricultural and Forest Meteorology* 103 (1–2): 137–57. https://doi.org/10.1016/S0168-1923(00)00108-8.

Hosseini, Mehdi, and Heather McNairn. 2017. "Using Multi-Polarization C- and L-Band Synthetic Aperture Radar to Estimate Biomass and Soil Moisture of Wheat Fields." *International Journal of Applied Earth Observation and Geoinformation* 58: 50–64. https://doi.org/10.1016/j.jag.2017.01.006.

Javed, Faizan, Gregory S.H. Chan, Andrey V. Savkin, Paul M. Middleton, Philip Malouf, Elizabeth Steel, James Mackie, and Nigel H. Lovell. 2009. "RBF Kernel Based Support Vector Regression to Estimate the Blood Volume and Heart Rate Responses during Hemodialysis." *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, 4352–55. https://doi.org/10.1109/IEMBS.2009.5332739.

Kross, Angela, Heather McNairn, David Lapen, Mark Sunohara, and Catherine Champagne. 2015. "Assessment of RapidEye Vegetation Indices for Estimation of Leaf Area Index and Biomass in Corn and Soybean Crops." *International Journal of Applied Earth Observation and Geoinformation* 34 (1): 235–48. https://doi.org/10.1016/j.jag.2014.08.002.

Lee, Jong-Sen., and Eric. Pottier. 2009. *Polarimetric Radar Imaging : From Basics to Applications*. CRC Press.

Li, Huapeng, Ce Zhang, Shuqing Zhang, and Peter M. Atkinson. 2019. "Full Year Crop Monitoring and Separability Assessment with Fully-Polarimetric L-Band UAVSAR: A Case Study in the Sacramento Valley, California." *International Journal of Applied Earth Observation and Geoinformation* 74 (August 2018): 45–56. https://doi.org/10.1016/j.jag.2018.08.024.

Liu, Chao, Shi Qun Yin, Miao Zhang, Yi Zeng, and Jiang Yue Liu. 2014. "An Improved Grid Search Algorithm for Parameters Optimization on SVM." *Applied Mechanics and Materials* 644–650 (2): 2216–19. https://doi.org/10.4028/www.scientific.net/AMM.644-650.2216.

Liu, Jiangui, Elizabeth Pattey, and Guillaume Jégo. 2012. "Assessment of Vegetation Indices for Regional Crop Green LAI Estimation from Landsat Images over Multiple Growing Seasons." *Remote Sensing of Environment* 123: 347–58. https://doi.org/10.1016/j.rse.2012.04.002.

Liu, Jiangui, Elizabeth Pattey, John R. Miller, Heather McNairn, Anne Smith, and Baoxin Hu. 2010. "Estimating Crop Stresses, Aboveground Dry Biomass and Yield of Corn Using Multi-Temporal Optical Data Combined with a Radiation Use Efficiency Model." *Remote Sensing of Environment* 114 (6): 1167–77. https://doi.org/10.1016/j.rse.2010.01.004.

McNairn, Heather, Xianfeng Jiao, Anna Pacheco, Abhijit Sinha, Weikai Tan, and Yifeng Li. 2018. "Estimating Canola Phenology Using Synthetic Aperture Radar." *Remote Sensing of Environment* 219 (October): 196–205. https://doi.org/10.1016/j.rse.2018.10.012.

Reisi-gahrouei, Omid, Saeid Homayouni, Heather Mcnairn, and Mehdi Hosseini. 2019. "Crop Biomass Estimation Using Multi Regression Analysis and Neural Networks from Multitemporal L-Band Polarimetric Synthetic Aperture Radar Data." *International Journal of Remote Sensing* 0 (0): 1–19. https://doi.org/10.1080/01431161.2019.1594436.

Wang, Li'ai, Xudong Zhou, Xinkai Zhu, Zhaodi Dong, and Wenshan Guo. 2016. "Estimation of Biomass in Wheat Using Random Forest Regression Algorithm and Remote Sensing Data." *Crop Journal* 4 (3): 212–19. https://doi.org/10.1016/j.cj.2016.01.008.

Ward, K., R. Scarth, J. K. Vessey, and J. K. Daun. 1995. "Chlorophyll Degradation in Summer Oilseed Rape and Summer Turnip Rape during Seed Ripening." *Canadian Journal of Plant Science* 75 (2): 413–20. https://doi.org/10.4141/cjps95-069.

Willmott, CJ, and K Matsuura. 2005. "Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance." *Climate Research* 30 (1): 79–82. https://doi.org/10.3354/cr030079.

Wiseman, Grant, Heather McNairn, Saeid Homayouni, and Jiali Shang. 2014. "RADARSAT-2 Polarimetric SAR Response to Crop Biomass for Agricultural Production Monitoring." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (11): 4461–71. https://doi.org/10.1109/JSTARS.2014.2322311.

Yamaguchi, Y., T. Moriyama, M. Ishido, and H. Yamada. 2005. "Four-Component Scattering Model for Polarimetric SAR Image Decomposition." *IEEE Transactions on Geoscience and Remote Sensing* 43 (8): 1699–1706. https://doi.org/10.1109/TGRS.2005.852084.

Yue, Jibo, Guijun Yang, Changchun Li, Zhenhai Li, Yanjie Wang, Haikuan Feng, and Bo Xu. 2017. "Estimation of Winter Wheat Above-Ground Biomass Using Unmanned Aerial Vehicle-Based Snapshot Hyperspectral Sensor and Crop Height Improved Models." *Remote Sensing* 9 (7). https://doi.org/10.3390/rs9070708.