

## SOYBEAN YIELD FORECAST USING DUAL-POLARIMETRIC C-BAND SYNTHETIC APERTURE RADAR

Mehdi Hosseini<sup>1\*</sup>, Inbal Becker-Reshef<sup>1</sup>, Ritvik Sahajpal<sup>1</sup>, Pedro Lafluf<sup>2</sup>, Guillermo Leale<sup>2,3</sup>, Estefania Puricelli<sup>1</sup>, Sergii Skakun<sup>1</sup>, Heather McNairn<sup>4</sup>

<sup>1</sup>NASA-Harvest, Department of Geographical Sciences, University of Maryland, College Park, MD, USA - [mhosseini@umd.edu](mailto:mhosseini@umd.edu)

<sup>2</sup>SIMA

<sup>3</sup>Departamento Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional, Facultad Regional Rosario, Argentina

<sup>4</sup>Science and Technology Branch, Agriculture and Agri-Food Canada, Ottawa, ON, Canada

**KEY WORDS:** Soybean, Yield, Sentinel-1, Polarimetry, Neural Network

### ABSTRACT:

Crop yield forecast is important, but determining productivity is an on-going challenge for agricultural communities including policy makers. An increase in the number of satellites, improved temporal and spatial resolutions and more open data policies, are leading to wider interest by the agricultural sector in exploiting space-based data at moderate spatial resolutions (10-30 m) and varying wavelengths (optical and microwave) for crop yield monitoring. This study evaluated Sentinel-1 dual-polarimetric data to forecast soybean yields one month before the harvest at field scales over a site in central Argentina. Specifically, polarimetric features were extracted from the Sentinel-1 data using the M-Chi decomposition. An Artificial Neural Network (ANN) model was trained using a time series of the single-bounce and volume scattering parameters derived from M-Chi from stacks of data acquired during the growing season. To estimate soybean yield from the ANN model, an innovative iterative retrieval method was developed. This retrieval approach improved the accuracies of the soybean yield forecast delivering a final coefficient of determination ( $R^2$ ) of 0.81 with a root mean square error (RMSE) of 755.81 kg/ha and mean absolute error (MAE) of 581.65 kg/ha. These accuracies demonstrate a high potential of Synthetic Aperture Radar (SAR) data from Sentinel-1 for soybean yield forecast at field scales.

### 1. INTRODUCTION

Significant effort has been invested in assessing the use of remote sensing data to provide estimates of crop yields at national and regional scales, information critical for monitoring food security and driving policy decisions (Ahmad et al., 2018; Donohue et al., 2018; Franch et al., 2019; Khaki et al., 2021). However, crop yield estimation at field scales is more challenging partly due to difficulties in accessing ancillary high resolution soil moisture data and information on site specific management such as fertilizer application. High quality data on field scale yield is also difficult to acquire.

Studies on satellite-based crop yield estimation have been primarily carried out using coarse (250 m – 1 km) to moderate (10–30 m) resolution optical satellite data, utilizing vegetation indices or surface reflectance as features (Burke and Lobell, 2017; Skakun et al., 2019). Yet there are challenges with using optical data for crop yield monitoring. These sensors are obstructed by clouds, cloud shadows and haze which reduce the temporal frequency of useful data and impedes the implementation of optical-based monitoring given the rapidly changing crop canopy and need to capture development throughout the growing season. The relatively small wavelengths used in optical sensing also limit sensing of crop properties to the top of the canopy, with less scattering originating from within the canopy. Synthetic Aperture Radar (SAR) satellites are an alternative source of data and are not impacted by these limitations (Hosseini and McNairn, 2017). SARs are active sensors which propagate microwave signals at longer wavelengths, unimpeded by cloud cover and as such, are able to provide images of the Earth even during cloudy periods. Also, microwave scattering is sensitive to the structure of crop canopies and given that canopy structure changes with advancing phenology, SARs are able to track crop development during the entire active growing season and into late stages of crop growth (Hosseini and McNairn, 2017). These imaging advantages,

coupled with open data policies for sensors such as Sentinel-1, create opportunity to advance efforts to estimate crop yield and model productivity forecasts.

Previous SAR studies were primarily focused on assessing the use of linear co-polarization and cross-polarization intensities to estimate yields (Setiyono et al., 2019), although some studies examined the phase information from quad-polarimetric SARs for crop monitoring (Homayouni et al., 2019). Unlike quad-polarimetric capable SAR sensors (RADARSAT-2, ALOS-2 PALSAR-2 and TerraSAR-X), Sentinel-1 acquisitions over land are typically limited to dual-polarization in the Interferometric Wide (IW) mode. As such, not all methods applied to fully polarimetric imagery, including many target decompositions, can be applied to Sentinel-1 data. This study investigates the decomposition of dual-pol SAR data for crop yield estimation. We show that the M-Chi decomposition method, that was originally developed for compact polarimetry SAR data (Raney et al., 2012), can be effectively applied to Sentinel-1 data. The derived M-Chi parameters, as well as the backscatter intensities extracted from Sentinel-1, are tested for soybean yield estimation at field scales. Machine learning algorithms have been widely used in recent years for crop monitoring applications (Reisi-Gahrouei et al., 2020; Mandal et al., 2019). In this study, the Artificial Neural Network (ANN) algorithm is trained and tested for soybean yield estimation.

### 2. STUDY AREA AND DATASET

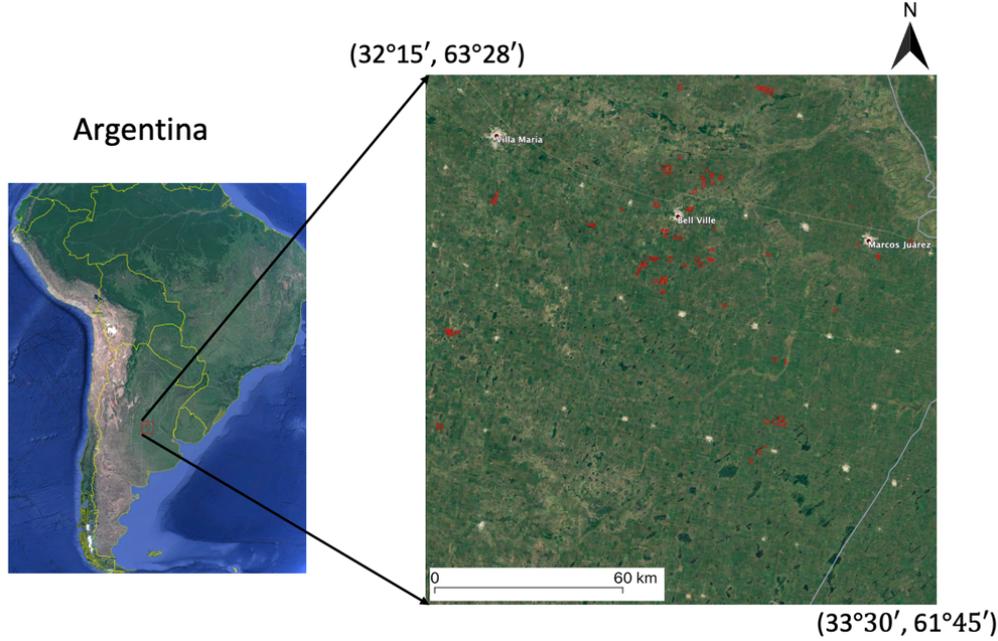
The study area is located in central Argentina (Figure 1). The yield data were collected for the 2017-2018 growing season for 58 soybean fields which are distributed across an area of about 150 km by 220 km. The yield values are in the range of 1.25 t/ha to 5.5 t/ha.

Sentinel-1 data acquired in the Interferometric Wide (IW) were used in this study. Forty Ground Range Detected (GRD) and Single Look Complex (SLC) products were downloaded and

\*Corresponding author.

preprocessed. GRD data were converted to VV and VH intensities. Images were filtered using a boxcar 5 by 5 speckle filter to reduce speckle noise. The Range Doppler Terrain Correction was applied to ortho-rectify the data stacks. The SLC products were also radiometrically and geometrically corrected

using the European Space Agency (ESA) Sentinel-1 Toolbox software. The M-Chi decomposition was applied to the SLC products to derive single bounce, double bounce and volume scattering components. All images were resampled to a 10 m spatial resolution.



**Figure 1.** Study area in central Argentina. The soybean fields are shown in red polygons.

### 3. METHODOLOGY

#### 3.1 M-Chi decomposition

In the Interferometric Wide (IW) swath mode, Sentinel-1 transmits a vertically or horizontally polarized signal and receives the orthogonal backscattered signals in both vertical and horizontal polarizations. The phase between the two received polarizations is retained and can thus be exploited. Quad-polarimetric (QP) SAR satellites both transmit and receive in two orthogonal polarizations (with phase recorded), leading to a 4-element scattering matrix (Eq. 1). Dual-polarimetric (DP) Sentinel-1 (such as VV, VH) has a 2-element scattering matrix (Eq. 2):

$$[S]_{QP} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \quad (1)$$

$$[S]_{DP} = [S_{VH} \ S_{VV}] \quad (2)$$

where  $[S]_{QP}$  is the scattering matrix for quad-polarimetric SAR,  $[S]_{DP}$  is the scattering matrix for dual-polarimetric SAR, and  $S_{XY}$  is the element of the scattering matrix in X-transmit and Y-received polarizations.

Well known and widely used decomposition algorithms such as Freeman-Durden, Yamaguchi, Cloude-Pottier and Van-Zyl were developed for quad-polarimetric SAR data using the 9-element covariance matrix (Eq. 3) (Homayouni et al., 2019). Dual-pol SAR has a 4-element covariance matrix (Eq. 4) and consequently, the polarimetric decompositions applicable to the dual-pol SAR are limited.

$$[C]_{QP} = \begin{bmatrix} \langle S_{HH}S_{HH}^* \rangle & \langle S_{HH}S_{VV}^* \rangle & \langle S_{HH}S_{VH}^* \rangle \\ \langle S_{VV}S_{HH}^* \rangle & \langle S_{VV}S_{VV}^* \rangle & \langle S_{VV}S_{VH}^* \rangle \\ \langle S_{VH}S_{HH}^* \rangle & \langle S_{VH}S_{VV}^* \rangle & \langle S_{VH}S_{VH}^* \rangle \end{bmatrix} \quad (3)$$

$$[C]_{DP} = \begin{bmatrix} \langle S_{VV}S_{VV}^* \rangle & \langle S_{VV}S_{VH}^* \rangle \\ \langle S_{VH}S_{VV}^* \rangle & \langle S_{VH}S_{VH}^* \rangle \end{bmatrix} \quad (4)$$

where  $C$  is the covariance matrix,  $\langle \rangle$  is the spatial average, and  $*$  is the complex conjugate.

The M-Chi decomposition was developed and used for compact-pol SAR (Raney et al., 2012). The features from M-Chi decompositions are calculated as follows.

$$B = ((mS_1 + S_4)/2)^{0.5} \quad (5)$$

$$R = ((mS_1 - S_4)/2)^{0.5} \quad (6)$$

$$G = (S_1(1 - m))^{0.5} \quad (7)$$

$$m = (S_2^2 + S_3^2 + S_4^2)^{0.5}/S_1 \quad (8)$$

where  $B$  is single-bounce and represents the polarimetric portion of the received signal after a single reflection from the target (soil or crop),  $R$  is double-bounce representing the portion of the signal received after two reflections from the soil-crop or crop-crop interaction, and  $G$  is the volume scattering of the received signal after multiple reflections from within crop canopies or between the soil and crop.  $m$  is the degree of polarization, and  $S_1$ ,  $S_2$ ,  $S_3$  and  $S_4$  are the Stokes parameters derived using the following equations:

$$S_1 = \langle |E_H|^2 \rangle + \langle |E_V|^2 \rangle \quad (9)$$

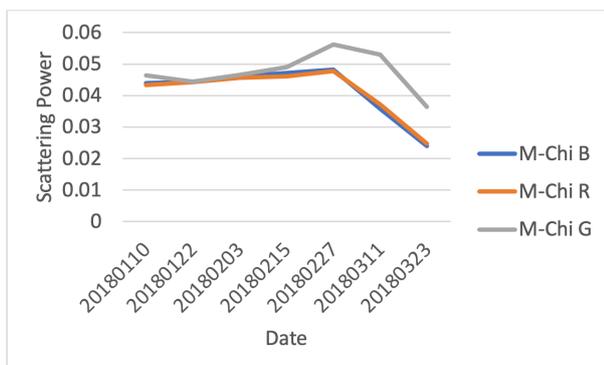
$$S_2 = \langle |E_H|^2 \rangle - \langle |E_V|^2 \rangle \quad (10)$$

$$S_3 = 2\langle E_H E_V \cos \phi_{HV} \rangle \quad (11)$$

$$S_4 = 2\langle E_H E_V \sin \phi_{HV} \rangle \quad (12)$$

$E_H$  and  $E_V$  are the horizontal and vertical components of the received signals and  $\Phi_{HV}$  is the phase difference between them. All four Stokes parameters can be derived from dual-polarized SAR satellites including data collected in the Sentinel-1 IW mode. The challenge in applying the M-Chi decomposition to dual-pol Sentinel-1 is that the ellipticity angle is close to zero, meaning that it is challenging to distinguish between single and double bounce targets (Raney, 2007). This presents an important limitation when the objective is to distinguish between different types of polarimetric scattering.

For most distributed targets, including those with a dominant vertical structure, single, double and volume scattering all contribute to total scattering. However, for agriculture targets one scattering mechanism typically dominates and the mix of mechanisms changes as the canopy develops (McNairn et al., 2002; McNairn et al., 2009). Here, we apply the M-Chi decomposition over agricultural fields two months after planting. Although a mixture of scattering mechanisms can be expected, volume scattering will dominate during this period of advanced canopy development. Our objective is not to define the dominated scattering but rather to evaluate if the scattering components B, R and G derived from Sentinel-1 SLC data (equations 5-7) are correlated with crop yield. Because the ellipticity angle for dual-pol SAR is close to zero, the powers of the single bounce (B) and double bounce (R) are very close, and the power of the volume scattering (G) is significantly higher and is illustrated in Figure 2. As such yield estimates from single-bounce and double-bounce components would be expected to be equivalent, and is verified in section IV.



**Figure 2.** Time-series of the average scattering powers of all 58 fields for the three scattering components of the M-Chi decomposition including single-bounce (M-Chi B), double-bounce (M-Chi R) and volume scattering (M-Chi G).

In this study the parameters derived from M-Chi decomposition, was applied to Sentinel-1 data, are used to train an ANN algorithm for yield estimation.

### 3.2 ANN Algorithm

The ANN algorithm used in this study is a feed-forward multi-layer perceptron (MLP) network that has one input layer, two hidden layers and one output layer. Each hidden layer has 10 neurons. The Sigmoid function was used as the cost function and

the Levenberg-Marquardt backpropagation algorithm (Yu and Wilamowski, 2011) was used to train the algorithm.

### 3.3 Feature Selection

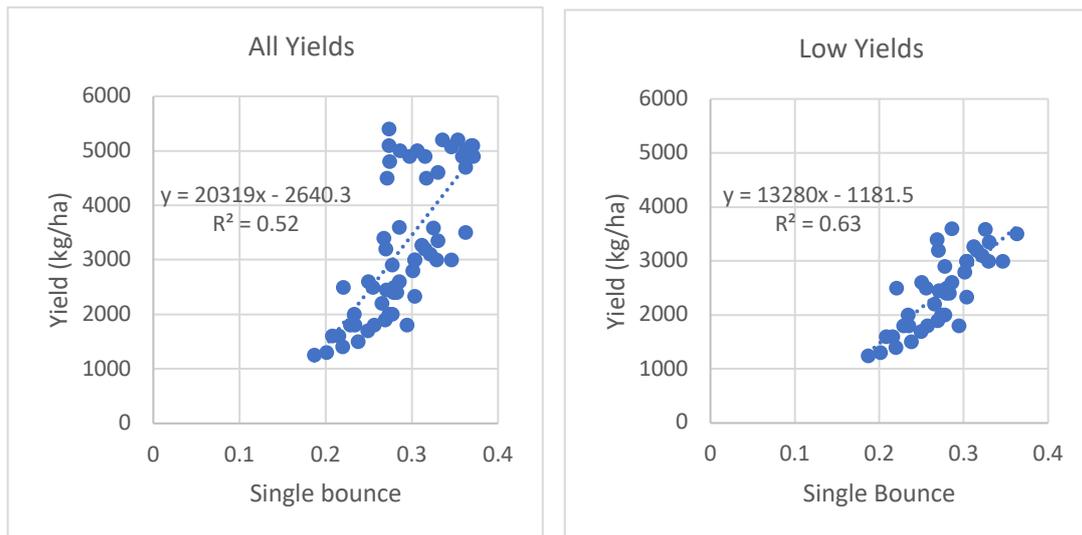
Following the application of the M-Chi decomposition to Sentinel-1 SLC products, the field measured yield of soybeans was correlated with the single-bounce and volume scattering parameters using first-order polynomial functions. As explained above, single-bounce and double-bounce are equivalent so we used just one of them. Compared to the volume scattering, the single-bounce parameter had the higher correlation with soybean yield. However, because volume scattering is the dominant scattering in mid to late growing season (Figure 2), we used both parameters as input features for training the model. Therefore, the median single-bounce scattering component, and the median volume scattering of all pixels for each of the 58 soybean fields, were used to train the Artificial Neural Network (ANN) model for yield estimation. To minimize the impact of soil on the SAR backscattered signal, the time-series commenced two months after the planting dates. Fields were planted late-November 2017 to early December 2017. The Sentinel-1 time-series started from February 2018.

Given the 12 day revisit of Sentinel-1A, seven images were available from February 2018 until soybean harvest. A 5-fold cross-validation approach (Kuhn and Johnson, 2013) was applied to train and validate the model. Therefore, in each fold, 80 percent of the points were selected randomly for the training and with the remainder of points used for the validation.

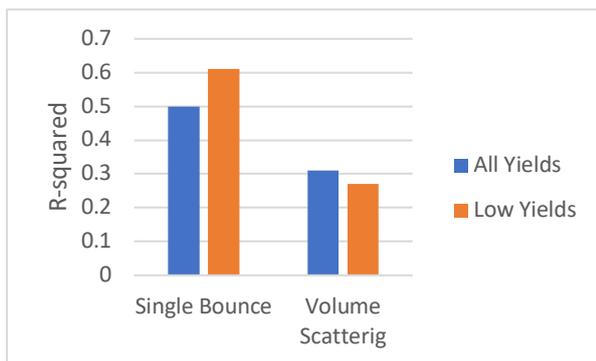
## 4. RESULTS

### 4.1 Evaluations of the Input Features

The correlation between soybean yield and SAR feature was not consistent across all ranges of yield. Correlation of determination ( $R^2$ ) between soybean yield and sum of the M-Chi single-bounce time-series from two months after planting date to one month before the crop harvest are shown in Figure 3 for the whole soybean yield range (1000- 5200 kg/ha) and the low soybean yield range (1000- 4000 kg/ha). The correlations between volume scattering and soybean yields are shown in Figure 4. While for the volume scattering, the correlation for total yield range is a bit higher than the lower yield range (i.e.  $R^2$  of 31 for the total yield range versus  $R^2$  of 27 for the lower yield range), higher correlation of 0.62 were observed for the lower yield range than the total yield range with the correlation of 0.52 for the single-bounce. Given this finding the ANN algorithm was trained for the entire range of soybean yield, as well as for the soybean yield range of 1000-4000 kg/ha using time-series of single-bounce and volume scattering from two months after planting date to one month before the harvest date for each field. These trained ANN models were then used to retrieve soybean yield. First, the ANN model was trained for the entire yield range to predict a preliminary yield value for each field. If the predicted yield was equal or less than 4000 kg/ha, a second model that was trained for this range of yield and rerun to estimate yield on these lower producing fields. If the initial predictions were above 4000 kg/ha, this estimate was considered final.



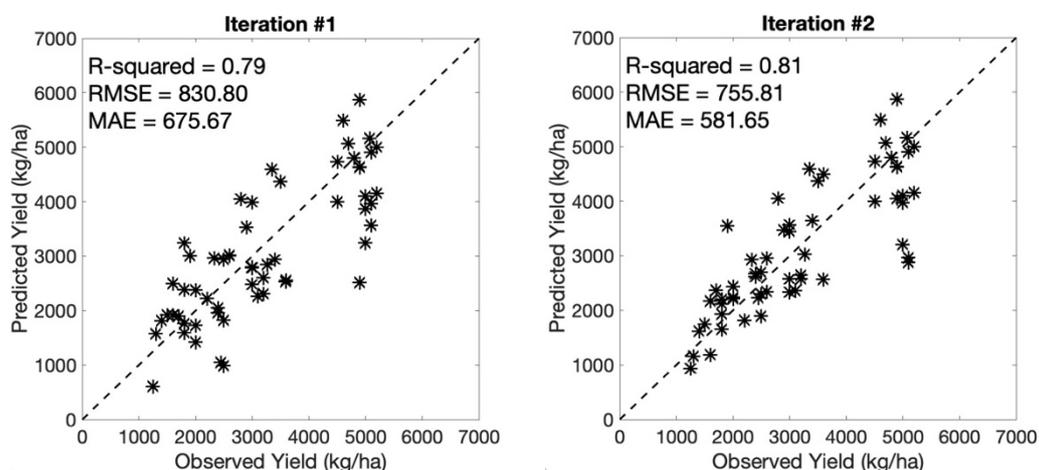
**Figure 3.** Correlation between single-bounce and soybean yield for (a) all yield values and (b) only low yield values.



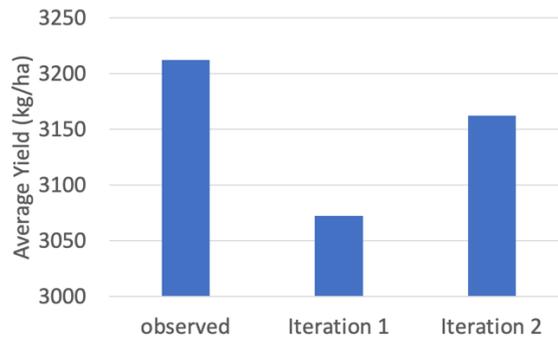
**Figure 4.** Correlation of determination between the SAR features and soybean yield.

#### 4.2 Validation of the ANN Model

The trained ANN algorithms were used for a 5-fold cross-validation. Using an independent dataset for the validation will make sure of not having the overfitting issue that might occur when using the Machine Learning algorithms. The accuracies of yield forecast are shown in Figure 5 for iterations 1 and 2. Using the ANN model trained for the whole yield range,  $R^2$  of 0.79, RMSE of 830.80 kg/ha and MAE of 675.67 kg/ha were derived. However, in the second iteration and by using the ANN model trained for the low yield range, the accuracies significantly improved to  $R^2$  of 0.81, RMSE of 755.81 kg/ha and MAE of 581.65 kg/ha. For many fields, after applying iteration 2, their RMSE were significantly increased. Figure 6 shows the average of the yield values for the 58 fields. It shows the average estimated yields are 3072 kg/ha and 3162 kg/ha from iterations 1 and 2, respectively. The observed average yield is 3212 kg/ha which shows 140 kg/ha (i.e. 4.4%) error for iteration 1 and 50 kg/ha (i.e. 1.5%) error for iteration 2. It shows that at the lower scale, the error reduced by 2.9% after applying the second iteration.



**Figure 5.** Accuracies of yield forecast for the first and second iterations.



**Figure 6.** Average yield values of all the fields derived from iterations 1 and 2 and the ground observations.

## 5. CONCLUSIONS

In this study, the M-Chi decomposition was applied to the dual-polarimetric Sentinel-1 data and the extracted parameters were used and tested for soybean yield forecast. Three M-Chi decomposition parameters including single-bounce, double-bounce and volume scattering were analyzed. Time-series of the single-bounce and volume scattering parameters for each field were used to train the ANN model. A new method was developed and proposed to improve the yield forecast accuracies using an iterative retrieval algorithm. The results showed that the Correlation of Determination ( $R^2$ ) improved by 2%, Root Mean Square Error (RMSE) of yield forecast decreased by about 75 kg/ha and the Mean Absolute Error (MAE) decreased by about 106 kg/ha by using the proposed iterative algorithm at field scale. When the estimated yields were upscaled to the whole region, the difference between the two iterations was about 2.9%. All the experiments showed a high potential of SAR data for soybean yield estimation at both field scale and regional scale.

## ACKNOWLEDGEMENT

This study was funded by NASA-Harvest program led by the Department of Geographical Sciences, University of Maryland, College Park, USA.

## REFERENCES

- Ahmad, I., Saeed, U., Fahad, M. et al., 2018, Yield Forecasting of Spring Maize Using Remote Sensing and Crop Modeling in Faisalabad-Punjab Pakistan, *Journal of Indian Society of Remote Sensing*, 46, 1701–1711.
- Burke, M. and Lobell, D.B., 2017, Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proceedings of the National Academy of Sciences of the United States of America*, 114(9), 2189–2194.
- Donohue, R. J., Lawes, R. A., Mata, G., Gobbett, D., & Ouzman, J., 2018, Towards a national, remote-sensing-based model for predicting field-scale crop yield. *Field Crops Research*, 227, 79–90.
- Franch, B., Vermote, E. F., Skakun, S., Roger, J. C., Becker-Reshef, I., Murphy, E., and Justice, C., 2019, Remote sensing based yield monitoring: Application to winter wheat in United States and Ukraine, 76, 112–127.
- Homayouni S., McNairn H., Hosseini M., Xiang J., Jarret P., 2019, Quad and compact multitemporal C-band PolSAR observations for crop characterization and monitoring, *International Journal of Earth Observation and Geoinformation*, 74, 78–87.
- Hosseini M., McNairn H., 2017, Using multi-polarization C- and L-band synthetic aperture radar to estimate wheat fields biomass and soil moisture, *International Journal of Earth Observation and Geoinformation*, 58, 50–64.
- Kuhn M., Johnson K., 2013, *Applied predictive modeling*. New York: Springer.
- Khaki S., Pham H., Wang L., 2020, Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning, *Scientific Reports*, 11 (1), 1–14.
- Mandal D., Hosseini M., McNairn H., Kumar V., Bhattacharya A., Rao Y. S., Mitchell S., Dingle Robertson R., Davidson A., Dabrowska-Zielinska K., 2019, An investigation of inversion methodologies to retrieve the Leaf Area Index of corn from C-Band backscatter, *International Journal of Earth Observation and Geoinformation*, 82, 1–11.
- McNairn, H., Duguay, C., Brisco, B., and Pultz, T.J. (2002). The effect of soil and crop residue characteristics on polarimetric radar response. *Remote Sensing of Environment*, 80, 308–320.
- McNairn, H., Shang, J., Jiao, X., and Champagne, C. (2009). The contribution of ALOS PALSAR multi-polarization and polarimetric data to crop classification. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47, pp. 3981–3992.
- Raney, R. K., 2007, Hybrid-Polarity SAR architecture, *IEEE Trans. Geosci. Remote. Sens.*, 45(11).
- Raney, R. K., Cahill, J. T. S., Patterson, G. W., and Bussey, D. B., 2012, The M-Chi Decomposition of Hybrid Dual-Polarimetric Radar Data with Application to Lunar Craters. *Journal of Geophysical Research: Planets*, 117(E12), 1–8.
- Reisi-Gahrouei O., Homayouni S., McNairn H., Hosseini M., 2020, Estimating Crop Biomass and Leaf Area from Multi-Temporal-Multispectral Imagery and Machine Learning Approaches, *Canadian Journal of Remote Sensing*, 46(1), 84–99.
- Setiyono, T. D., Quicho, E. D., Holecz, F. H., Khan, N. I., Romuga, G., Maunahan, A., Garcia, C., et al., 2019, Rice Yield Estimation Using Synthetic Aperture Radar (SAR) and the

ORYZA Crop Growth Model: Development and Application of the System in South and South-East Asian Countries. *International Journal of Remote Sensing*, 40(21), 1–32.

Skakun, S., Vermote, E., Franch, B., Roger, J.-C., Kussul, N., Ju, J., Masek, J., 2019, Winter Wheat Yield Assessment from Landsat 8 and Sentinel-2 Data: Incorporating Surface Reflectance, Through Phenological Fitting, into Regression Yield Models. *Remote Sensing*, 11(15), 1768.

Yu, H., Wilamowski, B. M., 2011, Levenberg–Marquardt Training, *Industrial Electronics Handbook*. vol. 5 – Intelligent Systems, ed: CRC Press, 12-1–12-15.