A hyperlocal machine learning approach to estimate NDVI from SAR for agriculture fields

Ran Pelta, Ofer Beeri, Rom Tarshish, and Tal Shilo

Manna Irrigation, Gvat, Israel

A paper from the Proceedings of the
15th International Conference on Precision Agriculture
June 26 – 29, 2022
Minneapolis, Minnesota, United States

Abstract: *The normalized difference vegetation index (NDVI) is commonly used in precision agriculture. The NDVI is a proxy for crop growth, health, leaf area index, crop cover, and more. Yet, when clouds are present, the NDVI cannot be calculated. Synthetic Aperture Radar (SAR), on the other hand, can penetrate clouds but is sensitive to different crop properties than the NDVI. Several SAR vegetation indices have been suggested to estimate NDVI via SAR, however, they tend to work for limited spatial and temporal settings. This study presents a hyperlocal machine learning approach to estimate NDVI from SAR images for agriculture fields. The approach utilized time series of past NDVI and multiple SAR indices to train a machine learning model each time a new SAR image is available over each field. Consequently, the model estimates the crop NDVI value from the current SAR image. Then, when the next SAR image is available, the model will re-learn the relationship (based on past data) which might have changed, thus, the model is kept up-to-date. The suggested approach was tested on 97 fields from 12 countries with 5 crop types. RMSE, R^2, and Bias of 0.07, 0.92, and 0.00, respectively, were achieved, expressing model usefulness and global applicability. The suggested approach can ensure a constant stream of NDVI values, regardless of clouds, which is crucial in cloudy areas and at specific times during the growing season such as when crops start their development stage.*

Keywords.
NDVI,SAR,machine learning,random forest,time-series,remote sensing

Introduction

The normalized difference vegetation index (NDVI) which was introduced almost 50 years ago (Rouse et al., 1974), is still the most common index to monitor vegetation in general and specifically in agriculture. The NDVI can help monitor whether vegetation is healthy or under stress, as well as observe and detect changes in vegetation due to various reasons such as natural disturbances or changes in plants' phenological.

However, in poor illumination conditions such as in the presence of clouds, NDVI cannot reliably be calculated because the sensor reading will contain reflected light from the clouds, therefore will not represent the true condition of the vegetation. As so, the window of opportunity to calculate high-quality NDVI is limited.

On the other hand, the Synthetic Aperture Radar (SAR) is not affected by illumination conditions and can penetrate clouds, thus creating an opportunity for remote sensing of the vegetation in all weather conditions.

The Sentinel-1 satellites (A and B) carry a C-band instrument, which provides a collection of data in all-weather, day or night with global coverage, is the most widely SAR dataset available.

Therefore, several published studies have attempted to estimate NDVI with Sentinel-1 data using the VV and VH bands. This was done by correlating SAR indices or SAR backscatter (the VV and VH bands) to NDVI (Frison et al., 2018; Holtgrave et al., 2020; Kaushik et al., 2022; Navarro et al., 2016; Veloso et al., 2017) or by using mathematical models to find the relationship between SAR and NDVI (Filgueiras et al., 2019; Mazza et al., 2018; Mohite et al., 2020).

Yet, indices or models with pre-defined coefficients are applicability confined, either because they were tested on a small number of crops, small spatial extent, or were developed for specific crops. Therefore, previous studies found different correlation strengths with different crops, phenological, NDVI values, soil types, etc., leading to the conclusion that none of the indices or models will work well when tested on a variety of crops, soils, or local conditions.

From a practical point of view, a method to estimate NDVI from SAR should be like NDVI in terms of global applicability, regardless of crop type or local conditions.

The approach suggested in this study seeks to provide such a solution as it makes use of multiple SAR indices simultaneously as well as producing a model per field, per point in time. By doing that, the suggested approach is able to account for the local conditions and changes of the field, therefore, keeping the model and its coefficients up to date based on the most relevant data.

Materials and Methods

Study sites

97 commercial fields from 12 countries with 5 crop types were selected for this study. The source of the study sites is the Manna Irrigation database (not publicly available). Table 1 shows the crops and the corresponding number of fields per crop used here.

Table 1. The number of fields per crop that were used in this study

	Crop Name	Number of fields
1	Avocado	20
2	Almonds	19
3	Cotton	20

4	Watermelon	18
5	Alfalfa	20

Remote Sensing dataset

The remote sensing dataset used in this study was obtained with Google Earth Engine (GEE) (Gorelick et al., 2017) Python API. NDVI was calculated via the NIR and red spectral bands of Sentinel-2 and Landsat-8, both processed to Level-2, surface reflectance. The NDVI was averaged per image per field thus generating NDVI times series for each field. Images with clouds, cirrus, or cloud shadows were removed from further analysis.

The SAR (i.e., Sentinel-1) time series (per field) were also obtained with GEE in the form of Interferometric Wide Swath Mode (IW) with dual polarization (VV+VH) and were acquired under level-1 processing as ground range detected (GRD).

Data processing and NDVI estimation from SAR

The data preprocessing and NDVI estimation from SAR are based on past NDVI and SAR time series. The data processing begins when NDVI is not available, but Sentinel-1 image is. The first step is to obtain NDVI and SAR indices (Table 2) time series for the last 365 days. The second step is to apply, per time series, a locally weighted regression algorithm (Atkeson et al., 1997) followed by daily interpolation with the assumption that changes in crop growth are gradual during short periods (i.e., between two consecutive images) (Fieuzal et al., 2013). Then, the time series is used for model training, where the NDVI is the dependent variable, and the SAR indices time series are the independent variables. The machine learning model used here is the random forest (RF) (Pedregosa et al., 2011). Once the model is trained, the Sentinel-1 data from the current image is used for the model estimation of NDVI. This process was executed for each available Sentinel-1 image for each of the 97 fields across 2021.

Table 2. The SAR indices used in this study

	Name	Full name	Formula	Source
1	PRVI	Polarimetric Radar Vegetation Index	$(1-\frac{VV}{VH+VV})*VH$	Chang et al., 2018
2	RFDI	Radar Forest Degradation Index	$\frac{VV-VH}{VH+VV}$	Flores et al., 2019
3	RVI	Radar Vegetation Index	$\frac{4*VH}{VH+VV}$	Trudel et al., 2012
4	SNI	Sentinel normalized index	$\frac{VH-VV}{VH+VV}$	Filgueiras et al., 2019
5	VH_VV_ratio	VH to VV ratio	$\frac{VH}{VV}$	Veloso et al., 2017
6	VV_VH_ratio	VV to VH ratio	$\frac{VV}{VH}$	Frison et al., 2018

Results and Discussion

Fig. 1 shows the overall accuracy of the model for the entire dataset, for matching dates between

NDVI and Sentinel-1, while Table 3 displays the results per crop.

Accuracy metrics of RMSE, R^2, and Bias of 0.07, 0.92, and 0.00 (Fig. 1), respectively express model robustness and global applicability.

Orchards (almonds and avocado) achieved better results than field crops (alfalfa, watermelon, and cotton). This can be attributed to the fact that orchard NDVI changes less than field crops during the growing season.

The crop with the lowest R^2 is alfalfa which can be explained by its short phenological cycles (~28 days), which might not be well captured by the NDVI due to a lack of sufficient images. Perhaps with higher NDVI frequency for model training, the results can be improved.

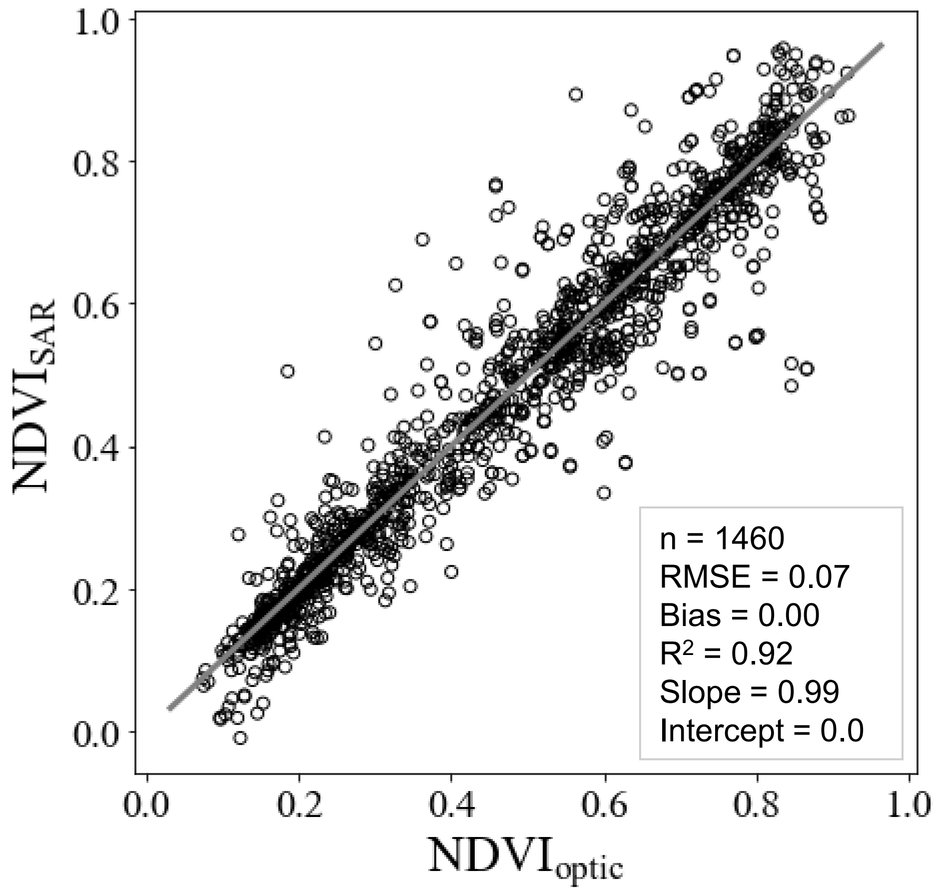

Fig 1. Overall model performance

Table 3. Model performance per crop

	Crop	n	RMSE	Bias	R^2
1	Almonds	375	0.03	0.00	0.96
2	Avocado	275	0.04	0.00	0.96
3	Cotton	189	0.05	-0.01	0.95
4	Watermelon	247	0.06	-0.01	0.91
5	Alfalfa	374	0.10	0.02	0.76

The relatively large and diverse dataset used in this study coupled with high performance validates the suggested approach.

Concluding from previous studies finding, the relationship between NDVI and SAR is not consistent across various crops, soil types, or phenological stages. Therefore, instead of using a model with fixed coefficients or adhering to a specific index, the suggested approach generates a new model per field, per point in time (i.e., per new available Sentinel-1 image), based on the most up-to-date NDVI-SAR field-specific relationship.

Hence, the suggested approach is always up-to-date, dynamic, and flexible enough to account for the hyperlocal changes in the field, thus achieving good results across various crops and locations.

Conclusions

This study proposed an approach to estimate NDVI from SAR (Sentinel-1) images for agricultural fields. As opposed to previous studies, the approach here did not focus on a specific SAR index, but rather use multiple SAR indices to estimate NDVI. A machine learning model (random forest) was employed to find the best combination of SAR indices per field, per point in time in order to estimate NDVI. This approach was validated with 97 commercial fields from 12 countries with 5 different crop types. High accuracy metrics disclose the applicability and usefulness of this approach to cases where cloud cover or low NDVI frequency are present. Consequently, ensure high frequent NDVI for agricultural fields which can assist with daily decision-making.

References

Atkeson, C. G., Moore, A. W., & Schaal, S. (1997). Locally Weighted Learning. In D. W. Aha (Ed.), *Lazy Learning* (pp. 11–73). Springer Netherlands. https://doi.org/10.1007/978-94-017-2053-3_2

Chang, J. G., Shoshany, M., & Oh, Y. (2018). Polarimetric Radar Vegetation Index for Biomass Estimation in Desert Fringe Ecosystems. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(12), 7102–7108. https://doi.org/10.1109/TGRS.2018.2848285

Fieuzal, R., Baup, F., & Marais-Sicre, C. (2013). Monitoring Wheat and Rapeseed by Using Synchronous Optical and Radar Satellite Data—From Temporal Signatures to Crop Parameters Estimation. *Advances in Remote Sensing*, *2*(2), 162–180. https://doi.org/10.4236/ars.2013.22020

Filgueiras, R., Mantovani, E. C., Althoff, D., Fernandes Filho, E. I., & Cunha, F. F. da. (2019). Crop NDVI Monitoring Based on Sentinel 1. *Remote Sensing*, *11*(12), 1441. https://doi.org/10.3390/rs11121441

Flores, A., Herndon, K., Thapa, R., & Cherrington, E. (2019). *The SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*. https://doi.org/10.25966/nr2c-s697

Frison, P.-L., Fruneau, B., Kmiha, S., Soudani, K., Dufrêne, E., Le Toan, T., Koleck, T., Villard, L., Mougin, E., & Rudant, J.-P. (2018). Potential of Sentinel-1 Data for Monitoring Temperate Mixed Forest Phenology. *Remote Sensing*, *10*(12), 2049. https://doi.org/10.3390/rs10122049

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Holtgrave, A.-K., Röder, N., Ackermann, A., Erasmi, S., & Kleinschmit, B. (2020). Comparing Sentinel-1 and -2 Data and Indices for Agricultural Land Use Monitoring. *Remote Sensing*, *12*(18), 2919. https://doi.org/10.3390/rs12182919

Kaushik, S. K., Mishra, V. N., Punia, M., Diwate, P., Sivasankar, T., & Soni, A. K. (2022). Crop Health Assessment Using Sentinel-1 SAR Time Series Data in a Part of Central India. *Remote Sensing in Earth Systems Sciences*. https://doi.org/10.1007/s41976-021-00064-z

Mazza, A., Gargiulo, M., Scarpa, G., & Gaetano, R. (2018). Estimating the NDVI from SAR by Convolutional Neural Networks. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 1954–1957. https://doi.org/10.1109/IGARSS.2018.8519459

Mohite, J. D., Sawant, S. A., Pandit, A., & Pappula, S. (2020). Investigating the Performance of Random Forest and Support Vector Regression for Estimation of Cloud-Free Ndvi Using SENTINEL-1 SAR Data. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43B3*, 1379–1383. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-1379-2020

Navarro, A., Rolim, J., Miguel, I., Catalão, J., Silva, J., Painho, M., & Vekerdy, Z. (2016). Crop Monitoring Based on SPOT-5 Take-5 and Sentinel-1A Data for the Estimation of Crop Water Requirements. *Remote Sensing*, *8*(6), 525. https://doi.org/10.3390/rs8060525

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, *12*(null), 2825–2830.

Rouse, J. W., Jr., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring Vegetation Systems in the Great Plains with Erts. *NASA Special Publication*, *351*, 309.

Trudel, M., Charbonneau, F., & Leconte, R. (2012). Using RADARSAT-2 polarimetric and ENVISAT-ASAR dual-polarization data for estimating soil moisture over agricultural fields. *Canadian Journal of Remote Sensing*, *38*(4), 514–527. https://doi.org/10.5589/m12-043

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F., & Ceschia, E. (2017). Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sensing of Environment*, *199*, 415–426. https://doi.org/10.1016/j.rse.2017.07.015