Modeling SAR Observables by Combining a Crop-Growth Model With Machine Learning

Nikaein, Tina; Lopez-Dekker, Paco; Steele-Dunne, Susan; Kumar, Vineet; Huber, Manuel

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Modeling SAR Observables by Combining a Crop-Growth Model With Machine Learning

Tina Nikaein , *Member, IEEE*, Paco Lopez-Dekker , *Senior Member, IEEE*,
Susan C. Steele-Dunne , *Member, IEEE*, Vineet Kumar , *Member, IEEE*, and Manuel Huber

Abstract—In this article, our aim is to estimate synthetic aperture radar (SAR) observables, such as backscatter in VV and VH polarizations, as well as the VH/VV ratio, cross ratio, and interferometric coherence in VV, from agricultural fields. In this study, we use the decision support system for agrotechnology transfer (DSSAT) crop-growth simulation model to simulate parcel-level phenological and growth parameters for over 1500 parcels of silage maize in the Netherlands. The crop model was calibrated using field data, including silage maize phenological phases, leaf area index, and above-ground dry biomass (AGB). The simulations incorporate fine-resolution gridded precipitation data and soil parameters to model the interaction between soil–plant–atmosphere and genotype in DSSAT. The crop variables produced by DSSAT are then used as inputs to a support vector regression model. This model is trained to simulate SAR observables in 2017, 2018, and 2019, and its performance is evaluated using independent fields in each of these years. The results show a close fit between modeled and observed SAR C-band observables. The importance of vegetation variables in the estimation of SAR observables is assessed. The AGB showed significant importance in the estimation of backscatter. This study demonstrates the potential value of combining crop-growth simulation models and machine learning to simulate SAR observables. For example, the SVR model developed here could be used as an observation operator in an assimilation context to constrain vegetation and soil water dynamics in a crop-growth model.

Index Terms—Crop, decision support system for agrotechnology transfer (DSSAT), forward model, synthetic aperture radar (SAR), Sentinel-1, silage maize, simulation.

I. INTRODUCTION

THE launch of the Sentinel-1 synthetic aperture radar (SAR) constellation [1] has provided unprecedented opportunities for agricultural monitoring. This is due to its dense time series of radar scattering and interferometric coherence in C-band, which offers a short revisit time of 6 days in Europe and 12 days globally. Additionally, the dataset is freely accessible. Satellite observations from active and passive microwave sensors are sensitive to vegetation structure [2] and water content [3].

Space-borne SAR observations have been used for many years for agricultural applications and vegetation monitoring because of microwave signal sensitivity to changes during the crop-growth period [4], [5]. Several studies highlighted the potential of Sentinel-1 SAR data for crop-growth monitoring [6], [7], [8], crop water content estimation and soil moisture mapping [9], [10], and parameter retrieval [11]. Vreugdenhil et al. [12] demonstrated using in situ observations that C-band backscatter from Sentinel-1 is sensitive to vegetation parameters such as vegetation water content (VWC), biomass, height, and leaf area index (LAI). In addition to monitoring dynamics directly, Sentinel-1 data could be assimilated into a crop-growth model to constrain the estimates of biogeophysical variables. The assimilation of Earth observation (EO) data in agricultural models has been demonstrated in several studies. For example, the assimilation of LAI and soil moisture derived from Sentinel-1 and Sentinel-2 data into the Word Food Studies (WOFOST) model to estimate crop yield was studied in [13]. Assimilating LAI and dry biomass from optical and SAR data into a model to estimate soybean yield was studied in [14]. In another study [15], they assimilated LAI from SAR product to decision support system for agrotechnology transfer (DSSAT) for rice yield estimation. These studies collectively demonstrate that the assimilation of EO data or products can be used to improve yield estimates. In addition, assimilation could provide improved estimates of the growth and development of the crop to support agricultural management and decision making.

To integrate SAR observations with modeled soil moisture and vegetation, two approaches can be employed. The first approach involves converting the SAR signal into retrievals of geophysical variables using change detection algorithms [16] or other methods [17]. However, this approach is limited by the availability of high-resolution retrievals from the C-band with global coverage. The second approach focuses on estimating and combining the backscatter components of soil moisture and vegetation to simulate the expected signal at the sensor level using a backscatter model as a forward model. This study specifically adopts the second approach, utilizing a forward operator to convert model simulations into a backscatter signal. Our aim is to assimilate Sentinel-1 observables directly rather than retrieved parameters. This approach allows us to utilize

all the information related to the incidence angle dependence while avoiding any potential for cross-correlated errors between retrievals and model simulations [18]. The objective is to prepare for a data assimilation system, where SAR observations will be utilized to update crop model simulations of soil moisture and biomass [19].

To exploit Sentinel-1 backscatter and coherence for agricultural monitoring, we need to be able to model the observables given a description of the soil and vegetation. Furthermore, it is essential to be able to understand and quantify the effect of crop biogeophysical variables on SAR observables. The most common way to model radar observables is using a model, the complexity and requirements of which can vary considerably. The most widely used model is the water cloud model (WCM). The WCM is a semiempirical parameterized model to relate the normalized radar cross section (NRCS) to the characteristics of the vegetation and the surface. The model is relatively simple, as it models the vegetation as a collection of identical water droplets randomly distributed within the canopy [20], [21]. As a recent example, Rains et al. [22] used the WCM as a measurement operator to assimilate Sentinel-1 data into the global land evaporation Amsterdam model (GLEAM). More sophisticated models, such as Michigan microwave canopy scattering (MIMICS) [23], can also be used to simulate microwave observables [24]. However, they require descriptions of vegetation geometry, architecture, and dielectric properties that are seldom available. An alternative to using the WCM as a forward operator is employing machine-learning techniques. Machine learning offers advantages due to its flexibility, adaptability, and ability to handle nonlinear relationships. It can extract meaningful insights directly from data and generalize well to new instances. It is important to note that they also have limitations. They may require large amounts of high-quality labeled data for training.

This study aims to provide a model capable of simulating Sentinel-1 observables that require readily available crop descriptors. We propose circumventing the limited availability of in situ data by using a crop-growth model instead. This is a significant step toward developing a robust and reliable system for assimilating high-resolution Sentinel-1 observables over vegetation areas. Davitt et al. [24] demonstrated that DSSAT could be used to provide a description of the growing crop to be used as input to MIMICS. Here, instead, we will use this description as input to a machine-learning model to map the crop descriptors to the SAR observables. The biophysical parameters generated by the crop-growth model reflect the state of the crop at a given time, which reflects all the past inputs to the model. The radar observables depend only on the state of the crop and the soil at the time the data are acquired. Our approach, of using modeled land surface states as input to a machine-learning model has been used to map snow and land parameters to brightness temperature [25], to map land surface variables to advanced scatterometer backscatter, slope, and curvature [26], and recently, to map daily 1-km AquaCrop model biomass and surface soil moisture to backscatter [27].

The objectives of this study are as follows.
1) Demonstrate that a crop-growth model can be used together with a machine-learning model to simulate SAR
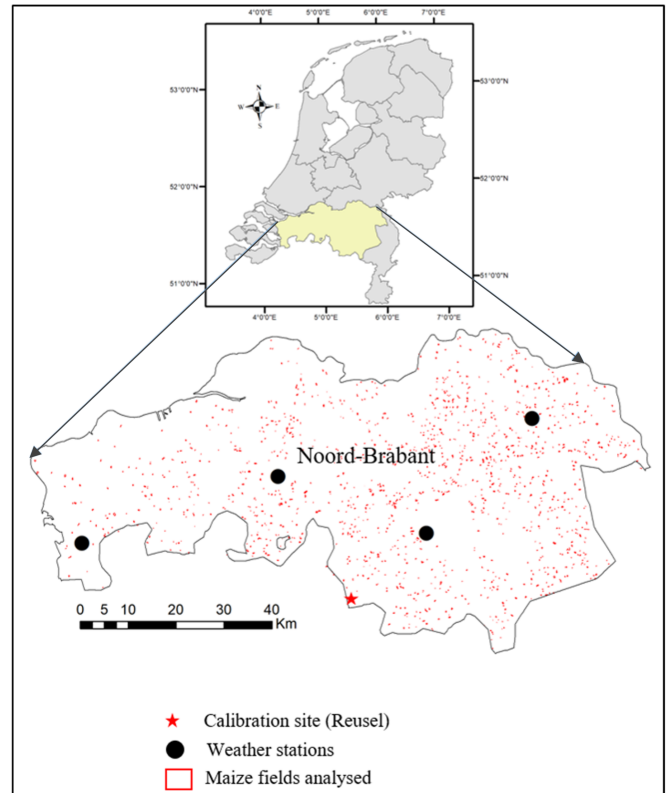
Fig. 1. Location of the study area. The maize fields analyzed in this study are shown in red. Black circles display the location of four automatic meteorological stations over the province. The red star indicates the location of the site used for calibration.

observables to optimize the future assimilation of SAR observables into the regional DSSAT model.
2) Show that the relationships between crop biogeophysical variables and modeled SAR observables are physically plausible.
3) Discuss/identify the potential limitations of applying this technique in agricultural applications.

The analysis will be conducted over 1500 maize fields in the Netherlands. The DSSAT model will be used to simulate a description of the growing vegetation in terms of LAI, AGB, surface soil moisture (SM_S), and root zone soil moisture (SM_R). To establish the link between the observables and the biophysical parameters, we use a support vector regression (SVR) model. Feature analysis, the minimum redundancy maximum relevance (MRMR), will be used to examine the sensitivity of the SAR observables to the DSSAT variables.

II. STUDY AREA

The study is performed in the Noord-Brabant province (5081 km^2) in the South of The Netherlands. Province boundaries and location of agricultural fields with their associated crop types were retrieved from the Basisregistratie Gewaspercelen (BRP) [28], an open national database of crop parcel boundaries and crop type in The Netherlands. The average temperature in this area varies between 2 °C and 24 °C and the average annual

Fig. 2. DSSAT-SVR SAR observables simulation workflow.
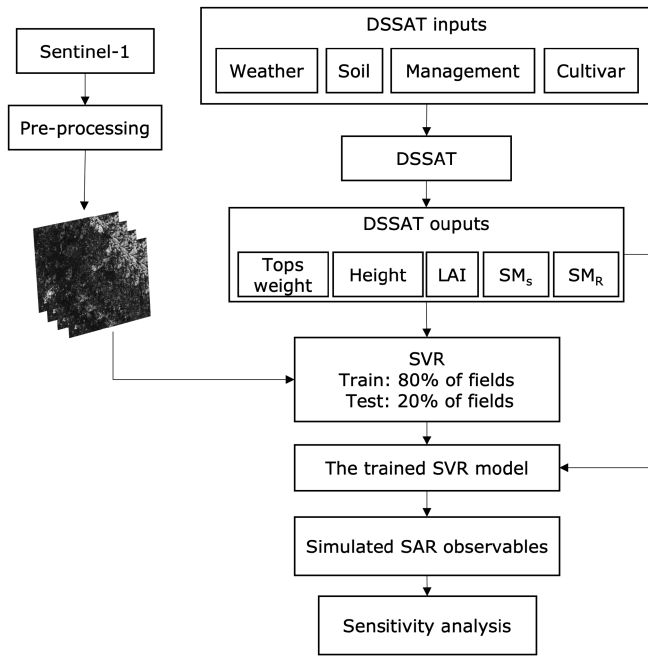
precipitation is around 646 mm. Mean monthly sunshine hours range from 5 to 14 h [29].

The study area location and the spatial distribution of the maize fields are shown in Fig. 1. Almost all maize grown by farmers in The Netherlands is silage maize. It is planted between mid-April and the beginning of May, and the emergence is in mid-May. It is left to ripen in the field, and is harvested in September [30]. Silage maize is grown in approximately 20% (about 20000#) of parcels in Noord-Brabant. 10% of these fields were used for this study, ensuring that they were randomly spread all over the province (see Fig. 1).

For calibration, data were available at an experimental site in Reusel, 51.319°N, 5.173°E, in 2019 [31]. The field is on sandy soil and can be irrigated with a gun sprinkler irrigation system. In 2019, the maize was watered twice during the summer, relying on the rain during the rest of the growth period.

III. DATA AND METHODOLOGY

Fig. 2 provides an overview of the workflow followed in this study.

A. Decision Support System for Agrotechnology Transfer (DSSAT)

Crop-growth models, such as the DSSAT model, simulate the interaction between plants and their environment, in daily steps, to estimate the growth, development, and yield of different crops [32], [33]. Basically, they are scientific tools with a set of equations that take environmental information as inputs and determine the phenological development, and growth stage as outputs and these require precise parameterization. DSSAT has been used for years for different crops and regions across the world [34]. The CERES-Maize module (Crop Environment Resource Synthesis) [35] is part of DSSAT v4.7, and is one of the most broadly used maize models. The CERES-Maize module has been used and evaluated in several studies in different parts of the world. For instance, in [36], DSSAT performance was validated, and then used to determine the best management practices of nitrogen fertilization and irrigation. In another study, the impact of climate variability on maize in the semiarid area by CERES-Maize module was assessed [37].

B. Input Data for DSSAT

DSSAT requires a minimum set of input data to simulate crop growth: daily weather information during the growth period; soil characteristics of the area; crop management data (sowing and harvest date, row spacing, irrigation, and fertilization information); and cultivar coefficients.

1) Weather Data: The required weather data including daily solar radiation (SRAD), precipitation, and maximum and minimum air temperature (TMAX and TMIN), were obtained from Royal Dutch Meteorological Institute weather station data [29]. There are over 40 weather stations in The Netherlands, of which four are within the Noord-Brabant province. The point-wise meteorological data was interpolated to raster format using inverse distance weighting spatial interpolation.

Precipitation is a key parameter. So, as an alternative to interpolated gauge data at a limited number of locations, the daily precipitation sum data in gridded format measured on approximately 300 locations of a voluntary network over The Netherlands have been used to obtain precipitation data with higher spatial resolution [38]. The gridded data are not available for all the other features. Daily weather files in the required DSSAT file format were generated for each crop parcel, and each year.

2) Soil Data: Bulk density, soil organic carbon, sand, silt, clay fractions, coarse fragments, soil pH, and total nitrogen were obtained from the ISRIC's global Soil Information System (SoilGrids) [39]. Soil data at six different depth layers 0–5, 5–10, 15–30, 30–60, 60–100, and 100–200 cm were obtained with a 250-m grid spacing from SoilGrids250 m [40].

Soil properties required by DSSAT that are not available in SoilGrids250 m, are obtained from HarvestChoice HC27 [41]. HC27 is a soil database containing 27 soil profiles that are provided by considering only three criteria: soil texture, organic carbon content (proxy for soil fertility), and rooting depth (water availability proxy). In this study, values for soil color, albedo, evaporation limit, drainage rate, runoff curve number, mineralization factor, photosynthesis factor, pH in buffer determination method, extractable phosphorus determination code, and potassium determination method were obtained from the corresponding HC27 soil profiles.

We calculated the values for saturated hydraulic conductivity, saturation, drained upper limit, and lower limit by pedo-transfer functions following [42], [43], and [44].

Table I lists the most relevant soil parameters and the range of values assumed in this study.

TABLE I
SOIL PARAMETERS FOR DSSAT

Parameters	Range
Depth (cm)	5–200
Porosity (cm^3/cm^3)	0.376–0.417
Field capacity (cm^3/cm^3)	0.155–0.187
Wilting point (cm^3/cm^3)	0.069–0.097
Bulk density(g/cm^3)	1.11–1.51
Organic carbon (%)	1.33–6.55
Total nitrogen (%)	0.16–0.48

TABLE II
MAIZE INPUT MANAGEMENT PARAMETERS INTO DSSAT MODEL FROM
NOORD-BRABANT PROVINCE

Parameters	Value
Density ($\#/m^2$)	8
Row spacing (cm)	75
Depth (cm)	12
Planting date	Beginning of May
Flowering date	End of July
Physiological maturity date	End of August
Harvest date	End of September

3) Management and Genetics: Table II provides details on planting, emergence, harvest dates, and planting density assumed in this study. In Noord-Brabant, they prepare the field in April, the growing season of maize starts in May and lasts until the end of September. During the simulation, water stress simulation is enabled but all the other nutrient stresses are turned OFF as all the crops growing over the area, are well fertilized. The planting density was measured by [31], and these are typical values for maize planted in this area.

The DSSAT model requires some calibration. Specifically, we need to define a set of genetic coefficients controlling the growth, development, and yield of the crop such as ecotype and cultivar coefficients [35]. Ecotype coefficients are a set of coefficients for a group of cultivars (cultivar means a type of cultivated crop) that show similar responses to environmental conditions. Genetic coefficients in DSSAT have been calibrated for maize grown in tropical and semiarid areas [45]. However, the day length, radiation use efficiency, and temperature mean that different kinds of maize are grown in Northern Europe. Furthermore, the maize grown in the Netherlands is grown for silage. Therefore, some ecotype coefficients, such as radiation use efficiency (RUE) [46] and canopy light extinction coefficient for daily photosynthetically active radiation (PAR) [47], [48], have to be tuned. These are important drivers controlling the maximum value of the growth parameters [49].

The *medium season* cultivar, was selected as an initial maize cultivar in the area and the calibration started based on those primary values. The cultivar coefficients are calibrated by comparison between observed and simulated variables. First, we calibrate the phenology coefficients (P1, P2, and P5 in Table III) by finding the values that provide the best agreement between simulated and observed flowering and physiological maturity dates. Then, we calibrate the growth coefficients (G2 and G3) in order to minimize the root mean square difference between

TABLE III
DSSAT MAIZE CULTIVAR COEFFICIENTS DEFINITION AND CALIBRATED
VALUES

Cultivar parameters	Value
P1: Thermal time from emergence to end of juvenile phase.	140
P2: Photoperiod sensitivity coefficient.	0.856
P5: Thermal time from silking to physiological maturity.	890.7
G2: Potential kernel number.	983.3
G3: Potential kernel growth rate (mg/day).	7.09
PHINT: Phylochron interval; the interval in thermal time (degree days) between successive leaf tip appearances.	38.9

observed and simulated LAI and dry biomass for the calibration field. Calibration data were available at an experimental site in Reusel [31]. Table III shows the calibrated cultivar coefficients for CERES-Maize.

The key biophysical parameters of maize, like LAI, canopy height, AGB, SM_S, and SM_R provided by DSSAT, are entered as inputs to the SVR forward model to simulate the SAR observables.

C. SAR and Optical Data

Sentinel-1 C-band Interferometric Wide swath data with a 6-day repeat cycle (relative orbit 37) have been used. For this relative orbit, the incidence angle range varies between 36 to 40° over the study area. We used the spatially averaged parcel-level NRCS in VV and VH polarizations (σ_{VV}^0 and σ_{VH}^0), the cross ratio ($CR = \sigma_{VH}^0/\sigma_{VV}^0$), and the interferometric coherence in VV polarization. All these radar observables were extracted from the Agricultural SandboxNL database [50], [51]. This database was generated by utilizing the openly available annual BRP vector layers and Sentinel-1 SAR ground range detected (GRD) data over The Netherlands in the Google Earth Engine (GEE) using data collected between 2017 and 2019. The GEE-provided Sentinel-1 GRD images are preprocessed with orbit file update, radiometric calibration, border, thermal noise correction, and terrain correction [52]. The Agricultural SandboxNL database contains parcel averaged Sentinel-1 backscatter (σ_{VV}^0, σ_{VH}^0, and CR) and associated attributes (local incidence angle, azimuth look angle, and pixel count) for six different relative orbits (37, 110, 139, 15, 88, and 161).

The interferometric coherence is defined as the normalized cross-correlation between two coregistered SAR images. The coherence is defined as [53]

$$\gamma_{est} = \frac{\langle S_1 S_2^* \rangle}{\sqrt{\langle |S_1|^2 \rangle \langle |S_2|^2 \rangle}} \tag{1}$$

where S_1 and S_2 represent two coregistered complex images, and $*$ denotes the complex conjugate operator, and $< \cdot >$ represents the spatial averaging operator. We have estimated the coherence values implementing standard InSAR preprocessing steps [54] in ESA SNAP software [55] for VV polarization with a 6-day repeat cycle. The parcel-level spatially averaged coherence product is generated similarly to the Sentinel-1 SAR backscatter in the Agricultural SandboxNL database.

The coherence values range from 0 to 1, where low values correspond to high decorrelation between the two acquisitions.

High coherence is achieved when the physical properties and position of scatterers remain the same between two acquisitions. This happens during the bare soil or after harvest, which makes the coherence a valuable indicator to detect agricultural events.

The SandboxNL was extended to include parcel-level averaged Sentinel-2 optical data. In this study, we used these optical data to obtain estimates of LAI over our study area. The parcel-level LAI values for maize were estimated from the normalized difference vegetation index (NDVI) using the relationship for maize adopted in [56] and [57].

D. Support Vector Regression (SVR)

To model the relation between the biophysical parameters and SAR observables (NRCS and coherence), we used an SVR [58] algorithm. SVR has been employed in previous studies for different applications. For example, in [22], the performance of the WCM and SVR as a forward model was compared, in order to assimilate Sentinel-1 data into the GLEAM. Their results show the capability of machine learning as an alternative to semiempirical models to predict backscatter. Additionally, soil moisture was retrieved by Ezzahar et al. [59] through the inversion of both the theoretical integral equation model (IEM) and the semiempirical model (Oh), and the results were compared with SVR. They showed that the data-driven machine-learning approach outperforms the other mentioned models. As mentioned in [60], SVR has limited complexity in the training phase and produces high accuracy with the less computational load. The purpose of this article is not to compare machine-learning algorithms nor to determine the best algorithm. In this study, we chose to use a proven machine-learning algorithm as a data-driven forward modeling technique to predict SAR observables based on its positive track record in similar applications [22], [25].

The SVR model is fed with LAI, dry biomass, SM_S, and SM_R from the DSSAT model to predict NRCS in VV and VH polarizations, and the CR. In the case of coherence, canopy height from DSSAT is also added to the inputs, and SVR is fed with the values corresponding to the dates of the first acquisition in the interferometric pair and with the differences between the values of the first and second acquisitions.

We trained and tested the SVR model in the following three ways:

1) training with data of individual years and testing on the same year (same-year);
2) training on individual years and testing on multiple years (cross-year);
3) training with multiple-year data and testing on multiple years (multiyear).

In all cases, we follow standard practices by first training the algorithm to find the best model fit, and then, testing the model on the separate test datasets. In cases 1 and 3, the algorithm is trained using 80% of the fields, and the remaining 20% fields are used for test of the model, with the fields being randomly assigned to the training and testing sets.

Each field at each Sentinel-1 acquisition time counts as an individual sample. During the growth period of each year (planting to harvest), 25 SAR acquisitions are available. We applied tenfold cross-validation to avoid overfitting. It means that the model is trained using nine of the folds and validated on the remaining part. Still, the 20% test data are kept and used exclusively for the final assessment. Grid-search is used in order to obtain the hyperparameters with the highest cross-validation accuracy. The cost parameter C and the hyperparameter γ have to be tuned. The best (C, γ) values are used to train and generate the model. The values of all input variables are scaled so that they are between 0 and 1. The optimum kernel was the radial basis function kernel, which is defined as

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \tag{2}$$

where x_i and x_j are the a pair of samples.

E. Surface Roughness

Rough surface scattering contributes to the observed radar signal. This contribution is dominant at the beginning of the growth period, when fields are mostly bare or covered with a small amount of vegetation [21]. Rough surface scattering is controlled by the dielectric constant, which in turn depends on the soil composition and the soil moisture content, and on the roughness spectrum. The latter is not represented in any of the DSSAT variables, which implies that field-to-field variations of the rough surface scattering term cannot be modeled with the available biophysical parameters. This contributes to interfield differences in the observed NRCS values during the planting and emergence period.

Conceptually, we can assume that the rough surface contribution to the radar observables can be directly measured in the period between planting and emergence. Therefore, we assume that the properties controlling this rough surface component, with the exception of the soil moisture, remain constant during the growth season (in particular during the initial period).

We consider a reference backscatter value for each parcel as a proxy for the effect of roughness/geometry on the variability between parcels and included that as a label for each parcel (along with other parameters for the parcel). This reference backscatter value is calculated as the mean NRCS value for the parcel in the three acquisitions following the planting date.

F. Evaluation of Model Performance

Five error metrics are used to evaluate the performance of the regression model. The simulated and observed backscatter and coherence values are compared using standard statistical metrics as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3}$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4}$$

and

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{5}$$
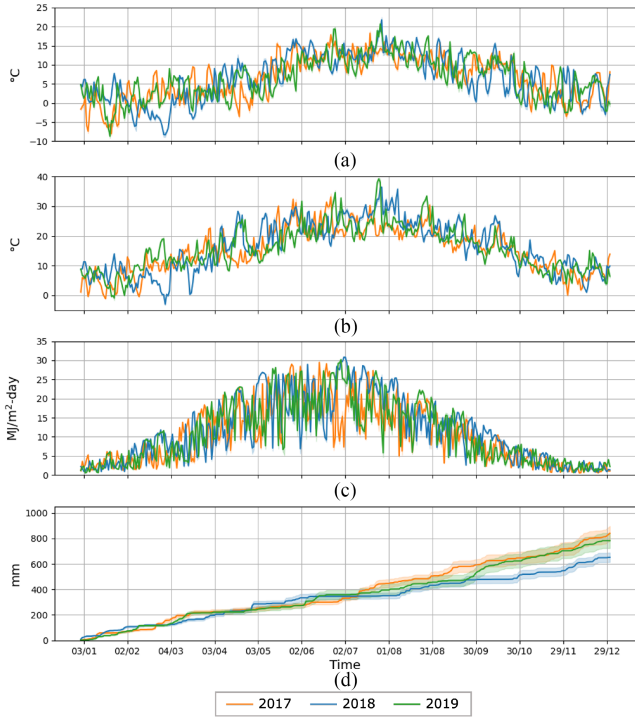
Fig. 3.   Daily time-series of meteorological forcing inputs to the DSSAT model. (a) Minimum temperature. (b) Maximum temperature. (c) Solar radiation. (d) Cumulative precipitation. Solid lines show the mean value of the variable over studied maize fields, and the bounded area represents one standard deviation of the variables.
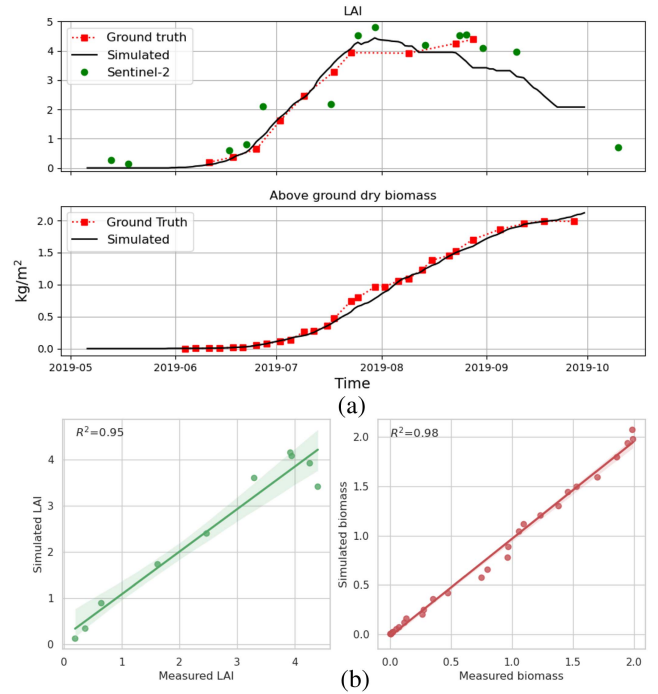

Fig. 4.   DSSAT model calibration. (a) (Top) Time-series fit between: estimated LAI from Sentinel-2 data, simulated LAI from DSSAT model, and in situ measurements; (down) simulated biomass from DSSAT and ground measurements. (b) Comparison between simulated and measured LAI and biomass.

where y_i represents the ith observation, $\hat{y}_i$ is the ith predicted value, and n is the number of observations. The mean absolute error (MAE) gives the average absolute difference between the predicted and the actual values. The mean square error (MSE) is used to measure the error of the model in simulating SAR observables [61] that is magnifying large errors. The coefficient of determination, R^2, is used to show the capability of the model in explaining the variation of the actual data. In an ideal case, R^2 is equal to 1. In addition, we use Pearson and Spearman's correlation, which shows the correlation between predicted and observed backscatter.

G. Feature Analysis

In order to understand and quantify the importance of different features in defining the regression model, we used a feature analysis algorithm. The MRMR algorithm is applied to maximize the relevance of a feature set with the dependent variable and minimize the redundancy in a feature set [62]. The MRMR algorithm searches to find an optimal subset of features (S) that maximize V, the relevance of a feature set with response variable (y), where this relevance is quantified through the mean value of the mutual information (I) as

$$V = \frac{1}{|S|} \sum_{x \in S} I(x, y) \tag{6}$$

where $|S|$ is the number of features in S. The redundancy, W, is quantified by the mean of the mutual information between the features within the set

$$W = \frac{1}{|S|^2} \sum_{x,z \in S} I(x, z). \tag{7}$$

The MRMR algorithm ranks features by using the mutual information quotient (MIQ) value as follows:

$$\text{MIQ} = \frac{V}{W}. \tag{8}$$

IV. RESULTS AND DISCUSSION

A. DSSAT Calibration and Performance

Fig. 3 shows three years of meteorological data, including TMIN, TMAX, SRAD, and cumulative precipitation in daily steps. DSSAT needs to be calibrated for the specific crop variety of interest. The model was calibrated using data collected during field experiment in Reusel, Noord-Brabant (see Fig. 1 and Section II). As illustrated in Fig. 4, after calibration, DSSAT simulated the LAI and biomass with 95% and 98% accuracy, respectively. Fig. 4(a) compares the time series of predicted and observed LAI and AGB. LAI estimates were derived from Sentinel-2 NDVI observation and from in situ measurements. NDVI-derived and simulated LAI have a downward trend after the end of August, while the LAI value from field measurement increases. As reported in [63], the field measurements of LAI were obtained by multiplying the averaged leaf area by the plant density. Consequently, the loss of LAI (related to primary productivity) due to leaf degradation after the crop reaches maturity is not reflected in the field data. The quality of
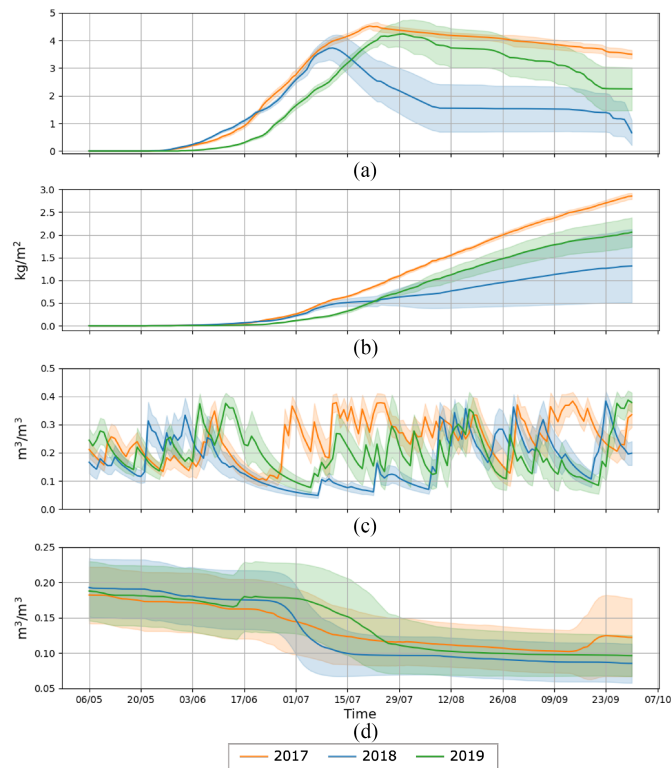
Fig. 5. Vegetation and soil parameters during the growing season in daily steps from the DSSAT model that are considered relevant inputs to simulate SAR observables. (a) LAI. (b) Above ground dry biomass. (c) Surface soil moisture. (d) Root zone soil moisture. Solid lines show the mean value of the variable over studied maize fields and the bounded area represents one standard deviation of the variable.

NDVI-derived LAI estimates is limited by the spatial resolution of the Sentinel-2 data and by the density of the crop, as the visible and near-infrared reflectances depend also on bare soil exposure [64]. Fig. 4(b) displays the simulated LAI and biomass versus the predicted values. Generally, there is a good agreement between the simulated and observed values for LAI and biomass, showing that the model is well calibrated.

Fig. 5 provides an interannual comparison between DSSAT-simulated crop biogeophysical parameter outputs of the maize fields for normal (2017) and drought-affected (2018 and 2019) years. In 2018, Europe experienced a very hot and dry summer. Drought in 2018 and 2019 affected agricultural production in The Netherlands. The 2018 drought influenced the groundwater levels, and hence, crop production [65]. The influence of the drought on Sentinel-1 SAR observables was discussed by [30].

Starting in July 2018, reduced soil moisture levels led to a significant drop in the simulated LAI and the rate of growth of the simulated dry biomass. The maximum daily average of LAI reaches approximately 3.5, compared to around 4.5 in 2017. Similarly, the biomass accumulation was substantially lower. Fig. 5 also illustrates that, in our case, the simulated LAI and AGB have higher variances during drought periods than in normal conditions. For example, this is clearly visible in 2018 for LAI, after it reaches its highest value in July. This illustrates that the anomaly in root zone soil moisture was sufficient to constrain crop growth.

B. Modeled Versus Measured Radar Observables

Fig. 6 provides the comparison of NRCS in VV and VH, cross ratio (CR, VH/VV), and interferometric coherence VV between DSSAT-SVR estimated (in red) and Sentinel-1 observations (in blue). The results are calculated for the independent test datasets. In this figure, training and test data are from the same year in this case.

Before crop emergence, radar backscatter is controlled by the surface roughness and moisture content of the exposed soil. During this period, sudden variations of the NRCS are caused by precipitation events. Starting from late May, when maize enters the leaf development stage, radar backscatter increases as the plant grows. During the stem elongation stage to tassel initiation in July, LAI rapidly reaches its maximum, leading to an increase in both co- and cross-polarized NRCS and in the CR. Once the crop reaches maximum LAI, the sensitivity of backscatter to growth decreases. The fluctuations in backscatter are probably due to the rain events on 6 June 2017, 1 and 7 June 2018, and 28 May and 5 June 2019. When maize reaches physiological maturity, in the last week of August, radar backscatter begins to decrease until harvest. This happens due to the decrease in VWC. The interfield variability ranges from 1 to 4 dB. Other studies also established similar temporal behavior of the radar observables for maize fields [24], [66]. In 2018, dry conditions resulted in an earlier ripening and harvest, leading to a shorter growing season compared to 2017 [30]. This is evident in the observed NRCS values, particularly in the cross-pol channel and the CR. The jump in NRCS co-pol in August 2018 was caused by the rain event on 7 August. Fig. 5(c) shows that in July and the beginning of August, surface soil moisture was lower than in 2017 and a rain event on 7 August 2018 led to higher soil moisture values.

As expected, coherence values are higher before the maize emergence. After crop emergence, coherence drops quickly due to temporal decorrelation associated with crop growth [67]. The coherence value remains low (<0.3) through the vegetative period of maize. The lower panels in Fig. 6 illustrate the reasonable agreement between estimated and observed SAR observables.

In Section III-E, it is argued that the average backscatter of three acquisitions after planting can serve as an indicator of the rough surface contribution to the backscatter for each parcel. This variable, referred to as offset, is included along with LAI, biomass, SM_S, and SM_R in the analysis. To assess the impact of the offset parameter, the SVR model was trained and tested with and without taking it into account. Fig. 7 shows the time series of the difference between the observed and the estimated VV backscatter with and without using the offset, when the model is trained and tested on normal 2017 data. This figure illustrates that including the offset reduces the difference between the estimated and observed backscatter at the beginning of the vegetative period when the total backscatter is still sensitive to surface scattering. Similar results were obtained in other years and channels, as shown in supplementary Fig. S1. We calculate statistics for the entire growing period and also for the period where the surface scattering contribution is expected to be significant. This period covers the start of the vegetative
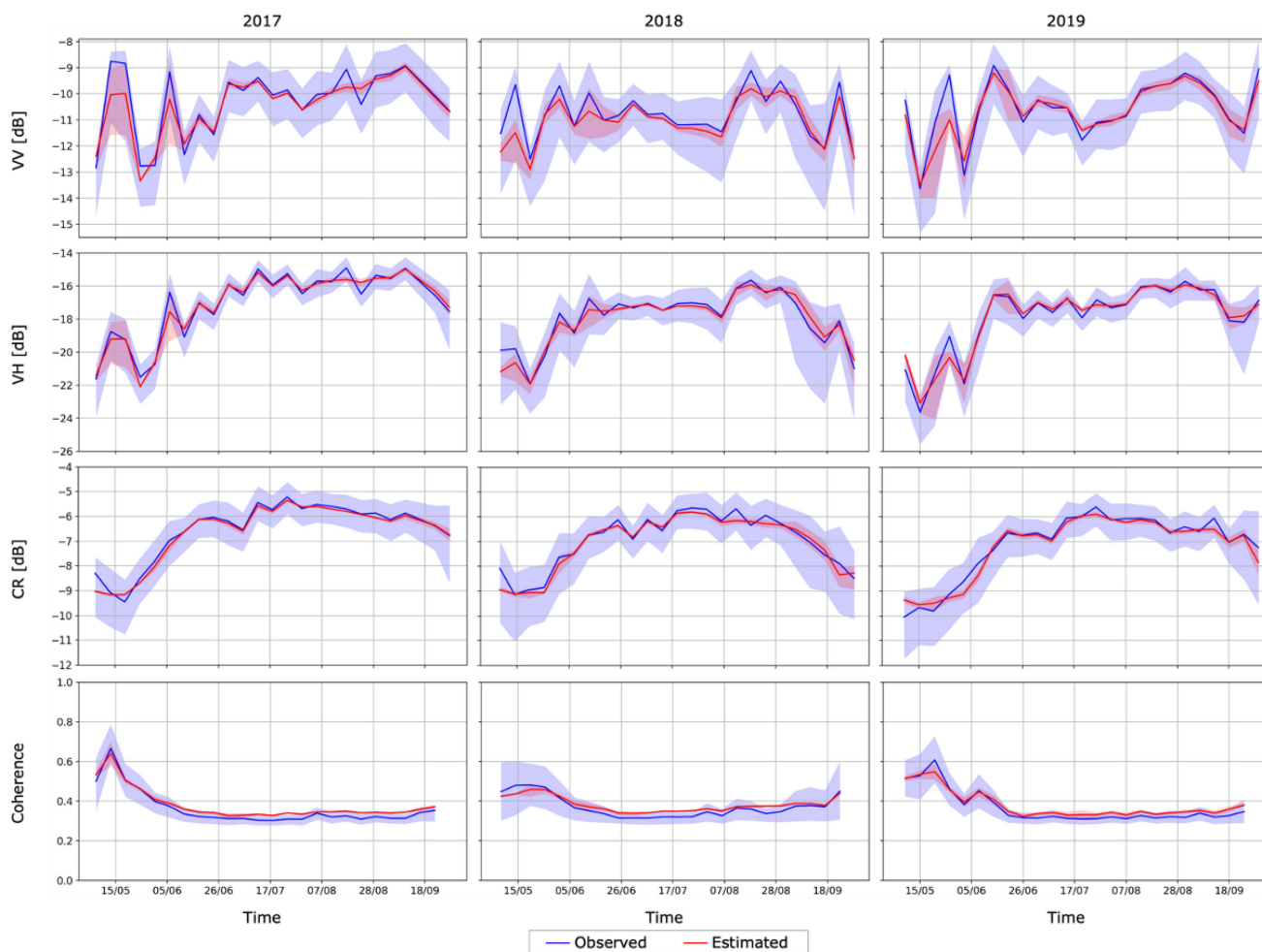
Fig. 6. Comparison of NRCS in VV, VH, cross-ratio (CR, VH/VV), and interferometric coherence VV between DSSAT-SVR estimated and Sentinel-1 observations. Each column is associated with a different year. Solid lines indicate the mean value of the feature over maize parcels in the test set, and the bounded area shows the 20th–80th percentiles. Training and test data are from the same year in this case.
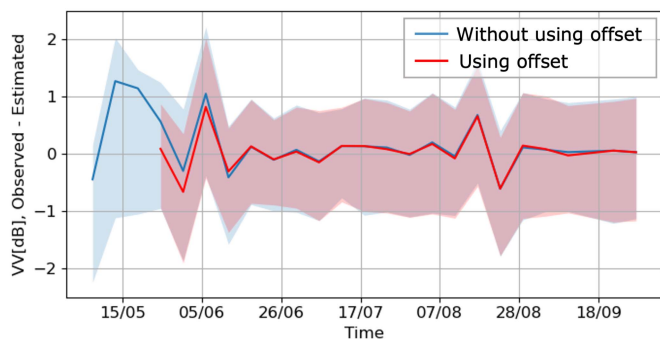

Fig. 7. Time series difference between observed and estimated backscatter VV in 2017 with and without using the offset.

stage, from emergence, typically the last week of May, until the beginning of tassel emergence, around the first week of July each year. Including the offset reduces the error, improves the accuracy, and increases the correlation coefficients. For example, the R^2 score increases from 0.45, 0.15, and 0.27 to 0.55, 0.30, and 0.44 at the beginning of the vegetative period for 2017, 2018,

and 2019, respectively. Supplementary Fig. S2 shows standard regression performance metrics for backscatter using train and test sets from the same year, both with and without the offset at the beginning of the vegetative period. Because the offset is included to account for, e.g., roughness in the surface scattering, all results discussed hereafter have been obtained including the offset parameter.

Tables IV and V display the MAE and Pearson correlation for different years and channels over the beginning of the vegetative period and the whole growing period (the values within the parentheses). The first column of the tables indicates the year in which the train data were selected. Other statistical metrics are presented in supplementary Tables S1– S5.

Given the significantly different conditions during the three years considered, we expect low model accuracy in the case that the training dataset does not include data from the year to which the SVR model is applied. This is confirmed by the results presented. For instance, using 2018 VV and VH backscatter data to train the model and applying it to 2017 results in a high MAE and very low correlation. Likewise, using VV and VH backscatter in 2017 as training data to predict backscatter in

TABLE IV

MEAN ABSOLUTE ERROR (MAE) BETWEEN ESTIMATED AND OBSERVED σ_{VV}^0, σ_{VH}^0, CR, AND γ_{VV}, OVER THE BEGINNING OF THE VEGETATIVE PERIOD AND THE WHOLE GROWING PERIOD (THE VALUES WITHIN THE PARENTHESIS)

2017 MAE	σ_{VV}^0	σ_{VH}^0	CR	γ_{VV}
2017	1.01 (0.98)	1.05 (0.93)	0.78 (0.77)	0.05
2018	2.89 (2.64)	2.53 (2.03)	0.91 (0.98)	0.07
2019	1.47 (1.37)	1.40 (1.34)	0.91 (0.92)	0.06
2018 + 2019	1.64 (1.42)	1.53 (1.31)	0.90 (0.89)	0.06
2017+2018+2019	1.17 (1.05)	1.45 (1.32)	0.80 (0.78)	0.05
2018 MAE				
2017	1.88 (1.91)	1.99 (2.44)	1.09 (1.92)	0.15
2018	1.04 (1.18)	1.18 (1.33)	0.95 (1.02)	0.07
2019	1.38 (1.70)	1.65 (1.70)	1.07 (1.18)	0.10
2017 + 2019	1.42 (1.55)	1.64 (1.60)	1.05 (1.18)	0.10
2017+2018+2019	1.14 (1.23)	1.28 (1.37)	0.95 (1.02)	0.07
2019 MAE				
2017	1.72 (1.64)	1.93 (1.66)	1.05 (1.05)	0.07
2018	1.46 (1.76)	1.53 (1.66)	0.98 (0.97)	0.06
2019	1.12 (1.00)	1.18 (1.00)	0.82 (0.83)	0.05
2017 + 2018	1.46 (1.30)	1.43 (1.26)	0.98 (0.96)	0.06
2017+2018+2019	1.20 (1.06)	1.25 (1.05)	0.84 (0.85)	0.05

TABLE V

PEARSON CORRELATION BETWEEN ESTIMATED AND OBSERVED σ_{VV}^0, σ_{VH}^0, CR, AND γ_{VV}, OVER THE BEGINNING OF THE VEGETATIVE PERIOD AND THE WHOLE GROWING PERIOD (THE VALUES WITHIN THE PARENTHESIS)

2017 Pearson	σ_{VV}^0	σ_{VH}^0	CR	γ_{VV}
2017	0.75 (0.69)	0.80 (0.80)	0.68 (0.63)	0.77
2018	-0.2 (-0.09)	0.09 (0.40)	0.61 (0.54)	0.50
2019	0.45 (0.33)	0.69 (0.57)	0.56 (0.50)	0.57
2018 + 2019	0.27 (0.30)	0.63 (0.68)	0.57 (0.53)	0.57
2017+2018+2019	0.67 (0.63)	0.67 (0.70)	0.67 (0.63)	0.75
2018 Pearson				
2017	0.17 (-0.02)	0.16 (0.08)	0.49 (0.15)	0.10
2018	0.55 (0.53)	0.61 (0.60)	0.58 (0.55)	0.50
2019	0.33 (0.21)	0.42 (0.33)	0.50 (0.35)	0.04
2017 + 2019	0.21 (0.21)	0.33 (0.32)	0.50 (0.36)	0.17
2017+2018+2019	0.44 (0.48)	0.53 (0.58)	0.60 (0.56)	0.50
2019 Pearson				
2017	0.44 (0.22)	0.45 (0.39)	0.64 (0.47)	0.54
2018	0.45 (0.19)	0.59 (0.40)	0.65 (0.57)	0.52
2019	0.67 (0.63)	0.74 (0.71)	0.75 (0.67)	0.74
2017 + 2018	0.50 (0.40)	0.62 (0.59)	0.66 (0.58)	0.58
2017+2018+2019	0.62 (0.58)	0.71 (0.69)	0.74 (0.66)	0.70

2018, produces similar results. This happens because the trained model is applied to sets of inputs for which it has not been trained. As an illustrative example, Fig. 8 shows the bivariate histogram of LAI and AGB for each year. The figure shows how in this 2-D subspace of our 4-D parameter space, the crop state follows similar trajectories during the beginning of the vegetative periods, where both LAI and AGB grow rapidly, after which we can observe distinct trajectories. The consequence is that some regions of the parameter space are only visited in particular years.

Training the model with backscatter in 2018 to predict it in 2019 performed reasonably well during the start of the vegetative stage of 2019, as these two periods suffer from drought. The CR produces better results as it suppresses the influence of soil moisture. The CR generally follows the overall temporal behavior of LAI. Notably, year-to-year transferability is improved in the CR
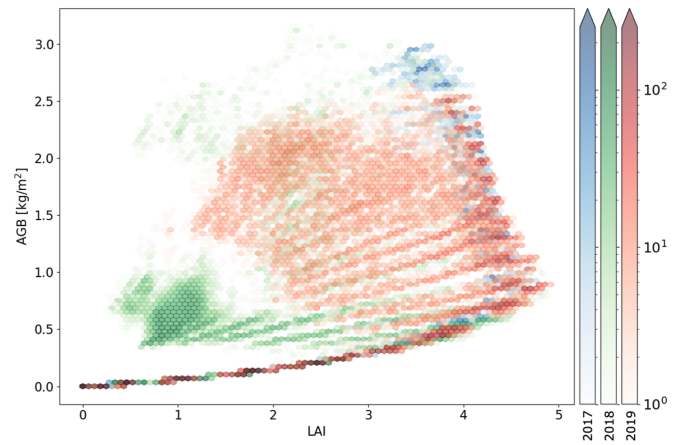

Fig. 8. Bivariate histogram of LAI and dry biomass for 2017 (blue), 2018 (green), and 2019 (red). The color bars show the normalized density of samples on a logarithmic scale for each year.

results. When the model is trained on multiple years (2017, 2018, and 2019), it performs almost as well as when it is trained on the same year.

The CR is not an independent observable, while we already have σ_{VH}^0 and σ_{VV}^0. However, this does not necessarily mean that the best possible forward models of σ_{VH}^0 and σ_{VV}^0 would give the best possible model of the CR. In particular, the CR is often used because it partially suppresses some contributions, for example, soil moisture that affects σ_{VH}^0 and σ_{VV}^0 in similar ways. The quality of the forward model for the backscattering coefficients may be limited by its ability to correctly represent the effect of soil moisture effects or by errors in the soil moisture input data, which would result in errors in the ratio of these forward-modeled coefficients. In contrast, the CR forward model will be less affected by these errors.

We see a similar general behavior for the coherence, with a high correlation between the predicted and observed coherences in the same-year and multiyear cases but relatively low in the cross-year cases. In particular, the results for 2018 if the model is trained with 2017 or 2019 data are very poor. As discussed in Section IV-C, the trained coherence model is sensitive to the dry biomass difference. Indeed, one can expect the coherence to be higher if the biomass is not changing. Fig. 5 suggests that in 2018, for a significant number of fields, the dry biomass predicted by DSSAT stagnates around August, which can cause the model to predict high coherences during that period. When we train including 2018 data, the model *learns* to put much less emphasis on the biomass difference. In a physical conceptual model, we would take the biomass difference as a good indicator of coherence during the beginning of the growth period, but not anymore at later stages, where the wind-induced motion of the crop is sufficient to explain a low coherence. Other coherence statistics are presented in the supplementary Table S6.

Fig. 9 shows the time series of observed backscatter in different years and channels (in blue) and estimated backscatter from the combination of three years (multiyear) in green and estimated backscatter from a year that behaves differently with the estimating year (cross-year) 2017 in red and 2018 in black.
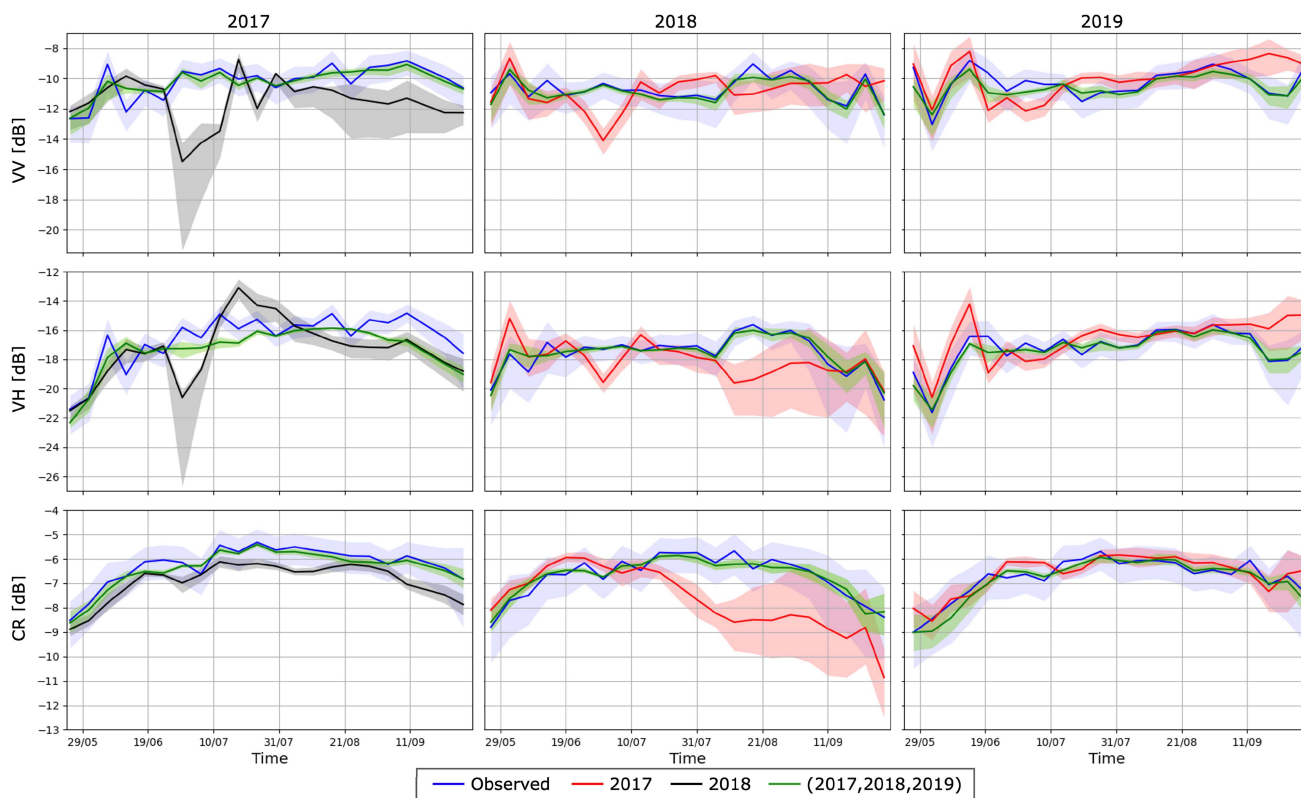
Fig. 9.   Time series of observed and estimated backscatter in 2017 (first column), 2018 (second column), and 2019 (third column) for different channels. The years that training data are selected from are shown in the legend.
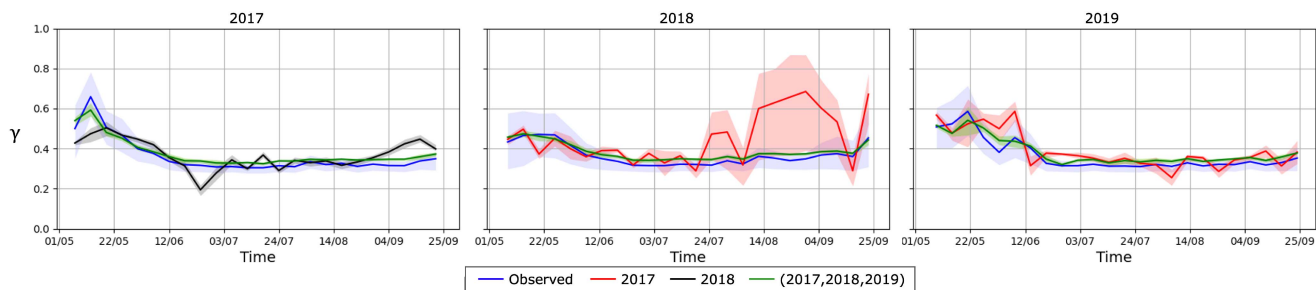

Fig. 10.   Time series of observed and estimated coherence VV in different years. The legend shows the year in which the training data belongs. Different year training and test data were used for the estimation of observables.

The transparent buffer shows 20th–80th percentiles. This figure demonstrates the impact of having partially disjoint sets of SVR input parameters (or DSSAT outputs) for different years. In early vegetative stages, interannual variation in the SM_S will cause errors in the predicted backscatter. For example, when we apply the 2018-trained model to 2017 inputs, we see large errors in the NRCS at times where the SM_S in 2018 was consistently lower than in 2017 (see Fig. 5). This problem does not affect the CR because it is much less sensitive to SM_S variations [68]. At later vegetative stages, the NRCS is much more controlled by LAI and dry biomass. This explains, for example, the large error in the prediction of all the 2018 observables, including CR, for a 2017-trained model: the low LAI values (see Fig. 8) of 2018

were never encountered in the 2017 training dataset. Similar results were obtained using all other combinations of training and testing years (see Supplementary Fig. S3)

Fig. 10 displays the same time series for the VV coherence. In general, the highest correlation and lowest error are observed when the training and testing data are from the same year or when data from all three years are used. As expected, the model wrongly shows high coherences for part of 2018 growth when we trained the model with 2017. This can be explained by considering the quite constant slope of 2017 biomass time series as compared to the wide range of biomass differences for 2018. Similar results were obtained for coherence from all training and testing year combinations (see Supplementary Fig. S5).

Fig. 11. Rank features for SVR with backscatter using the MRMR algorithm. The bars in each plot are positioned on the left side for same-year and on the right side for multiyear.
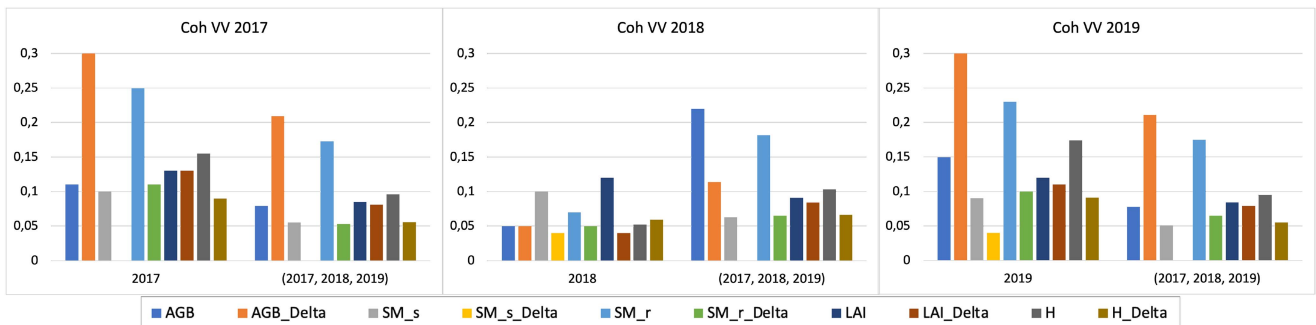

Fig. 12. Rank features for SVR with coherence VV using the MRMR algorithm. The bars in each plot are positioned on the left side for same-year and on the right side for multiyear. The variables with the "Delta" extension represent the difference between the dates of the pair of consecutive acquisitions used to create the interferograms.

C. Feature Analysis

The feature analysis aims to understand the drivers of SAR observables and to ensure that they are physically plausible. The MRMR algorithm is applied to assess the importance of the different variables in the regression model.

Fig. 11 displays the feature importance scores of the different parameters fed to the SVR model for the estimation of the

VV and VH backscatter and the CR for different years, in the same-year and multiyear cases. In general, dry biomass and the offset parameter have the highest scores. Significant correlations between dry biomass and C-band radar backscatter were also observed in [66]. In CR, because the sensitivity to soil moisture is minimized, we expect the higher importance to AGB always would be the case unless LAI and AGB are interchangeable so that one is picked up in the same-year case and the other

in the multiyear case. The impact of LAI in VV and VH was found to be minimal for all three years, possibly due to the strong correlation between LAI and dry biomass. Similarly, the influence of soil moisture was relatively low, likely because the analysis covers the entire growing season. The influence of soil moisture varies with the growth stage, as backscatter sensitivity to soil moisture decreases with increasing biomass. According to Fig. 12, the biomass difference is the most important factor for coherence VV across different years, as this parameter represents the overall growth throughout the entire period, making it a significant indicator. A nonzero difference in biomass implies a low coherence, which happens from emergence till harvest, while a surface soil moisture difference has the least impact (the feature importance scores of other training years are presented in supplementary Figs. S4 and S6).

V. Conclusion

This article aimed at demonstrating the feasibility of using machine-learning techniques to create a forward model linking biogeophysical crop parameters to C-band radar observables. This forward model serves as a bridge between the observed SAR data and the crop model simulations, enabling the integration of the two. We used a crop-growth model as an alternative to in situ data to provide crop descriptors over 1500 maize fields in the Netherlands. The MRMR was used to quantify the sensitivity of the SAR observables to the DSSAT variables. We demonstrate that the connections between crop biogeophysical variables, such as LAI, AGB, SM_S, and SM_R, and the modeled SAR observables, such as NRCS and coherence, are plausible and consistent with known physical principles of microwave remote sensing of vegetated surfaces.

In the early season, surface scattering plays an important role in the interaction with the soil surface, so the mean value of backscatter in three acquisitions after the planting date in each year is used as a proxy for information about the roughness, row geometry, and other static parameters that influence surface scattering. Adding this proxy improves the estimations in the early season. We demonstrate that the difference in environmental conditions (drought and nondrought situations) affects the model accuracy when the training dataset does not include data from the year to which the SVR model is applied. The reason is that there is not enough spatial intravariability within the area of study due to the similarity in soil texture and rainfall patterns across the province. However, DSSAT-SVR can estimate SAR observables with reasonable accuracy including the effect of surface roughness and using three years of training data. The CR shows better transferability from year to year as it minimizes the influence of soil moisture. Applying the method over a larger area with more heterogeneity or a longer time frame of the observations should result in an improved performance as the model would train with a much wider set of data.

Like all models, crop-growth model performance itself depends on the quality of the input data and the calibration of the model parameters. Therefore, there is always some degree of uncertainty associated with them. This uncertainty can be reduced, to some degree, by the availability of accurate meteorological forcing data at suitably fine spatial resolutions as well as in situ observations of biomass and LAI for model calibration.

Combining crop-growth models with machine-learning methods has the potential to estimate remote sensing observations without solely relying on ground measurements. Ongoing research will consider the suitability of this approach for anomaly detection in agricultural applications, and its use in a data assimilation context where Sentinel-1 data are used to constrain the state and parameters of the crop-growth model.

References

[1] R. Torres et al., "GMES Sentinel-1 mission," *Remote Sens. Environ.*, vol. 120, pp. 9–24, May 2012.

[2] G. Macelloni, S. Paloscia, P. Pampaloni, F. Marliani, and M. Gai, "The relationship between the backscattering coefficient and the biomass of narrow and broad leaf crops," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 4, pp. 873–884, Apr. 2001.

[3] A. G. Konings, K. Rao, and S. C. Steele-Dunne, "Macro to micro: Microwave remote sensing of plant water content for physiology and ecology," *New Phytologist*, vol. 223, no. 3, pp. 1166–1172, 2019.

[4] H. McNairn and J. Shang, "A review of multitemporal synthetic aperture radar (SAR) for crop monitoring," in *Multitemporal Remote Sensing: Methods and Applications (Remote Sensing and Digital Image Processing)*, Y. Ban, Ed. Cham, Switzerland: Springer, 2016, pp. 317–340.

[5] S. C. Steele-Dunne, H. McNairn, A. Monsivais-Huertero, J. Judge, P.-W. Liu, and K. Papathanassiou, "Radar remote sensing of agricultural canopies: A review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 2249–2273, May 2017.

[6] A. Veloso et al., "Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications," *Remote Sens. Environ.*, vol. 199, pp. 415–426, Sep. 2017.

[7] S. Khabbazan et al., "Crop monitoring using Sentinel-1 data: A case study from The Netherlands," *Remote Sens.*, vol. 11, no. 16, Jan. 2019, Art. no. 1887.

[8] K. Harfenmeister, D. Spengler, and C. Weltzien, "Analyzing temporal and spatial characteristics of crop parameters using Sentinel-1 backscatter data," *Remote Sens.*, vol. 11, no. 13, Jan. 2019, Art. no. 1569.

[9] D. Han et al., "Crop water content of winter wheat revealed with Sentinel-1 and Sentinel-2 imagery," *Sensors*, vol. 19, no. 18, Jan. 2019, Art. no. 4013.

[10] L. Pulvirenti et al., "A surface soil moisture mapping service at national (Italian) scale based on Sentinel-1 data," *Environ. Model. Softw.*, vol. 102, pp. 13–28, Apr. 2018.

[11] D. Mandal et al., "An investigation of inversion methodologies to retrieve the leaf area index of corn from C-band SAR data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 82, Oct. 2019, Art. no. 101893.

[12] M. Vreugdenhil et al., "Sensitivity of Sentinel-1 backscatter to vegetation dynamics: An Austrian case study," *Remote Sens.*, vol. 10, no. 9, Sep. 2018, Art. no. 1396.

[13] H. Pan, Z. Chen, A. de Wit, and J. Ren, "Joint assimilation of leaf area index and soil moisture from Sentinel-1 and Sentinel-2 data into the WOFOST model for winter wheat yield estimation," *Sensors*, vol. 19, no. 14, Jul. 2019, Art. no. 3161.

[14] J. Betbeder, R. Fieuzal, and F. Baup, "Assimilation of LAI and dry biomass data from optical and SAR images into an agro-meteorological model to estimate soybean yield," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2540–2553, Jun. 2016.

[15] S. Pazhanivelan et al., "Spatial rice yield estimation using multiple linear regression analysis, semi-physical approach and assimilating SAR satellite derived products with DSSAT crop simulation model," *Agronomy*, vol. 12, no. 9, Sep. 2022, Art. no. 2008.

[16] B. Bauer-Marschallinger et al., "Toward global soil moisture monitoring with Sentinel-1: Harnessing assets and overcoming obstacles," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 520–539, Jan. 2019.

[17] N. Efremova, M. E. A. Seddik, and E. Erten, "Soil moisture estimation using Sentinel-1/-2 imagery coupled with CycleGAN for time-series gap filing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 4705111.

[18] G. J. M. D. Lannoy and R. H. Reichle, "Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 catchment land surface model for soil moisture estimation," *J. Hydrometeorol.*, vol. 17, no. 2, pp. 669–691, Feb. 2016.

[19] G. J. M. De Lannoy et al., "Perspective on satellite-based land data assimilation to estimate water cycle components in an era of advanced data availability and model sophistication," *Front. Water*, vol. 4, 2022, Art. no. 981745.

[20] E. P. W. Attema and F. T. Ulaby, "Vegetation modeled as a water cloud," *Radio Sci.*, vol. 13, no. 2, pp. 357–364, Mar. 1978.

[21] F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing: Active and Passive: From Theory to Applications*, vol. 3. Norwood, MA, USA: Artech House, 1986.

[22] D. Rains, H. Lievens, G. J. M. De Lannoy, M. F. McCabe, R. A. M. De Jeu, and D. G. Miralles, "Sentinel-1 backscatter assimilation using support vector regression or the water cloud model at European soil moisture sites," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jun. 2022, Art. no. 4013105.

[23] F. T. Ulaby, K. Sarabandi, K. McDonald, M. Whitt, and M. C. Dobson, "Michigan microwave canopy scattering model," *Int. J. Remote Sens.*, vol. 11, no. 7, pp. 1223–1253, Jul. 1990.

[24] A. Davitt, J. M. Winter, and K. McDonald, "Integrated crop growth and radiometric modeling to support sentinel synthetic aperture radar observations of agricultural fields," *J. Appl. Remote Sens.*, vol. 14, no. 4, Nov. 2020, Art. no. 044508.

[25] B. A. Forman and R. H. Reichle, "Using a support vector machine and a land surface model to estimate large-scale passive microwave brightness temperatures over snow-covered land in North America," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 9, pp. 4431–4441, Sep. 2015.

[26] X. Shan et al., "Towards constraining soil and vegetation dynamics in land surface models: Modeling ASCAT backscatter incidence-angle dependence with a deep neural network," *Remote Sens. Environ.*, vol. 279, Sep. 2022, Art. no. 113116.

[27] S. de Roos, L. Busschaert, H. Lievens, M. Bechtold, and G. J. M. De Lannoy, "Optimisation of AquaCrop backscatter simulations using Sentinel-1 observations," *Remote Sens. Environ.*, vol. 294, Aug. 2023, Art. no. 113621.

[28] Introductie-PDOK, Accessed: Feb. 21, 2022. [Online]. Available: https://www.pdok.nl/introductie/-/article/basisregistratie-gewaspercelen-brp

[29] KNMI, Accessed: Feb. 21, 2022. [Online]. Available: https://daggegevens.knmi.nl/

[30] M. Shorachi, V. Kumar, and S. C. Steele-Dunne, "Sentinel-1 SAR backscatter response to agricultural drought in The Netherlands," *Remote Sens.*, vol. 14, no. 10, Jan. 2022, Art. no. 2435.

[31] P. C. Vermunt, S. C. Steele-Dunne, S. Khabbazan, J. Judge, and N. C. Van de Giesen, "Extrapolating continuous vegetation water content to understand sub-daily backscatter variations," *Hydrol. Earth Syst. Sci.*, vol. 26, no. 5, pp. 1223–1241, Mar. 2022.

[32] J. W. Jones et al., "Decision support system for agrotechnology transfer: DSSAT V3," in *Understanding Options for Agricultural Production (Systems Approaches for Sustainable Agricultural Development)*, G. Y. Tsuji, G. Hoogenboom, and P. K. Thornton, Eds. Dordrecht, Netherlands: Springer, 1998, pp. 157–177.

[33] M. K. van Ittersum and M. Donatelli, "Modelling cropping systems of the symposium and preface to the special issues," *Eur. J. Agronomy*, vol. 18, no. 3, pp. 187–197, Jan. 2003.

[34] DSSAT, Accessed: Feb. 21, 2022. [Online]. Available: https://dssat.net/

[35] C. A. Jones and J. R. Kiniry, *CERES-Maize: A Simulation Model of Maize Growth and Development*. College Station, TX, USA: Texas A&M Univ. Press, 1986.

[36] W. Malik, R. Isla, and F. Dechmi, "DSSAT-CERES-maize modelling to improve irrigation and nitrogen management practices under mediterranean conditions," *Agricultural Water Manage.*, vol. 213, pp. 298–308, Mar. 2019.

[37] I. Ahmed, M. H. ur Rahman, S. Ahmed, J. Hussain, A. Ullah, and J. Judge, "Assessing the impact of climate variability on maize using simulation modeling under semi-arid environment of Punjab, Pakistan," *Environ. Sci. Pollut. Res.*, vol. 25, no. 28, pp. 28413–28430, Oct. 2018.

[38] T. Brandsma, "Comparison of automatic and manual precipitation networks in the Netherlands," KNMI, Tech. Rep., 2014.

[39] T. Hengl et al., "SoilGrids250 m: Global gridded soil information based on machine learning," *PLoS One*, vol. 12, no. 2, Feb. 2017, Art. no. e0169748.

[40] Soilgrid, Accessed: Feb. 21, 2022. [Online]. Available: https://soilgrids.org/

[41] J. Koo and J. Dimes, "HC27 generic soil profile database," Harvard Dataverse, V5, 2013, doi: 10.7910/DVN/90WJ9W.

[42] K. E. Saxton and W. J. Rawls, "Soil water characteristic estimates by texture and organic matter for hydrologic solutions," *Soil Sci. Soc. Amer. J.*, vol. 70, no. 5, pp. 1569–1578, 2006.

[43] E. Han, A. V. M. Ines, and J. Koo, "Development of a 10-Km resolution global soil profile dataset for crop modeling applications," *Environ. Model. Softw.*, vol. 119, pp. 70–83, Sep. 2019.

[44] E. Han, A. Ines, and J. Koo, "Global high-resolution soil profile database for crop modeling applications," *Harvard Dataverse*, vol. 1, pp. 1–37, Oct. 2015.

[45] J. Morel et al., "Challenges for simulating growth and phenology of silage maize in a nordic climate with APSIM," *Agronomy*, vol. 10, no. 5, May 2020, Art. no. 645.

[46] R. S. Loomis and J. S. Amthor, "Yield potential, plant assimilatory capacity, and metabolic efficiencies," *Crop Sci.*, vol. 39, no. 6, pp. 1584–1596, 1999.

[47] F. X. López-Cedrón, K. J. Boote, J. Piñeiro, and F. Sau, "Improving the CERES-Maize model ability to simulate water deficit impact on maize production and yield components," *Agronomy J.*, vol. 100, no. 2, pp. 296–307, 2008.

[48] J. Lacasa, T. J. Hefley, M. E. Otegui, and I. A. Ciampitti, "A practical guide to estimating the light extinction coefficient with nonlinear Modelsa case study on maize," *Plant Methods*, vol. 17, no. 1, Jun. 2021, Art. no. 60.

[49] F. Sau, K. J. Boote, W. McNair Bostick, J. W. Jones, and M. Inés Mínguez, "Testing and improving evapotranspiration and soil water balance of the DSSAT crop models," *Agronomy J.*, vol. 96, no. 5, pp. 1243–1257, 2004.

[50] V. Kumar, M. Huber, B. Rommen, and S. Steele-Dunne, "Agricultural SandboxNL database V1.0," 4TU ResearchData, 2022, doi: 10.4121/14438750.v2.

[51] V. Kumar, M. Huber, B. Rommen, and S. C. Steele-Dunne, "Agricultural SandboxNL: A national-scale database of parcel-level processed Sentinel-1 SAR data," *Sci. Data*, vol. 9, no. 1, Jul. 2022, Art. no. 402.

[52] Google, "Google Earth Engine data catalog, sentinel-1 GRD," Accessed: Feb. 21, 2022. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD

[53] R. Touzi, A. Lopes, J. Bruniquel, and P. Vachon, "Coherence estimation for SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 135–149, Jan. 1999.

[54] A. Braun and L. Veci, *TOPS Interferometry Tutorial*. Paris, France: ESA, 2021.

[55] ESA, "SNAP Toolbox," Accessed: Feb. 21, 2022. [Online]. Available: https://step.esa.int/main/download/snap-download/

[56] Y. Kang et al., "How universal is the relationship between remotely sensed vegetation indices and crop leaf area index? A global assessment," *Remote Sens.*, vol. 8, no. 7, Jul. 2016, Art. no. 597.

[57] S. I. Jiménez-Jiménez, M. d. J. Marcial-Pablo, W. Ojeda-Bustamante, E. Sifuentes-Ibarra, M. A. Inzunza-Ibarra, and I. Sánchez-Cohen, "VICAL: Global calculator to estimate vegetation indices for agricultural areas with landsat and Sentinel-2 data," *Agronomy*, vol. 12, no. 7, Jul. 2022, Art. no. 1518.

[58] M. C. Mozer, M. I. Jordan, and T. Petsche, *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 1997.

[59] J. Ezzahar et al., "Evaluation of backscattering models and support vector machine for the retrieval of bare soil moisture from Sentinel-1 data," *Remote Sens.*, vol. 12, no. 1, Jan. 2020, Art. no. 72.

[60] D. Rains, R. Sabia, D. Fernández-Prieto, M. Marconcini, and T. Katagis, "Extended analysis of SMOS salinity retrieval by using support vector regression (SVR)," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 2265–2268.

[61] C. J. Willmott, "Some comments on the evaluation of model performance," *Bull. Amer. Meteorological Soc.*, vol. 63, pp. 1309–1369, Nov. 1982.

[62] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J. Bioinf. Comput. Biol.*, vol. 3, no. 2, pp. 185–205, Apr. 2005.

[63] P. C. Vermunt et al., "Response of subdaily L-band backscatter to internal and surface canopy water dynamics," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7322–7337, Sep. 2021.

[64] S.-Y. Hong, K.-A. Sudduth, N.-R. Kitchen, C.-W. Fraisse, H.-L. Palm, and W.-J. Wiebold, "Comparison of remote sensing and crop growth models for estimating within-field LAI variability," *Korean J. Remote Sens.*, vol. 20, no. 3, pp. 175–188, 2004.

[65] M. H. J. van Huijgevoort, J. A. de Wit, and R. P. Bartholomeus, "Evaluation of 2018 drought and effectiveness of adaptation measures in the Netherlands," in *Proc. EGU Gen. Assem. Conf. Abstr.*, 2020, Art. no. 9166.

[66] G. Wiseman, H. McNairn, S. Homayouni, and J. Shang, "RADARSAT-2 polarimetric SAR response to crop biomass for agricultural production monitoring," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 11, pp. 4461–4471, Nov. 2014.

[67] Y. Morishita and R. F. Hanssen, "Temporal decorrelation in L-, C-, and X-band satellite radar interferometry for pasture on drained peat soils," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1096–1104, Feb. 2015.

[68] M. Vreugdenhil et al., "Sentinel-1 cross ratio and vegetation optical depth: A comparison over Europe," *Remote Sens.*, vol. 12, no. 20, Jan. 2020, Art. no. 3404.

Tina Nikaein (Member, IEEE) received the B.Sc. degree in surveying and geomatics engineering and the M.Sc. degree in remote sensing from Tehran University, Tehran, Iran, in 2015 and 2018, respectively. She is currently working toward the Ph.D. degree in monitoring agriculture using synthetic aperture radar (SAR) with the Department of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands.

Her main research interest include radar remote sensing, InSAR, and environmental monitoring.

Paco Lopez-Dekker (Senior Member, IEEE) was born in Nijmegen, The Netherlands, in 1972. He received the Ingeniero degree in telecommunication engineering from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 1997, the M.S. degree in electrical and computer engineering from the University of California, Irvine, CA, USA, in 1998, under the Balsells Fellowship, and the Ph.D. degree in clear-air imaging radar systems to study the atmospheric boundary layer from the University of Massachusetts, Amherst, MA, USA, in 2003.

In 2003, he joined Starlab Barcelona, where he worked on the development of GNSS-R sensors and techniques. From 2004 to 2006, he was a Visiting Professor with the Department of Telecommunications and Systems Engineering, Universitat Autonoma de Barcelona. In March 2006, he was awarded a Ramon y Cajal Grant to conduct pioneering research on bistatic synthetic aperture radar (SAR) at Remote Sensing Laboratory, UPC. At the university, he taught courses on signals and systems, signal processing, communications systems and radiation, and guided waves. Between November 2009 and August 2016, he led the SAR Missions Group, Microwaves and Radar Institute, German Aerospace Center, Wessling, Germany. Since September 2016, he has been an Associate Professor with the Department of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherland. He has been deeply involved in several radar mission proposals, and is the Lead Investigator of the Harmony ESA Earth Explorer 10 Mission candidate. He has coauthored 50 peer-reviewed journal articles and more than 100 conference contributions in a broad range of topics related to radar remote sensing.

Susan C. Steele-Dunne (Member, IEEE) received the S.M. and Ph.D. degrees in hydrology from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2002 and 2006, respectively.

She has been with the Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands, since 2008. She leads the M-WAVE Group, who perform research from field to global scales, combining in situ and space-borne sensors to improve our understanding of microwave interactions with vegetation. Her research interests include the use of data assimilation, modeling and machine learning to exploit spaceborne radar instruments for applications in ecosystem and agricultural monitoring.

Vineet Kumar (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from Uttar Pradesh Technical University, Lucknow, India, in 2010, the M.Tech. degree in remote sensing and geographic information system (GIS) from the Indian Institute of Remote Sensing, Dehradun, India, in 2013, and the Ph.D. degree in synthetic aperture radar (SAR) remote sensing from the Indian Institute of Technology Bombay, Mumbai, India, in 2019.

He worked as a Postdoctoral Researcher with the Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands, from 2019 to 2022. His research interests include full and compact SAR polarimetry, bio-geo physical parameter modeling and large-scale agriculture monitoring using cloud computing.

Dr. Kumar was the recipient of the Shastri Research Student Fellowship 2016 Award by the Shastri Indo-Canadian Research Institute, India.

Manuel Huber received the B.Sc. degree in applied earth sciences from the Technical University Munich, Munich, Germany in 2017, and the M.Sc. degree in environmental engineering he obtained in 2019 at the Delft University of Technology focusing on remote sensing and environmental monitoring, in 2019.

From 2020 to 2022, he worked as Young Graduate Trainee with the European Space Agency performing research and application development by exploiting spaceborn data using a range of different modeling approaches and machine learning methods.