Earth and Space Science

RESEARCH ARTICLE
10.1029/2020EA001363

Special Section:
Results from 10 Years of UAVSAR Observations

Performance Evaluation of UAVSAR and Simulated NISAR Data for Crop/Noncrop Classification Over Stoneville, MS

S. Kraatz[1], S. Rose[1], M.H. Cosh[2], N. Torbick[3], X. Huang[3], and P. Siqueira[1]

[1]Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA, [2]USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA, [3]Applied Geosolutions, Durham, NH, USA

Key Points:

- Crop/noncrop classifications were evaluated using Uninhabited Aerial Vehicle Synthetic Aperture Radar (SAR) and simulated NASA ISRO SAR data using three spatial resolutions and performance metrics
- Accuracy was close to or exceeded the NISAR science requirement of 80% over a wide range of CV_{thr} values (0.2 to 0.4)
- Optimal crop/noncrop delineating thresholds monotonically decreased with spatial resolution for each performance metric

Correspondence to:
S. Kraatz,
skraatz@umass.edu

Citation:
Kraatz, S., Rose, S., Cosh, M. H., Torbick, N., Huang, X., & Siqueira, P. (2021). Performance evaluation of UAVSAR and simulated NISAR data for crop/noncrop classification over Stoneville, MS. *Earth and Space Science*, *8*, e2020EA001363. https://doi.org/10.1029/2020EA001363

Received 23 JUL 2020
Accepted 12 NOV 2020

Abstract Synthetic Aperture Radar (SAR) data are well-suited for change detection over agricultural fields, owing to high spatiotemporal resolution and sensitivity to soil and vegetation. The goal of this work is to evaluate the science algorithm for the NASA ISRO SAR (NISAR) Cropland Area product using data collected by NASA's airborne Uninhabited Aerial Vehicle SAR (UAVSAR) platform and the simulated NISAR data derived from it. This study uses mode 129, which is to be used for global-scale mapping. The mode consists of an upper (129A) and lower band (129B), respectively having bandwidths of 20 and 5 MHz. This work uses 129A data because it has a four times finer range resolution compared to 129B. The NISAR algorithm uses the coefficient of variation (CV) to perform crop/noncrop classification at 100 m. We evaluate classifications using three accuracy metrics (overall accuracy, J-statistic, Cohen's Kappa) and spatial resolutions (10, 30, and 100 m) for crop/noncrop delineating CV thresholds (CV_{thr}) ranging from 0 to 1 in 0.01 increments. All but the 10 m 129A product exceeded NISAR's mission accuracy requirement of 80%. The UAVSAR 10 m data performed best, achieving maximum overall accuracy, J-statistic, and Kappa values of 85%, 0.62, and 0.60. The same metrics for the 129A product respectively are: 77%, 0.40, 0.36 at 10 m; 81%, 0.55, 0.49 at 30 m; 80%, 0.58, 0.50 at 100 m. We found that using a literature recommended CV_{thr} value of 0.5 yielded suboptimal accuracy (65%) at this site and that optimal CV_{thr} values monotonically decreased with decreasing spatial resolution.

1. Introduction

Timely and accurate large-scale data on agricultural activity are important for tracking and identifying management practices, cropland distribution, and for supporting food security programs. Needs of agricultural monitoring community involve a combination of high (<4 m) to moderate (<30 m) spatial resolutions, frequent revisit time, and open access, operational coverage for large spatial extents (Fritz et al., 2019). Because a majority of agricultural fields are just over 2 hectares in size, spatial resolutions should ideally be on the order of 100 m or less to adequately capture agricultural conditions and change (Yan & Roy, 2016). Revisit times on the order of weeks or less are desirable, because agricultural fields may undergo substantial change on diurnal or greater timescale due to processes such as tilling or precipitation (McNairn & Brisco, 2004). Collectively, since agriculture is closely tied to global markets and food security, there is a strong need for accurately monitoring agricultural activity at global scale (Becker-Reshef et al., 2019).

Since retrievals made by Synthetic Aperture Radar (SAR) systems now meet or exceed these needs, they are well-suited for large-scale agricultural monitoring. Spaceborne SAR, such as the European Space Agency's (ESA) Sentinel-1 can map the Earth once every 6 to 12-days at moderate spatial resolution (Torres et al., 2012). SAR data provide valuable information useful for cropland identification, crop type classification and estimating yield (Betbeder et al., 2016; Huang et al., 2019; Whelen & Siqueira, 2018). For example, these data can be used to estimate biomass using backscatter magnitude, crop heights using interferometry and crop structure using polarimetry (Erten et al., 2016; Ferrazzoli et al., 1997; Wiseman et al., 2014). Also, unlike optical sensors, SAR can collect high quality data day and night and is largely unimpacted by atmospheric conditions, therefore having excellent potential for collecting dense time series. For these reasons, plus an increasing adoption of open data access policies, cheaper cloud computing resources, and a steady pipeline of future platforms, there has been a rapid increase in the use of SAR data sets with regards to agricultural applications and decision support systems.

The NASA ISRO SAR (NISAR) mission, slated for deployment in 2022, is designed to meet specific science requirements for applications pertaining to Ecosystems (e.g., agriculture, biomass), solid Earth and the cryosphere. NISAR will have frequent global observations, repeating approximately every 12 days, with L-band (1.26 GHz) having global coverage and S-band (3.2 GHz) collecting data specific areas, such as India (NISAR Science Team, 2020). NISAR's focus on agriculture is reflected in its science requirement of having the capability of routinely producing a global cropland area product and close collaboration with the United States Department of Agriculture (USDA), GEO Global Agricultural Monitoring (GEOGLAM), Joint Experiment for Crop Assessment and Monitoring (JECAM), and other stakeholders for calibration and validation efforts (NISAR Science Team, 2020).

Accuracy requirements of NISAR's Cropland Area product are to be met using an algorithm based on the coefficient of variation (CV), computed over time at each pixel, to identify those locations with relatively higher change from crop growth stages and field activities. The theory behind using the temporal CV is that areas experiencing greater change over time are consistent with agricultural activity (cropland) and areas that do not change as much are considered noncrop such as built-up areas or forests. Waterbodies also tend to have large CV values, mostly owing to having low signal-to-noise due to predominantly specular scattering. It is therefore important to obtain a waterbody mask to remove classifications over water. Crop and noncrop areas are then delineated by comparing the temporal CV values to a threshold (CV_{thr}). The use of a single delineating threshold for crop and noncrop had been tested in several prior studies, usually achieving accuracies in the 80% range. Whelen and Siqueira (2017) applied the temporal CV approach to AgriSAR and ALOS PALSAR data collected over Germany and Minnesota, respectively (Whelen & Siqueira, 2017). While ALOS PALSAR has a revisit period of 46 days, the AgriSAR data consisted of a dense time series collected on 12 dates between April and August 2006. Classification accuracies of 87%, 79%, and 78% were achieved for AgriSAR L-band, AgriSAR C-band, and ALOS PALSAR, respectively. They also applied the temporal CV approach in a subsequent study, but using ALOS PALSAR data over 11 agricultural sites within the contiguous United States (Whelen & Siqueira, 2018). That study used three approaches for setting CV_{thr} values for crop/noncrop classifications: it was set according to (1) the threshold of maximum separation of histograms of CV values for crop and noncrop classes, (2) the threshold corresponding to the largest Youden J-statistic value (Youden, 1950), and (3) using a fixed CV_{thr} value of 0.5. Averaged over the 11 sites investigated, the approaches performed nearly identical, all of them falling in a 74%–75% range.

Apart from Whelen and Siqueira (2017), the temporal CV approach had not been tested over agricultural areas using temporally dense (i.e., approximately biweekly) L-band time series or simulated NISAR data. There is an important need to use dense timeseries for computing CV values, because the biweekly observations would be better able to capture processes and management such as tilling, irrigation and rain, vegetation growth, and harvesting and result in more representative CV values. Furthermore, CV and ergo, the optimal CV_{thr} values for crop/noncrop delineation, will also depend on factors such as: the frequency of observation, cropping practices, landscape heterogeneity, and the spatial resolution of the data. Therefore, there had been a strong need to collect dense L-band time series over hydrologically dynamic regions to evaluate the NISAR science algorithms such as that to be used to calculate cropland area (Chapman et al., 2019). Thus, NASA conducted the "Uninhabited Aerial Vehicle SAR (UAVSAR) AM/PM Campaign" throughout summer 2019, consisting of repeat flights approximately every 2 weeks over forests and agricultural areas in the Southeast at local times of approximately 6 a.m. and 6 p.m. (Chapman et al., 2019). Data obtained during this campaign were also used to generate simulated NISAR products that more closely resemble the quality of data to be collected by NISAR (JPL, 2020).

Simulated NISAR data had not yet been studied in context of generating a cropland area product. This work focuses on providing first estimates on likely accuracy levels of NISAR's crop/noncrop classifications using approximately biweekly simulated NISAR data over an agricultural area near Stoneville, MS. Additionally, we also explore how CV_{thr} values depend on spatial resolution, and how accuracy metrics differ when using the best available data (UAVSAR) compared to the simulated NISAR products on 10×10 m grids. The simulated NISAR data were evaluated at three spatial resolutions: (1) 10 m which approximately corresponds to the 12×6.2 m slant-range grid spacing of the simulated NISAR multi look complex (MLC) product; (2) 100 m which corresponds to that of NISAR's Cropland Area product; and (3) 30 m which is an in-between spatial resolution that matches that of the annually produced USDA Cropland Data Layer (CDL) which

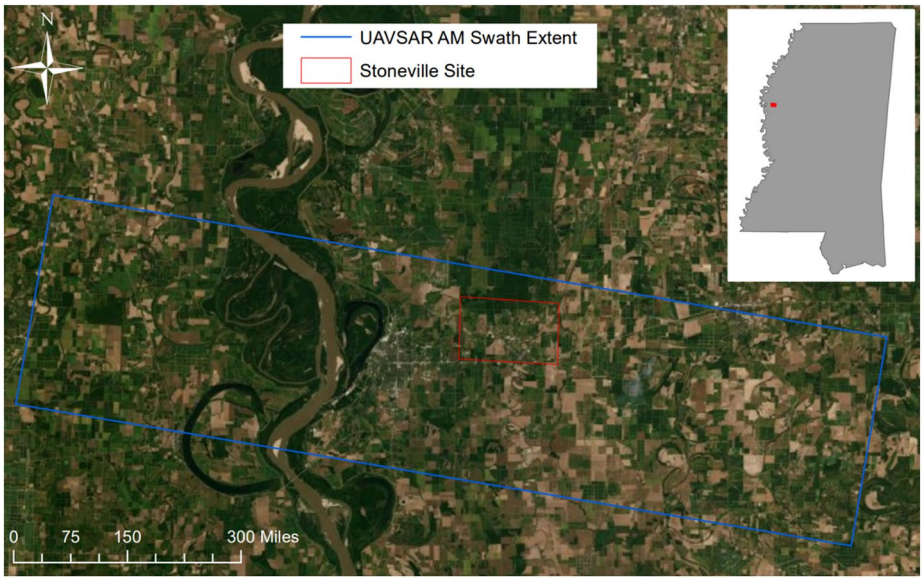

Figure 1. The study area is located in the western part of Mississippi, close to Stoneville. The swath of SAR data collected during the 6 a.m. is indicated by the blue box. The red box indicates the region of interest this study used for crop/noncrop classification. The site is located between an incidence angle of 33–47 degrees, matching the range to be covered by NISAR. SAR, Synthetic Aperture Radar; NISAR, NASA ISRO SAR; UAVSAR, Uninhabited Aerial Vehicle SAR.

this study uses as reference data set for evaluating the crop/noncrop classifications (Boryan et al., 2011). Our hypotheses are: (1) UAVSAR data would perform the best due to these data having less noise and finer spatial resolution than any of the simulated data, (2) that classification performance deteriorates with spatial resolution and (3) that optimal CV_{thr} values have some, but limited, dependence on spatial resolution.

2. Study Area, Data Sets, and Processing

2.1. Study Area

The study area is located in the Lower Mississippi River Basin and is part of the Big Sunflower River watershed within the Yazoo River Basin. The study area ("Stoneville Site" in Figure 1) is located near Stoneville in Washington County, MS. The site mainly consists of agricultural fields and riparian- and bottomland-wetlands which make a useful feasibility experiment given their dynamic range. Agriculture predominantly consists of soybeans, corn, and cotton. USDA Agricultural Research Service's (ARS) Crop Production Systems Research Unit farms are also located within the marked region. These research fields are routinely monitored for crop (e.g., growth stage) and soil conditions (e.g., soil moisture), and therefore are a valuable resource for evaluating NISAR algorithms.

2.2. USDA CDL

This study uses the CDL as reference for the accuracy assessments. The CDL is produced using observations from various spaceborne platforms such as Landsat 8 and Sentinel 2, provided on 30 × 30 m grids, and encompasses 106 different crops (Boryan et al., 2011). CDL accuracy with respect to ground truth is estimated to be in the 85%–95% range for major crop types. The CDL of the prior year is usually made available early the following year, is freely available to the public, and is produced operationally by the USDA National Agricultural Statistics Service (NASS). CDL data from 1997 onwards can be obtained at https://nassgeodata.gmu.edu/CropScape/.

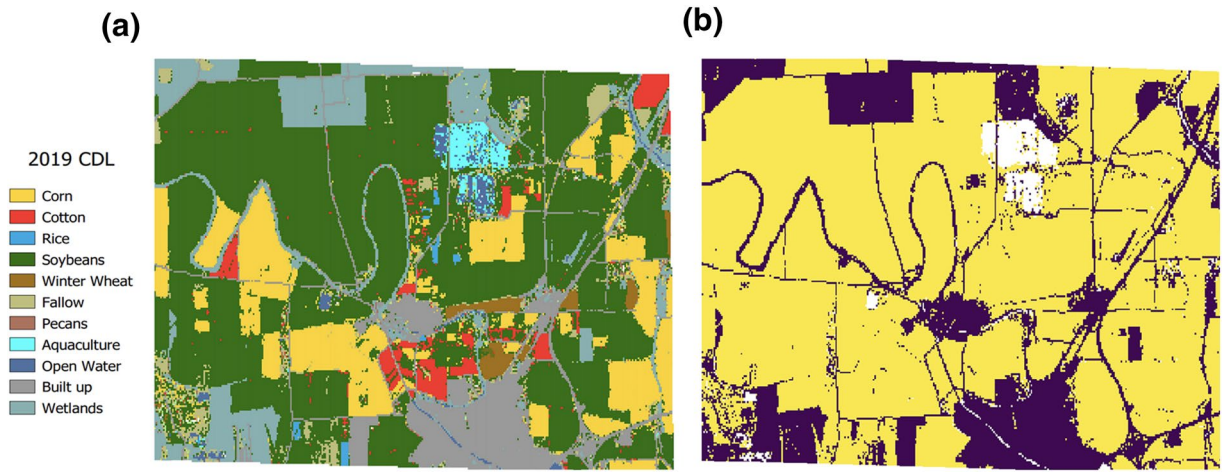

Figure 2. The 2019 USDA/NASS Crop Data Layer at study site (a) and the binary crop (yellow) noncrop (purple) classification (b) according to Table 1. Some CDL classes for which CV approach is known to not perform applicable were masked. CDL, Cropland Data Layer; CV, coefficient of variation; NASS, National Agricultural Statistics Service; USDA, United States Department of Agriculture.

Figure 2 shows the original and the binary crop/noncrop result obtained from assigning CDL layers to crop/noncrop (Table 1). The image consists of a total of 718828 pixels (1,054 columns and 682 rows) of which 501378 (69.7 %) and 166978 (23.2 %) pixels are crop and noncrop, respectively. The remaining 50472 (7.0%) pixels were masked if they fell into the CDL classes indicated in Table 1 and according to whether pixel locations fell outside the "Stoneville Site" shapefile (Figure 1). It is necessary to mask out open water and orchard crops because the temporal CV is not well-suited to these land cover classes. Temporal CV is not expected to be able to detect substantial change over orchards, since the difference between a harvested tree to one with fruits is expected to be negligible at L-band, and the soils underneath the trees do not vary from bare/tilled though vegetated over a season. We also masked CDL classes related to open water, because these areas have a low signal-to-noise ratio and may often indicate relatively large CV values that would be falsely classified as crop. It is important to note that the CDL contains three categorization codes referring to wetlands: "Wetlands" with code 87 under "Other"; "Woody Wetlands" and "Herbaceous Wetlands" with codes of 190 and 195, respectively. Because this study masked any items falling under "Other," code 87 was masked. The wetlands shown in Figure 2 correspond to both codes 190 and 195, making up 96% and 4% of the wetland pixels, respectively.

After excluding masked cells from comparisons 668356 "valid" pixels remain, with a percent breakdown of 75% for crop and 25% for noncrop. Because the CDL is categorical and posted at 30 × 30 m grids, to make comparisons at 10, 30, and 100 m, we subdivided each CDL grid into nine 10 × 10 m grids.

Unfortunately, the CDL is limited in spatial extend to CONUS. While other land cover data sets are available at global scale, they might not be updated as frequently, have as high spatial resolution or as fine of a granulation with respect to the crop and noncrop classes (Friedl et al., 2002; Loveland et al., 2000). Since all global data sets include flags for water, and waterbodies also are readily identified using radar data, it is relatively straightforward to reproduce comparable waterbody masking globally. However, orchards are not routinely included as a land cover type in the global data sets and will therefore present an additional source of error (misclassification as a noncrop).

2.3. UAVSAR and Simulated NISAR Data

SAR data were collected by the UAVSAR as part of the NISAR UAVSAR AM/PM campaign. The NISAR UAVSAR AM/PM was conducted in the Southeast during the crop season of 2019. Key features of this campaign are high frequency repeat observations at L-band (1.26 GHz), occurring approximately every two weeks and at NISAR observing times of 6 a.m. and 6 p.m. (local time). The campaign's main purpose was

Table 1
Rules Used for Reclassifying the 2019 USDA NASS Crop Data Layer Into Binary Crop/Noncrop Grids

CDL class ranges	CDL class type	Crop/noncrop/masked
1–60	Crop (e.g., cotton, rice ...)	Crop
61–65	Noncrop (e.g., fallow, forest ...)	Noncrop
66–80	Tree crops (e.g., cherry, peach ...)	Masked
81–109	Other (e.g., water, clouds ...)	Masked
110, 112–195	NLCD classes (e.g., developed, forest, wetland)	Noncrop
196–255	Crops (e.g., carrots, garlic ...)	Crop
111	Open water	Masked

Abbreviations: CDL, Cropland Data Layer; NASS, National Agricultural Statistics Service; USDA, United States Department of Agriculture.

to obtain data for testing and developing NISAR ecosystem science algorithms, and hence predominantly focused on agricultural and forested targets (Chapman et al., 2019).

UAVSAR is NASA Jet Propulsion Laboratory's (JPL) airborne platform providing high signal-to-noise and fully polarimetric SAR observations at various frequencies (e.g., L-band at 1.26 GHz) and at 80 MHz bandwidth (Rosen et al., 2006). All UAVSAR data are freely available and provided as MLC data in slant-range coordinates and in ground-range projected (GRD) format (https://uavsar.jpl.nasa.gov/). Single look complex (SLC) data and interferometric data sets are also available for download, but only for some of UAVSAR's flight lines and dates.

Simulated NISAR data are generated from UAVSAR data. UAVSAR observations have relatively higher signal-to-noise and spatial resolution compared to what will be collected by NISAR. Therefore, additional steps were taken by NASA JPL to generate simulated NISAR products using UAVSAR SLCs as input, mainly by reducing spatial resolution and adding noise (JPL, 2020). Simulated data are available for each of NISAR's observing mode (129, 138, and 143). Each mode is split into two frequency bands (A or B) with the bandwidth of each depending on mode. This study uses frequency A of observing mode 129 (i.e., 129A), because: (1) mode 129 is NISAR's main observing mode and will have nearly global coverage and (2) 129A has a four times better spatial resolution in range compared to 129B, because the former has a bandwidth of 20 MHz compared to 5 MHz for the latter.

The simulated data sets also have additional features compared to other UAVSAR data provided by NASA JPL. For example, the simulated NISAR GRD products of a given flight line are all coregistered and posted on a common grid, facilitating time series analysis. Furthermore, the GRD product also comes with a radiometric terrain correction (RTC) calibration file. The RTC calibration file has the same dimensions as the GRD data and contains the calibration factors to be applied at each grid cell in order to obtain radiometric and terrain corrected γ^0 backscatter in linear power units.

The MLC data have units of power and are the result of taking the complex products between the different SLC data. This study only uses the "HVHV" complex product, which is calculated by HV*conjugate(HV). The HVHV data had also been multilooked using 12 pixels in azimuth (along-track of airplane) and 3 pixels in range (perpendicular to azimuth). We only tested crop/noncrop delineation using the HVHV data, because previous studies indicated that this complex product performed best for agricultural crop/noncrop classifications (McNairn & Shang, 2016; Whelen & Siqueira, 2017).

The study area described in Section 2.1 is covered by the UAVSAR Stoneville 6 a.m. flight line. Seven approximately biweekly images were used to compute CV values at each pixel (Table 2). However, ahead of computing CV values (Section 3.1), we applied further data processing steps as described below (Section 2.3.3).

Table 2
The File Names of the UAVSAR and Simulated NISAR Datasets Used as Inputs (2019), Omitting File Extension ".mlc"

Date	UAVSAR	Simulated NISAR
6-Jun	NISARA_27900_19033_000_190606_L090HVHV_CX_02	NISARA_27900_19033_000_190606_L090HVHV_CX_129A_03
20-Jun	NISARA_27900_19038_003_190620_L090HVHV_CX_02	NISARA_27900_19038_003_190620_L090HVHV_CX_129A_03
16-Jul	NISARA_27900_19048_001_190716_L090HVHV_CX_01	NISARA_27900_19048_001_190716_L090HVHV_CX_129A_02
25-Jul	NISARA_27900_19051_001_190725_L090HVHV_CX_01	NISARA_27900_19051_001_190725_L090HVHV_CX_129A_02
12-Aug	NISARA_27900_19053_013_190812_L090HVHV_CX_01	NISARA_27900_19053_013_190812_L090HVHV_CX_129A_02
23-Sep	NISARA_27900_19069_001_190923_L090HVHV_CX_01	NISARA_27900_19069_001_190923_L090HVHV_CX_129A_02
30-Sep	NISARA_27900_19070_003_190930_L090HVHV_CX_01	NISARA_27900_19070_003_190930_L090HVHV_CX_129A_02

Notes. The naming fields (in order) consist of the campaign name ("NISARA"), the aircraft heading value for Stoneville (27900), flight id number (e.g., 19033), data take number (e.g., 000), the date of observation (e.g., 190606), the frequency band, steering angle polarization (e.g., L090HVHV), the cross talk calibration flag ("CX"), and the processing version number (e.g., "_02").
Abbreviations: NISAR, NASA ISRO SAR; UAVSAR, Uninhabited Aerial Vehicle SAR.

2.4. SAR Data Processing

The input SAR data consist of (1) the high bandwidth UAVSAR (80 MHz) and (2) the relatively lower bandwidth (20 MHz) and noisier simulated NISAR products that was generated from the UAVSAR data (Section 2.3). Users may find it more convenient to use the GRD product rather than the MLC because the data had already been coregistered and the RTC calibration are also provided. However, this study used a separate workflow using MLC data as input to produce coregistered radiometrically and terrain corrected imagery, because: (1) this methodology is applicable to all available UAVSAR data, because the RTC calibration files are only available for the simulated data sets of the NISAR AM/PM campaign; and (2) we were able to ensure that an identical workflow is used for processing both the UAVSAR and simulated NISAR data.

Data were subjected to further processing using the InSAR Scientific Computing Environment (ISCE) Docker Tools (IDT) (Kraatz et al., 2020). IDT is a dockerized version of ISCE 2.3.1, additionally modified for automated processing of UAVSAR MLC time series data and is available at https://github.com/UMass-MIRSL/isce_docker_tools. It automatically applies processing Steps 1 through 5 (Figure 3).

The first processing step is to map the slant-range data to the DEM using a forward geometry mapping via the ISCE function "topo.py." This step determines the geographic coordinates and other information such as flat earth and local incidence angles at every pixel. The former is the incidence angle with respect to the ellipsoid, while the latter is the incidence angle with respect to the terrain (DEM). This step is only conducted for the reference image, here the earliest data collected (June 6, 2019). All subsequent data will also

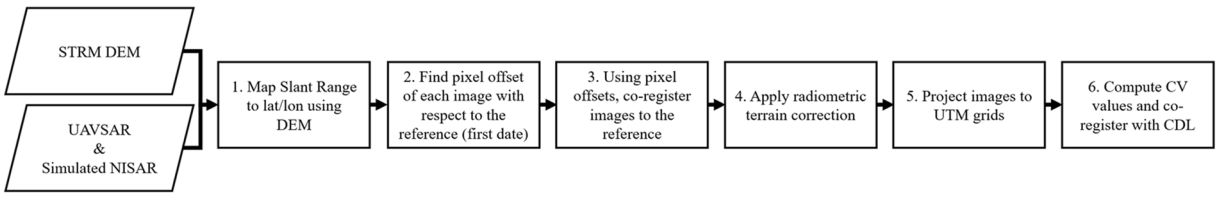

Figure 3. Flowchart showing the six data processing steps. Steps 1 through 5 were applied using the InSAR Scientific Computing Environment software via an automated workflow (ISCE Docker Tools). CDL, Cropland Data Layer; CV, coefficient of variation; ISCE, InSAR Scientific Computing Environment; NISAR, NASA ISRO SAR; UAVSAR, Uninhabited Aerial Vehicle SAR; UTM, Universal Transverse Mercator.

be posted on this grid. This step also required a digital elevation model, and we used and the Shuttle Radar Topography Mission (SRTM) 30 m data.

The second step relates to coregistering the subsequent imagery to the reference. Here we use the ISCE function "geo2rdr.py," to determine the pixel offsets in range and azimuth. This is a decimal number, because offsets need to be known to a 100th or better of a pixel for when for when this ISCE function is used to prepare data for interferometry. One reason for this offset is that subsequent UAVSAR flights are not exactly repeated. Another reason for often substantial offsets in azimuth (100s of rows) is that the SAR instrument does not record data for the same duration and at the same flight velocity.

The third processing step uses the pixel offsets to coregister each image to the reference image using "resampleSLC.py." Fourth, we use the flat earth and local incidence angles that were obtained in step 1 to apply the RTC and to project the results into γ_0 at every pixel (Small, 2011; Ulander, 1996). Fifth, the slant range MLCs were projected to Universal Transverse Mercator (UTM) zone 15N using 10 × 10, 30 × 30 × 100 m grids. Steps 1, 4, and 5 use routines that are specific to IDT and provided as python scripts at the GitHub link referenced above.

Sixth, we then computed CV values at each spatial resolution (Section 3.1). Because the CDL (30 × 30 m) and CV grids (10 m, 30 m, 100 m) have different spatial resolutions all data were subdivided into 10 × 10 m grids for making comparisons. Because the CDL and CV gridded data were produced independently, their grids are offset with respect to each other. Therefore, we coregistered the data sets by interpolating the CV values to the CDL grids.

3. Methods

This section focuses on the evaluation of the algorithm to be used to meet NISAR's Level 2 Cropland Area product science requirements. The algorithm is to have 80% accurate crop/noncrop classifications when evaluated at 100 × 100 m spatial resolution (NISAR Science Team, 2020). The algorithm is based on the temporal CV approach which had also been used in past studies (Whelen & Siqueira, 2017, 2018). The basic premise of the CV approach for delineating crop and noncrop areas is that actively cultivated lands (i.e., crop) experience more substantial change over time as compared to unmanaged lands (i.e., noncrop). Because of differences due to the types of crops, soil conditions and management practices within a given region of interest, we do not have a priori knowledge of what CV_{thr} values provide accurate crop/noncrop delineation. Therefore, we iterate through a range of CV_{thr} values, from 0.00 through 1.00 in 0.01 increments and classify each valid pixel of the CV image as crop or noncrop (Section 3.1). For each CV_{thr} value, we use the results of the confusion matrix to compute three different accuracy metrics: the overall accuracy, Youden's J-statistic, and Cohen's Kappa (Section 3.2).

3.1. Crop Classification Using the Coefficient of Variation

The CV metric represents the amount of variation in HVHV backscatter over time, with higher values indicating greater variation, and is calculated as

$$\mathrm{CV} = \frac{\sigma}{\mu} \tag{1}$$

where σ and μ respectively are the standard deviation and mean values of the backscatter data at each pixel, computed over time.

Crop/noncrop classification is determined using a fixed CV_{thr} value, applied to the region of interest where

$$\mathrm{CV}_{\mathrm{pixel}} \begin{cases} < \mathrm{CV}_{\mathrm{thr\ for\ ROI}}, \mathrm{noncrop} \\ \geq \mathrm{CV}_{\mathrm{thr\ for\ ROI}}, \mathrm{crop} \end{cases} \tag{2}$$

Table 3
Confusion Matrix for Crop/Noncrop Classification

Model (the SAR-based classifications)	Reference (the CDL)	
	Crop	Noncrop
Crop	True positive (TP)	False positive (FP)
Noncrop	False negative (FN)	True negative (TN)

Abbreviations: CDL, Cropland Data Layer; SAR, Synthetic Aperture Radar.

3.2. Performance Metrics

Our crop/noncrop results (Equation 2) are compared to the reference data set (the CDL, Section 40) using a confusion matrix. The confusion matrix tabulates the number of pixels for which both data sets agreed on the crop and noncrop pixels, respectively the true positive (TP) and true negative (TN) counts. It also tabulates classification errors where our classifications indicated crop but the CDL did not (false positive, FP) where our classifications indicated noncrop but the CDL did not (false negative, FN).

The Overall Accuracy is calculated from the confusion matrix (Table 3) as:

$$\text{Accuracy} = 100 \times \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{FN} + \text{TN})} \tag{3}$$

Youden's J-statistic indicates the threshold for which the difference between the true positive (Sensitivity) and false positive (1-Specificity) rate is the greatest (Youden, 1950; Habibzadeh et al., 2016). In that context, it gives an equal weight to false positive and false negative values. Youden's J-statistic is calculated as

$$J = \text{Sensitivity} + \text{Specificity} - 1 \tag{4}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{6}$$

According to this performance metric, the optimal CV_{thr} value is the one that yields the greatest value of J. In a visual representation, J is the vertical distance between a 1:1 line (the line of no discrimination) to a point on the Receiver Operating Characteristic (ROC) curve. The ROC curve is the result from plotting the Sensitivity versus 1-Specificity. The ROC curve allows for easy interpretation of classification performance. Classifications are generally poor if the curve falls relatively close to the 1:1 line.

Cohen's Kappa indicates the threshold for which two data sets show the best agreement, also attempting to account for random chance using standard assumptions (Cohen, 1960; McHugh, 2012). As in McHugh (2012), we calculate Kappa in terms of the four confusion matrix categories (TP, TN, FP, FN) shown in Table 3:

$$\text{Kappa} = \frac{p_O - p_e}{1 - p_e} \tag{7}$$

where p_O is the observed proportionate agreement, given by

$$p_O = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{8}$$

and p_e is the overall random agreement probability, given by

$$p_e = p_Y + p_N \tag{9}$$

where p_Y and p_N respectively are the expected probability of random agreement and disagreement, given by

$$p_Y = \frac{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN})}{(\text{TP} + \text{FP} + \text{FN} + \text{TN})^2} \tag{10}$$

$$p_N = \frac{(\mathrm{FN} + \mathrm{TN}) \times (\mathrm{FP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN})^2} \tag{11}$$

Kappa may range between −1.0 and 1.0. Values below zero indicate poor agreement while 1.0 represents a perfect agreement between the CDL and the SAR-based crop/noncrop classifications.

4. Results and Discussion

CV values of UAVSAR and simulated NISAR data range from about 0 to 2.5 (Figure 4). The UAVSAR retrievals have low noise and can clearly delineate individual fields and other small features at high contrast (Figure 4a). As expected, the simulated NISAR data are noisier and have less contrast (Figures 4b–4d). There is a considerable amount of speckle in the simulated NISAR data at 10 m (Figure 4b): intrafield CV variations are quite large, which may be problematic for accurate crop/noncrop classification. The speckle is not entirely surprising as the simulated NISAR MLCs had only been multilooked twice in azimuth with respect to the SLC and have a 12 (azimuth) × 6.2 m (range) posting. For comparison, UAVSAR's SLC resolution is 0.6 (azimuth) × 1.67 m (range) and the MLC product had been multilooked 12 × in azimuth and 3 × in range for a posting of 7.2 m by 5 m.

In general for the simulated NISAR data—especially at 10 m—features have lower contrast because the CV values have a smaller dynamic range across the scene (Figure 4b). Scene-wide CV values of the 10 m simulated NISAR data (mean: 0.63, median: 0.62) are smaller than those obtained from UAVSAR data (mean: 0.64, median: 0.66). Furthermore, for the simulated NISAR data CV values monotonically decrease with spatial resolution, and there is nearly no difference between their mean and median values over the scene: the mean/median CV values are 0.53 and 0.46 at 30 m and 100 m, respectively. These statistics indicate that finer-scale products also show relatively more frequent and higher valued CV pixels. At the same time, there is still noticeable speckle even when using 30 m grids. Due to reduced speckle and also larger pixel sizes at 100 m (Figure 4d), even fields that had large variability in CV at finer resolutions now have relatively uniform CV values, that is, the bright green field approximately centered at pixel 300,175. Because the subfield variations had been smoothened out at 100 m, and the field has much greater than average CV values (1.8 vs. 0.46), this should lead to relatively improved crop/noncrop delineation when applying a single CV_{thr} value over the entire scene.

ROC curves (Section 3.2) also provide valuable insight on the relative performance of crop/noncrop classifications using the different input data sets (Figure 5). ROC curves of the UAVSAR (Figure 5a) and 10 m simulated NISAR classifications (Figure 5b) have the greatest and least separation from the line of no discrimination (the 1:1 line), respectively. This indicates that classifications using the UAVSAR data perform best while the 10 m simulated NISAR data perform worst. Classifications using the 30 m (Figure 5c) and 100 m (Figure 5d) simulated NISAR data perform about equally and have a degree of separation from the 1:1 line falling closer to that of the UAVSAR retrievals. The notable difference between the 30 and 100 m classifications is that to obtain the same vertical separation from the 1:1 line at 100 m, a smaller CV_{thr} value must be used at the coarser resolution. This is result is consistent with the above-noted decrease of mean and median CV values at coarser spatial resolutions.

Figure 6 shows the results of pixel-wise comparisons between the UAVSAR and simulated NISAR data, compared to the CDL at specific CV_{thr} values. For all data sets, using small CV_{thr} values lead to crop overestimation errors, while high CV_{thr} value lead to crop omission errors. Visual inspection of the result indicates that the best CV_{thr} values occur around 0.4, 0.5, 0.4, and 0.3 respectively for UAVSAR and simulated NISAR data at 10 m, 30 m, and 100 m. The latter three values are consistent with the results of the ROC plots and the monotonically decreasing CV values at coarser resolution. Furthermore, visual inspection also shows relatively clearly that UAVSAR data perform the best, 10 m simulated NISAR the worst, and that the 30 and 100 m simulated NISAR results are comparable to one another and fall somewhere in between results obtained using the 10 m and UAVSAR data.

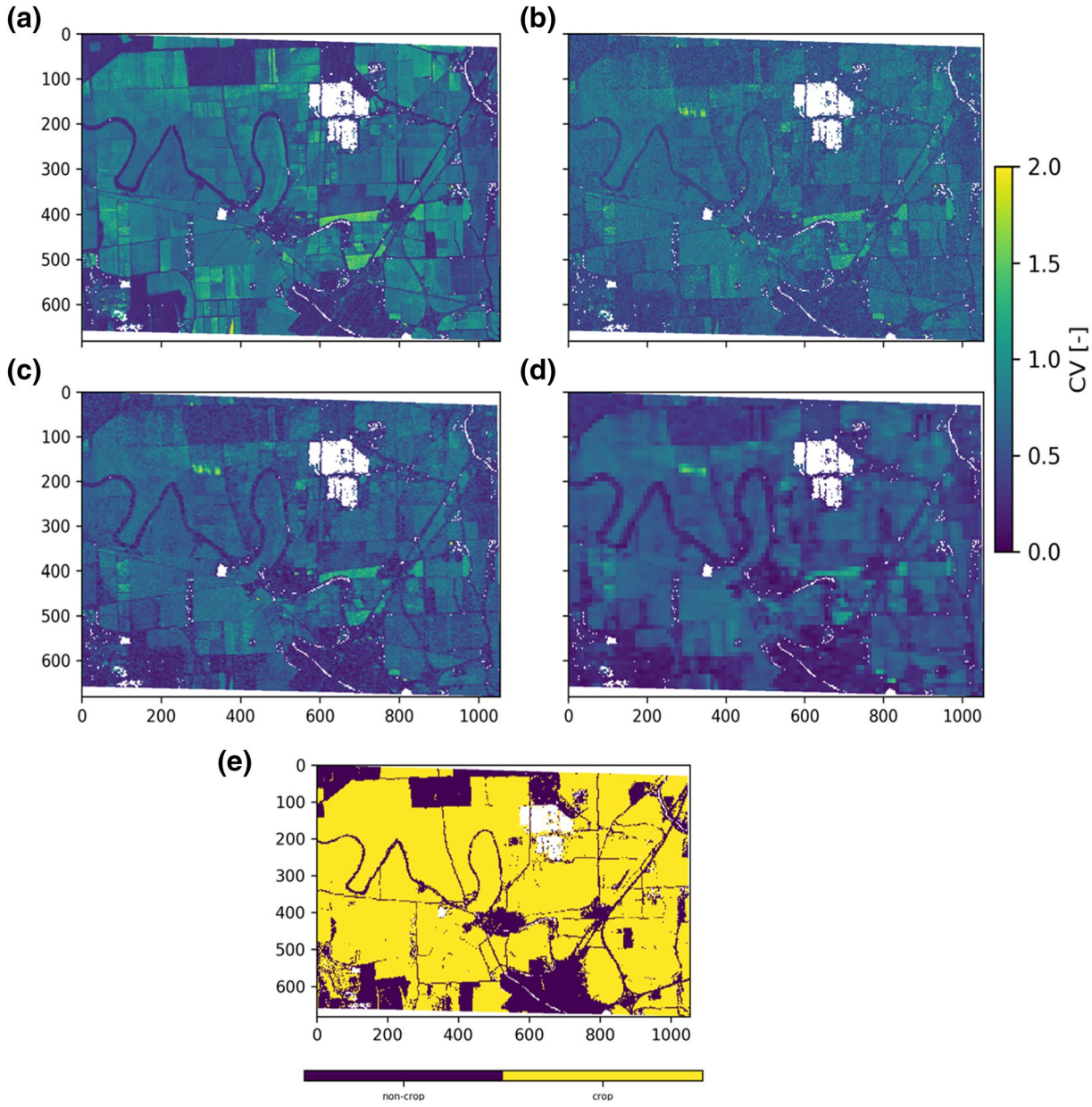

Figure 4. CV values for (a) UAVSAR 10 × 10 m and (b–d) simulated NISAR data interpolated to 10, 30, and 100 m grids and (e) the CDL classified as crop and noncrop according to Table 1. Some CDL land covers (e.g., aquaculture) were masked according to Table 1 and are shown in white. CDL, Cropland Data Layer; CV, coefficient of variation; UAVSAR, Uninhabited Aerial Vehicle SAR.

A breakdown of optimal CV_{thr} values by accuracy metric reveals that values obtained for Youden's J-statistic and Cohen' Kappa are comparable, whereas optimal CV_{thr} according to accuracy fall relatively lower (Table 4). There is a large range of optimal CV_{thr} values for the 10 m simulated NISAR data, indicating that the best agreement occurs when using CV_{thr} of 0.4, 0.6, and 0.5 for accuracy, J-statistic, and Kappa, respectively. Figure 6 shows that using CV_{thr} of 0.4 leads to many misclassifications of noncrop as crop. But as 75% of the landscape is crop, the relative contribution of these errors is relatively smaller and high accuracy is achieved. Whereas the J-statistic weighs both types of errors equally, resulting in a map that appears to have somewhat comparable rates of crop omissions and overdetections, as proportion of the total crop and noncrop classes: at a CV_{thr} of 0.57, 48972 out of 166978 noncrop grids (29%) are misclassified as crop,

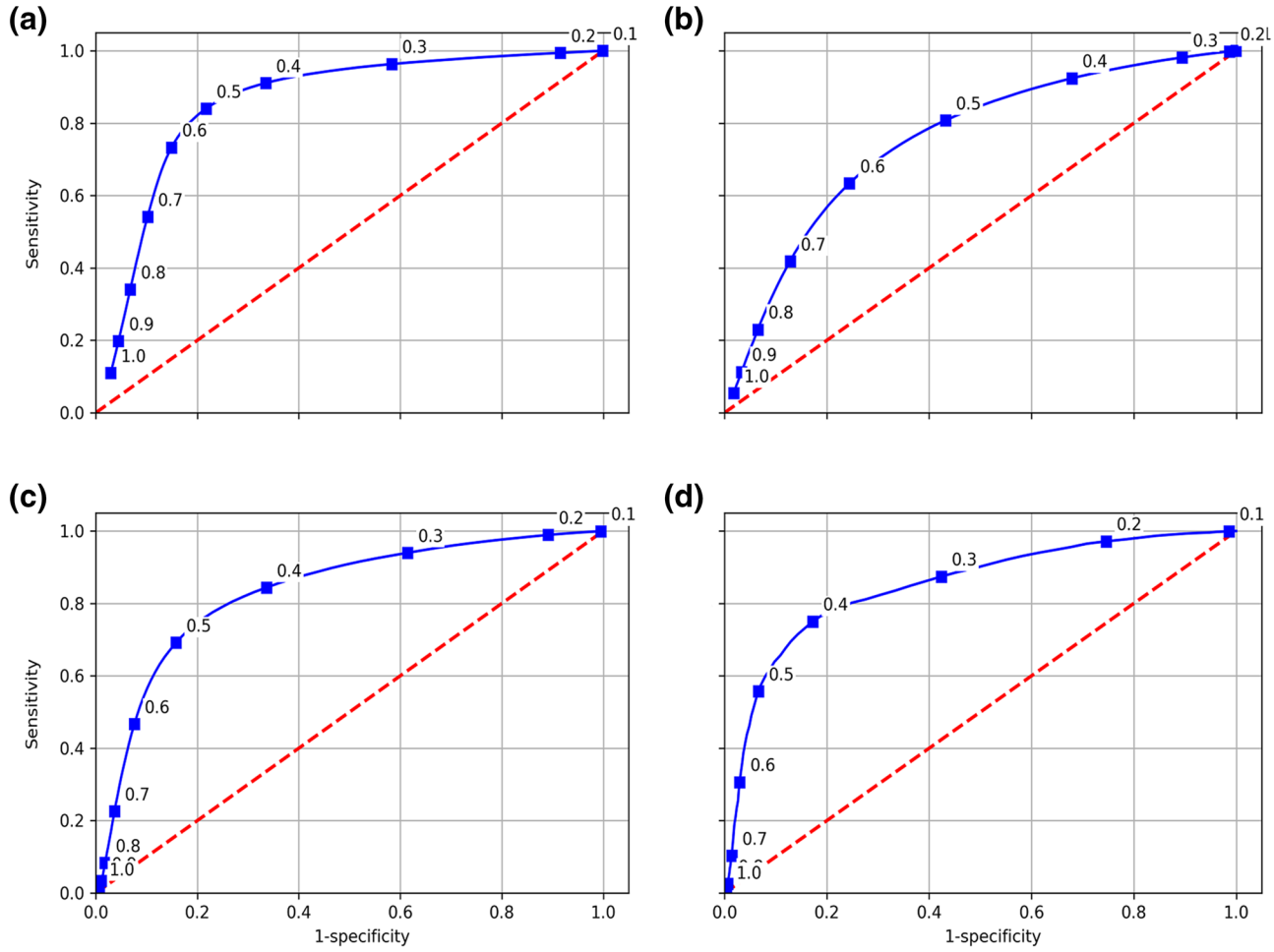

Figure 5. Receiver operating characteristic curves for (a) UAVSAR at 10 × 10 m and (b–d) simulated NISAR at 10 × 10, 30 × 30 × 100 m spatial resolution. The marked points on the curve indicate the Sensitivity and 1-Specificity values for the annotated value of the crop/noncrop delineating threshold (CV_{thr}). The 1:1 line indicates the line of no discrimination. CV, coefficient of variation; UAVSAR, Uninhabited Aerial Vehicle SAR.

while 154095 out of 501378 crop grids (31%) are misclassified as noncrop. In the 10 m simulated NISAR case, Kappa settles on a value in the middle, but approaches the threshold identified by the J-statistic with coarser resolution.

Plots of the three performance metrics versus CV_{thr} reveal that accuracy is relatively less sensitive to changes in CV_{thr}, compared to the other metrics (Figure 7). Overall accuracy remains fairly constant until CV_{thr} values exceed 0.5. Whereas, the J-statistic and Kappa values have much sharper peaks, occurring at comparable CV_{thr} values. Results for J-statistic and Kappa are close to zero for CV_{thr} values less and greater than 0.2 and 0.8, respectively. The figure also clearly shows how optimal values trend toward smaller CV_{thr} with coarser spatial resolutions: there is a particularly large contrast between the locations of peaks of the simulated NISAR data at 10 m as compared to 30 and 100 m: for J-statistic and Kappa, there is a CV_{thr} difference of about 0.1 or greater between them.

Figure 7 shows that all but the simulated NISAR classifications at 10 m are able to meet the 80% accuracy requirement. Results using the 30 and 100 m data only barely exceed the 80% requirement. For the interpretation of accuracy over the region of interest, it is important to point out that the CDL indicated a breakdown between crop and noncrop of 75% and 25%. Thus in the trivial cases of "assuming" everything is a crop ($CV_{thr} = 0$) or noncrop (high CV_{thr} such as $CV_{thr} = 1.0$), accuracy will be equal to the crop/noncrop breakdown. This is supported by the J-statistic and Kappa metrics being close to zero in these cases, indicating random chance rather than predictive skill. We also note that there appears to be a relatively large range

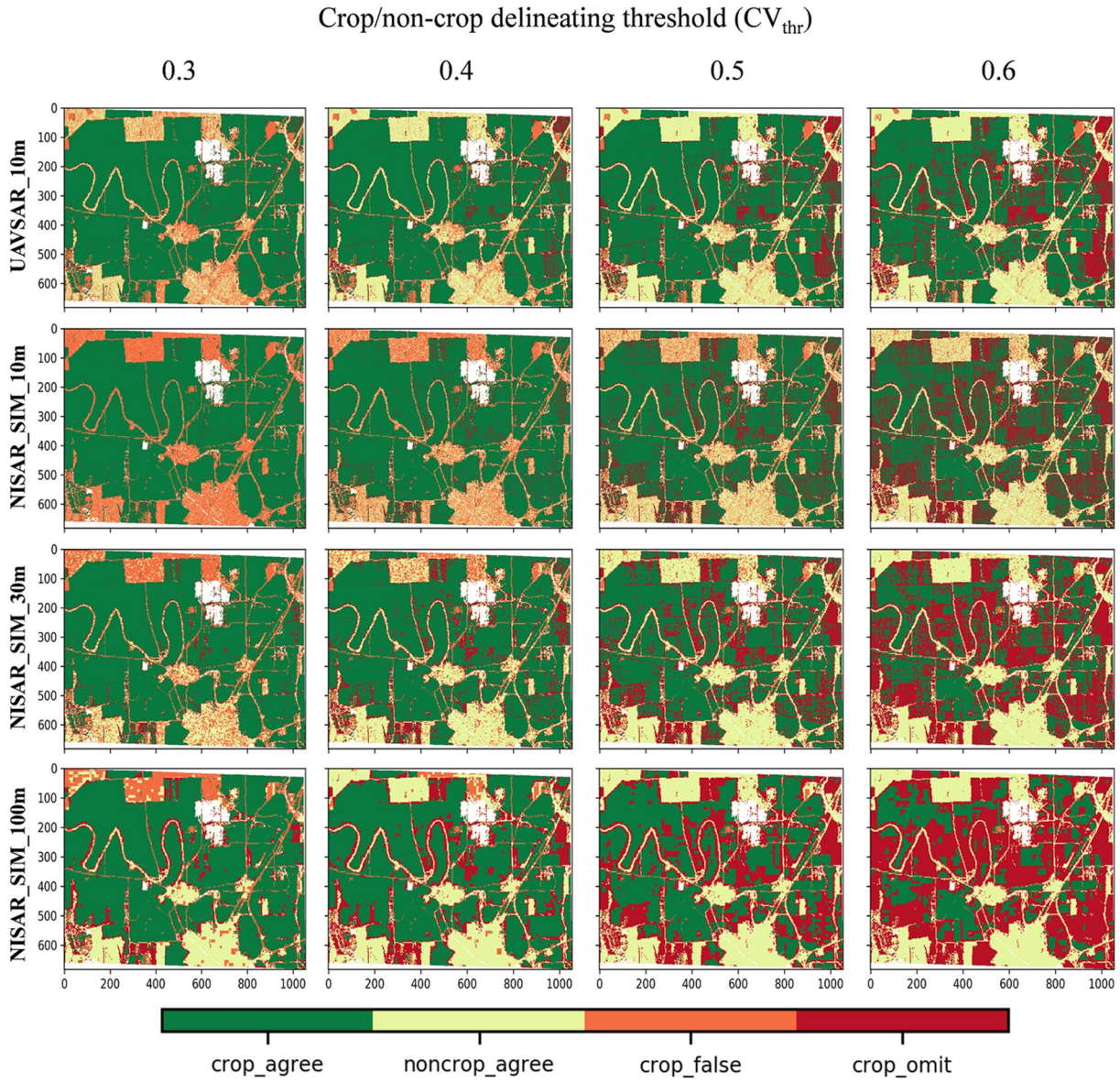

Figure 6. Crop/noncrop confusion matrix results for the UAVSAR (10 × 10 m) and simulated NISAR (10 × 10, 30 × 30 × 100 m) products (rows) as function of crop/noncrop delineating threshold (CV_{thr}, columns). Results show that to maintain the best classification accuracy with spatial resolution, it is necessary to employ smaller CV_{thr} values as the spatial resolutions become coarser. CV, coefficient of variation; NISAR, NASA ISRO SAR; UAVSAR, Uninhabited Aerial Vehicle SAR.

of CV_{thr} values, approximately ranging from 0.2 to 0.4, that could also be used to achieve accuracies close to or greater than the 80% requirement. This indicates that there could be a considerable leeway in obtaining good accuracies even when only using a single CV_{thr} value over a large region. Given NISAR mission requirements of 80% at 1 hectare scale, the simulated NISAR products meet these goals.

While our findings are generally consistent with our initial hypotheses, we note that classification performance only slightly decreased between using 30 (81%) and 100 m (80%) data. More notable is that the optimal CV_{thr} values, greatly varied according to spatial resolution and accuracy metric used, and that a wide range of CV_{thr} values ranging from 0.2 to 0.4, could be used to achieve near 80% accuracy.

Table 4
Optimal CV_{thr} Values as Function of Performance Metric (Rows) and Product or Resolution (Columns)

Optimal CV_{thr}	UAVSAR	N_10	N_30	N_100
Accuracy	0.39 (85%)	0.4 (77%)	0.34 (81%)	0.26 (80%)
J-statistic	0.5 (0.62)	0.57 (0.4)	0.46 (0.55)	0.4 (0.58)
Kappa	0.43 (0.59)	0.5 (0.36)	0.41 (0.49)	0.38 (0.50)

Notes. The values in parentheses is the maximum value of the respective metric, with Accuracy reported as percentage and J-statistic and Kappa are reported on a 0.00 to 1.00 scale. N_10, N_30, and N_100 respectively refer to the simulated NISAR data at 10, 30, and 100 m grids.

Abbreviations: CV, coefficient of variation; UAVSAR, Uninhabited Aerial Vehicle SAR.

5. Conclusions

This main goal of this work was to evaluate the proposed algorithm for generating crop/noncrop classifications using NISAR data. Crop/noncrop classifications were evaluated at spatial resolutions consistent with the finest NISAR products (around 10 m), that of the CDL (30 m) and that at which NISAR crop/noncrop classifications are to be evaluated (100 m). Using UAVSAR at 10 m resolution, we achieved crop/noncrop classification accuracy of 85%. Using the best currently available approximation of data to be collected by NISAR, we found that the mission requirement of 80% could be met in all but the 10 m case (77%). Speckle was substantial when using the 12 × 6.2 m MLCs on 10 × 10 m grids, resulting in misclassifications. And while results were not substantially better using 30 (81%) or 100 m (80%) grids, we found that a fairly large range of CV_{thr} values (0.2–0.4) may be used to approach or exceed the mission accuracy requirement of 80%.

In addition to overall accuracy, this study also used Youden's J-statistic and Cohen's Kappa to measure classification performance. Both metrics were found to be substantially more sensitive to performance differences, compared to overall accuracy. Classifications were substantially worse using the simulated NISAR data at 10 m (J-statistic: 0.40, Kappa value: 0.36) compared to those made at 30 m (J-statistic: 0.55, Kappa: 0.49) and 100 m (J-statistic: 0.58, Kappa: 0.50). All performance metrics considered, there was little difference between a 30 and 100 m product, showing that accurate cropland products could also be produced at finer spatial resolutions.

Optimal CV_{thr} values varied considerably depending on which spatial resolution and the performance metric was used. Optimal CV_{thr} values decreased monotonically with spatial resolution, for every performance metric. At all spatial resolutions, optimal thresholds for overall accuracy were smallest, followed by those obtained for Kappa and the J-statistic. For example, using $CV_{thr} = 0.5$ at 100 m, overall accuracy was only 65%. Using $CV_{thr} = 0.26$ instead, improved overall accuracy by 15%. Therefore, it will likely be necessary to consider CV_{thr} values different from 0.5 to not only for maximizing the product's accuracy, but also for achieving the 80% accuracy mission requirement.

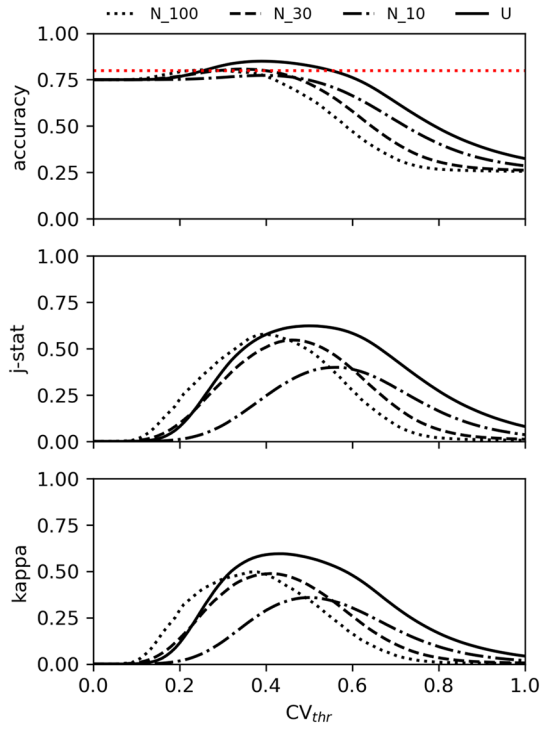

Figure 7. Plot of three performance metrics (Accuracy, J-statistic, and Cohen's Kappa) versus crop/noncrop delineating threshold for UAVSAR at 10 × 10 m ("U") and simulated NISAR at 10 × 10 m ("N_10"), 30 × 30 m ("N_30") and 100 × 100 m ("N_100") spatial resolution. UAVSAR, Uninhabited Aerial Vehicle SAR.

Author Contributions

S. Kraatz and P. Siqueira designed the study. S. Kraatz and S. Rose obtained and processed the data sets. S. Kraatz analyzed and interpreted the results. S. Kraatz wrote the manuscript with support of S. Rose, M. Cosh, N. Torbick, X. Huang, and P. Siqueira.

Data Availability Statement

The data supporting the conclusions of this manuscript are freely available. The UAVSAR and simulated NISAR data are available from NASA JPL at https://uavsar.jpl.nasa.gov/cgi-bin/data.pl using the search term "27900." This search results in data listed under Stoneville, MS according to dates. For each date, there will be one data labeled simulated NISAR data and another as UAVSAR data. To find the data sets used in this study, the user should then select the type of data and the data and click on the corresponding link. Within the new page that is opened, the user is then able to download the file of the same name as provided in Table 2.

The CDL layer used in this study may be obtained through CropScape at https://nassgeodata.gmu.edu/CropScape/, by selecting 2019 in

the Cropland Data Layers category. This layer can then be subset to the same extent used in this study by providing a polygon shapefile having these corner coordinates in the UTM 15 (EPSG: 32615) projection: (1) 688956.0519184095 3704431.626407656, (2) 699413.1631274965 3704131.230218298, (3) 699489.5084457516 3697620.260962357, (4) 688964.5833900904 3697847.959192821.

Acknowledgments
This work was supported by NASA grants 80NSSC19K1497, NNX16AK59G and USDA ARS grant 58-8042-8-072. The authors declare no conflict of interest. The authors thank the anonymous reviewers for their helpful comments.

References

Becker-Reshef, I., Barker, B., Humber, M., Puricelli, E., Sanchez, A., Sahajpal, R., et al. (2019). The GEOGLAM crop monitor for AMIS: Assessing crop conditions in the context of global markets. *Global Food Security*, *23*, 173–181.

Betbeder, J., Fieuzal, R., & Baup, F. (2016). Assimilation of LAI and dry biomass data from optical and SAR images into an agro-meteorological model to estimate soybean yield. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(6), 2540–2553.

Boryan, C., Yang, Z., Mueller, R., & Craig, M. (2011). Monitoring US agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International*, *26*(5), 341–358. https://doi.org/10.1080/10106049.2011.562309

Chapman, B., Siqueira, P., Saatchi, S., Simard, M., & Kellndorfer, J. (2019). *Initial results from the 2019 NISAR Ecosystem Cal/Val Exercise in the SE USA* (pp. 8641–8644). Paper presented at IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium. https://doi.org/10.1109/IGARSS.2019.8899227

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Erten, E., Lopez-Sanchez, J. M., Yuzugullu, O., & Hajnsek, I. (2016). Retrieval of agricultural crop height from space: A comparison of SAR techniques. *Remote Sensing of Environment*, *187*, 130–144.

Ferrazzoli, P., Paloscia, S., Pampaloni, P., Schiavon, G., Sigismondi, S., & Solimini, D. (1997). The potential of multifrequency polarimetric SAR in assessing agricultural and arboreous biomass. *IEEE Transactions on Geoscience and Remote Sensing*, *35*(1), 5–17.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., et al. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment*, *83*(1–2), 287–302.

Fritz, S., See, L., Bayas, J. C. L., Waldner, F., Jacques, D., Becker-Reshef, I., et al. (2019). A comparison of global agricultural monitoring systems and current gaps. *Agricultural Systems*, *168*, 258–272.

Habibzadeh, F., Habibzadeh, P., & Yadollahie, M. (2016). On determining the most appropriate test cut-off value: The case of tests with continuous results. *Biochemia Medica: Biochemia Medica*, *26*(3), 297–307.

Huang, X., Liao, C., Xing, M., Ziniti, B., Wang, J., Shang, J., et al. (2019). A multi-temporal binary-tree classification using polarimetric RADARSAT-2 imagery. *Remote Sensing of Environment*, *235*, 111478.

JPL, N. (2020). *Simulated NISAR data*. Retrieved from https://uavsar.jpl.nasa.gov/science/documents/nisar-sample-products.html

Kraatz, S., Siqueira, P., & Rose, S. (2020). *ISCE docker tools: Automated radiometric terrain correction and image coregistration of UAVSAR MLC data*. Paper presented at IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., & Merchant, J. W. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, *21*(6–7), 1303–1330.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica: Biochemia Medica*, *22*(3), 276–282.

McNairn, H., & Brisco, B. (2004). The application of C-band polarimetric SAR for agriculture: A review. *Canadian Journal of Remote Sensing*, *30*(3), 525–542.

McNairn, H., & Shang, J. (2016). A review of multitemporal synthetic aperture radar (SAR) for crop monitoring. In *Multitemporal remote sensing* (pp. 317–340). Springer.

NISAR Science Team. (2020). *NISAR science users' handbook*. Retrieved from https://nisar.jpl.nasa.gov/files/nisar/NISAR_Science_Users_Handbook.pdf

Rosen, P. A., Hensley, S., Wheeler, K., Sadowy, G., Miller, T., Shaffer, S., et al. (2006). *UAVSAR: A new NASA airborne SAR system for science and technology research* (p. 8). Paper presented at 2006 IEEE Conference on Radar, IEEE.

Small, D. (2011). Flattening gamma: Radiometric terrain correction for SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(8), 3081–3093.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., et al. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, *120*, 9–24.

Ulander, L. M. H. (1996). Radiometric slope correction of synthetic-aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, *34*(5), 1115–1122.

Whelen, T., & Siqueira, P. (2017). *Time series analysis of L-Band SAR for agricultural landcover classification* (pp. 5342–5345). Paper presented at 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE.

Whelen, T., & Siqueira, P. (2018). Coefficient of variation for use in crop area classification across multiple climates. *International Journal of Applied Earth Observation and Geoinformation*, *67*, 114–122.

Wiseman, G., McNairn, H., Homayouni, S., & Shang, J. (2014). RADARSAT-2 polarimetric SAR response to crop biomass for agricultural production monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(11), 4461–4471.

Yan, L., & Roy, D. P. (2016). Conterminous United States crop field size quantification from multi-temporal Landsat data. *Remote Sensing of Environment*, *172*, 67–86.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35.