Performance Evaluation of UAVSAR and Simulated NISAR Data for Crop/Non-crop Classification over Stoneville, MS

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

[1]Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003, USA

[2]USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705, USA

[3]Applied Geosolutions, Durham, NH 03857, USA

Corresponding author: Simon Kraatz (skraatz@umass.edu)

Key Points:

- Crop/non-crop classifications were evaluated using UAVSAR and simulated NISAR data using three spatial resolutions and performance metrics
- Optimal crop/non-crop delineating thresholds monotonically decreased with spatial resolution for each performance metric

Keywords: UAVSAR, NISAR, Cropland

Length: 7 figures + 4 tables = 11 units, plus 5416 words (including figure and table captions) = 11 units for a total of 22 out of 25 publication units.

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 UAVSAR and simulated NISAR data (129A). The NISAR algorithm uses the coefficient of variation (CV) to perform crop/non-crop 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/non-crop delineating CV thresholds (CV_{thr}) ranging from 0 to 1 in 0.01 increments. All but the 10 m 129A product exceeded the 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 was suboptimal (65%) and that optimal CV_{thr} values monotonically decreased with decreasing spatial resolution.

1 Introduction

Timely and accurate large-scale data on agricultural activity is 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 (<4m) to moderate (<30m) spatial resolutions, frequent revisit time, and open access, operational coverage for large spatial extents (Becker-Reshef et al 2019). Because a majority of agricultural fields are just over two 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 (H 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 (Fritz et al., 2019).

Since retrievals made by Synthetic Aperture Radar (SAR) systems now meet or exceed the above needs, they are well-suited for large-scale agricultural monitoring. Spaceborne SAR, such as the European Space Agency's (ESA) Sentinel-1 can map Earth once every 6 to 12-days at moderate spatial resolution (Torres et al., 2012). SAR data provides 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 (e.g. at C- band or longer), 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 datasets 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 (~12 day repeat) mainly using L-band but also features S-band over selected areas, mainly over 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), while other areas that do not change as much over time are considered non-crop, such as built or forest. Crop and non-crop 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 non-crop 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 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, respectively for AgriSAR L-band, AgriSAR C-band and ALOS. 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/non-crop classifications: it was set according to (1) the threshold of maximum separation of histograms of CV values for crop and non-crop classes, (2) the threshold corresponding to the largest Youden J-statistic value, 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% to 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 realistic CV values. Furthermore, computed CV values - and the ensuing CV_{thr} values to be used - will also depend on factors such as the frequency of observation, cropping practices and landscape heterogeneity, and the spatial resolution to which thresholding will be applied. 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 "UAVSAR AM/PM Campaign" throughout summer 2019, consisting of repeat flights every two weeks over forests and agricultural areas in the Southeast at local times of 6 a.m. and 6 p.m. (Chapman et al., 2019). Data obtained during this campaign was also used to generate simulated NISAR products that more closely resemble the quality of data to be collected by NISAR (JPL, 2020).

Furthermore, data resembling that to be collected by NISAR – the 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/non-crop 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 between the UAVSAR data compared to the simulated NISAR products on 10 m x 10 m grids. For simulated NISAR data, three spatial resolutions were investigated: (1) a 10 m spatial resolution that approximately corresponds to the 12 m x 6.2 m slant range grid spacing of the simulated NISAR MLC product; (2) a 100 m spatial resolution corresponding to that at which NISAR's Cropland Area product is to be evaluated; and (3) an in-between spatial resolution, here 30 m, to match the resolution of the annual USDA Cropland Data Layer (CDL) to be used as ground truth (Boryan et al., 2011). Our hypotheses are: (1) UAVSAR data would perform the best due to these observations having the highest quality, (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, Datasets 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.

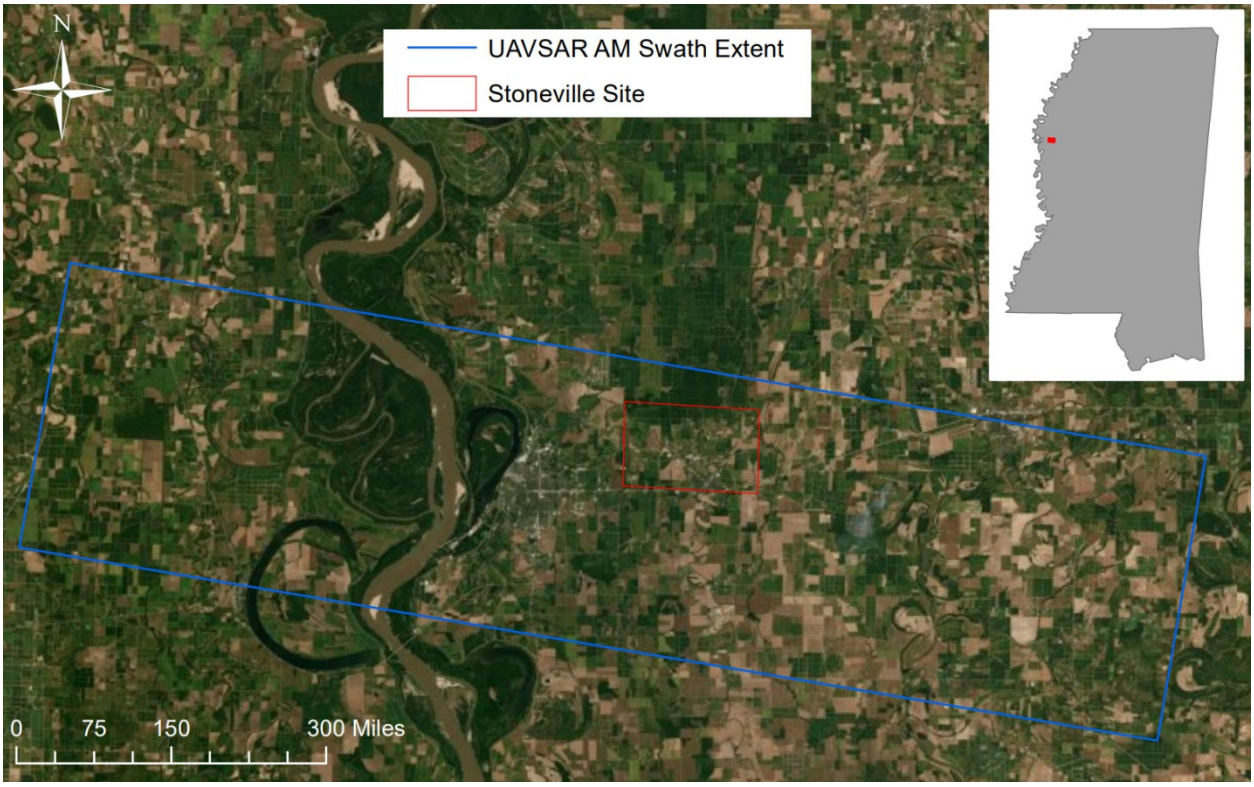

Figure 1. The study area is located in the western part of Mississippi, close to Stoneville. The swath of SAR data collected during the 6 a.m. UAVSAR flight (27900) is indicated in blue. The red box indicates the region of interest this study used for crop/non-crop classification. The site falls within a 33 to 47 degree incidence angle range, as to match those of NISAR.

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 m x 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 shows the original and the binary crop/non-crop result obtained from assigning CDL layers to crop/non-crop (Table 1). The image consists of a total of 718828 pixels (1054 columns and 682 rows) of which 501378 (69.7 %) and 166978 (23.2 %) pixels are crop and non-crop, 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 don’t 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. After excluding masked cells from comparisons 668356 'valid' pixels remain, with a percent breakdown of 75% for crop and 25% for non-crop. Because the CDL is categorical and posted at 30 m x 30 m grids, to make comparisons at 10, 30 and 100 m, we subdivided each CDL grid into nine 10 m x 10 m grids.

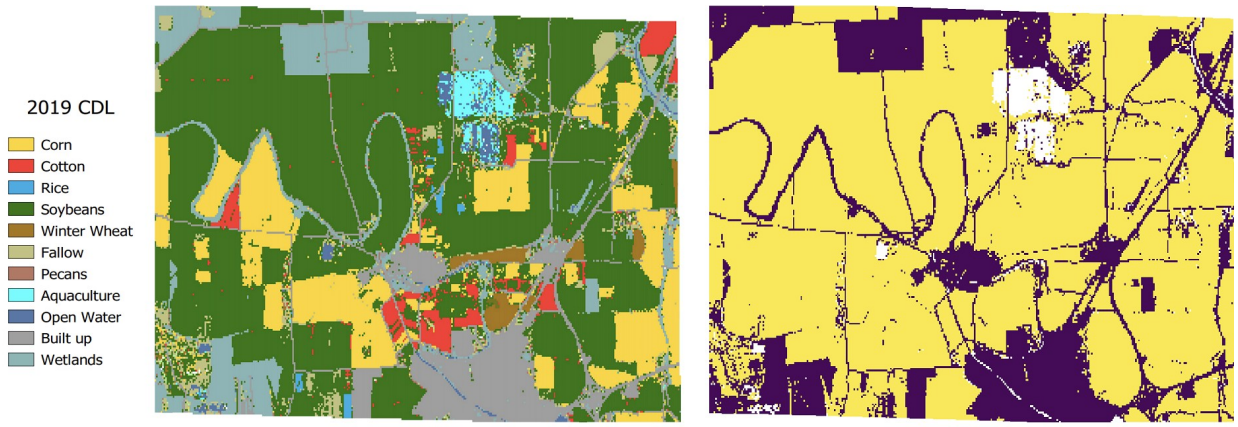

Figure 2. The 2019 USDA/NASS Crop Data Layer at study site (a), and the binary crop (yellow) non-crop (purple) classification (b) according to Table 1. Some CDL classes for which CV approach is known to not perform applicable were masked

Table 1. Rules used for re-classifying the 2019 USDA NASS Crop Data Layer into binary crop/ non-crop grids.

CDL Class Ranges	CDL Class Type	Crop/Non-Crop/Masked
1 to 60	Crop (e.g. Cotton, Rice …)	Crop
61 to 65	Non-crop (e.g. Fallow, Forest …)	Non-Crop
66 to 80	Tree crops (e.g. Cherry, Peach …)	Masked
81 to 109	Other (e.g. Water, Wetlands …)	Masked
110, 112 to 195	NLCD Classes (e.g. Developed, Forest)	Non-Crop
196 to 255	Crops (e.g. Carrots, Garlic …)	Crop
111	Open Water	Masked

2.3 UAVSAR and Simulated NISAR data

SAR data was collected by the Uninhabited Aerial Vehicle SAR (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 (~2 weeks) at L-band (1.26 GHz), occurring at NISAR observing times of 6 a.m. and 6 p.m. (local time). The campaign's main purpose was 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 (SNR) and fully polarimetric SAR observations (HH, HV, VV) 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 multi look complex (MLC) data in slant-range coordinates and ground-range projected (GRD) format (https://uavsar.jpl.nasa.gov/). Single look complex (SLC) data and interferometric datasets are also available for download, but only for some of UAVSAR's flight lines and dates.

Simulated NISAR data are generated from the UAVSAR observations. UAVSAR observations have relatively higher SNR 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 consists of two bands, referred to as Frequency A (for HH and HV) and Frequency B (for VH and VV). Specifically, we use Frequency A of observing mode 129 (129A) because of (1) its relatively higher spatial resolution compared to 129B and (2) it being NISAR's predominant observing mode covering most of the globe.

The simulated datasets 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 co-registered and posted on a common grid, facilitating time series analysis. Furthermore, the GRD product also comes with a radiometric terrain correction (RTC) calibration file. Although this makes the simulated GRD products attractive for processing, we opted to use the MLC data for the comparisons of UAVSAR and simulated NISAR data: (1) MLC data are available for all the UAVSAR flight lines and (2) we're able to ensure that an identical workflow is used for processing both UAVSAR and simulated NISAR data with respect to image co-registration and RTC.

The UAVSAR and simulated NISAR MLC data are the result of cross multiplying and multi looking the SLCs. The cross multiplications between the HH, HV, VV polarized yielding three complex valued and real valued results in linear units of power. Because previous studies indicated that the HV cross product performed best for agricultural crop/non-crop classifications, we only tested crop/non-crop delineation with the HVHV cross product (Heather McNairn & Shang, 2016; Rose et al., 2020; Whelen & Siqueira, 2017).

The study area described in Section 2.1 is covered by the UAVSAR Stoneville 6 a.m. flight line (27900). Seven approximately bi-weekly 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. UAVSAR and simulated NISAR datasets used as inputs (2019).

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

2.4 SAR Data Processing

The input data (UAVSAR and 129A) was subjected to further processing steps to improve the datasets with respect to image co-registration, RTC and re-projection from the source σ_0 to a γ_0 scattering plane (Figure 3).

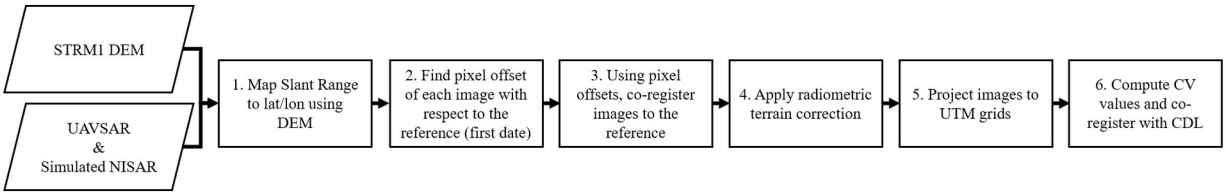

Figure 3. Flowchart showing the six data processing steps. Steps 1 through 5 were applied using the InSAR Scientific Computing Environment software via an automated workflow (ISCE Docker Tools).

First, we obtain the geometry of the reference image to which all others will be co-registered. The reference image was set as the earliest of the UAVSAR AM/PM campaign data for a given flightline (i.e. 6 June 2019 for UAVSAR Flightline 27900). This step also required a digital elevation model, and we used and the Shuttle Radar Topography Mission 30 m (SRTM1) data. Second, the pixel offsets of each subsequent image relative to the reference are determined. Third, using the pixel offsets as inputs, imagery is co-registered to the reference image. Fourth, the geometry information obtained in step 1 (e.g. the local and flat earth incidence angles) are used to apply RTC and output results in the γ_0 plane (Small, 2011; Ulander, 1996). Fifth, the slant range MLCs were projected to UTM grids at 10 m x 10 m, 30 m x 30 m and 100 m x 100 m spacing. Coarser spatial resolutions were investigated as the 129A MLC product appeared to be quite noisy due to speckle. Spatial resolutions were selected according to being consistent with the finest NISAR products (around 10 m), that of the CDL (30 m) and that at which NISAR's crop/non-crop classifications are to be evaluated (100 m). Sixth, we then computed CV values at each spatial resolution (Section 3.1). Because the CDL (30 m x 30 m) and CV grids (10 m, 30 m, 100 m) have different spatial resolutions these datasets were then subdivided into 10 m x 10 m grids for making comparisons. Also, because the 10 m CV grids were shifted relative to the CDL grids that had been subdivided into 10 m grids, we interpolated the CV datasets to the CDL grids. Therefore, all performance evaluations described in Section 3 are made using 10 m x 10 m grids.

Steps 1 through 5 are incorporated in our automated workflow using the InSAR Scientific Computing Environment (ISCE) Docker Tools (IDT) (Kraatz et al., 2020), available at https://github.com/UMassMIRSL/isce_docker_tools. IDT is a dockerized version of ISCE 2.3.1, additionally modified for automated processing of UAVSAR Multi Look Complex (MLC) time series data and applies processing Steps 1 through 5 in Figure 3.

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/non-crop classifications when evaluated at 100 m x 100 m spatial resolution. The algorithm is based on the temporal CV approach. The basic premise of the CV approach for delineating crop and

non-crop areas is that actively cultivated lands (i.e. crop) experience more substantial change over time as compared to unmanaged lands (i.e. non-crop) (Section 3.1). 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/non-crop delineation. Therefore, we iterate through a range of CV_{thr} values, from 0.00 through 1.00 in 0.01 increments (Section 3.2) and classify each valid pixel of the CV image as crop or non-crop. For each CV_{thr} value, we use the results of the confusion matrix (Section 3.3) to compute three different accuracy metrics: the overall accuracy (Section 3.4), Youden's J-statistic (Section 3.5) and Cohen's Kappa (Section 3.6).

3.1 Coefficient of Variation (CV)

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

$$CV_{HV} = \frac{\sigma_{HV}}{\mu_{HV}} \tag{1}$$

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

3.2 Crop/non-crop Classification

Crop/non-crop classification is determined using a fixed CV_{thr} value, applied to the region of interest (ROI), where

$$CV_{pixel} \begin{cases} ¿\, CV_{thr\ for\ ROI}, non-crop \\ \geq CV_{thr\ for\ ROI}, crop \end{cases} \tag{2}$$

3.3 Confusion Matrix

Our crop/non-crop results (Section 3.2) are compared to the ground truth (the CDL, Section 2.2) using a confusion matrix. The confusion matrix tabulates the number of pixels for which both datasets agreed on the crop and non-crop 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 non-crop but the CDL did not (false negative, FN).

Table 3. Confusion matrix for crop/non-crop classification.

Model (the SAR-based classifications)	Observed (the CDL)	
	Crop	Non-Crop
Crop	True Positive (TP)	False Positive (FP)
Non-crop	False Negative (FN)	True Negative (TN)

3.4 Overall Accuracy

Overall Accuracy is calculated from the confusion matrix (Section 3.3) as:

$$accuracy = \frac{100 * (TP + TN)}{(TP + FP + FN + TN)} \quad (3)$$

3.5 Youden's J-statistic

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 (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

$$J = Sensitivity + Specificity - 1 \quad (4)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

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 ROC curve. The ROC curve is the result from plotting of Sensitivity vs. 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.

3.6 Cohen's Kappa

Cohen's Kappa indicates the threshold for which two datasets 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 (Section 3.3):

$$Kappa = \frac{p_O - p_e}{1 - p_e} \quad (7)$$

where p_O is the observed proportionate agreement, given by

$$p_O = \frac{TP + \mathrm{TN}}{TP + FP + FN + TN} \quad (8)$$

and p_e is the overall random agreement probability, given by

$$p_e = p_Y + p_N \quad (9)$$

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

$$p_Y = \frac{(TP+FP)*(TP+FN)}{(TP+FP+FN+TN)^2} \quad (10)$$

$$p_N = \frac{(FN+TN)*(FP+TN)}{(TP+FP+FN+TN)^2} \quad (11)$$

Kappa may range between -1.0 to 1.0. Values below zero indicate poor agreement while 1.0 represents a perfect agreement between the CDL and the SAR-based crop/non-crop 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-d). There is a considerable amount of speckle in the simulated UAVSAR data at 10 m (Figure 4b): intra-field CV variations are quite large, which may be problematic for accurate crop/non-crop classification. The speckle is not entirely surprising as the simulated NISAR MLCs had only been multi-looked twice in azimuth with respect to the SLC and have a 12 m (azimuth) x 6.2 m (range) posting. For comparison, UAVSAR's SLC resolution is 0.6 m (azimuth) x 1.67 m (range) and the MLC product had been multi-looked 12 x in azimuth and 3 x in range for a posting of 7.2 m by 5 m.

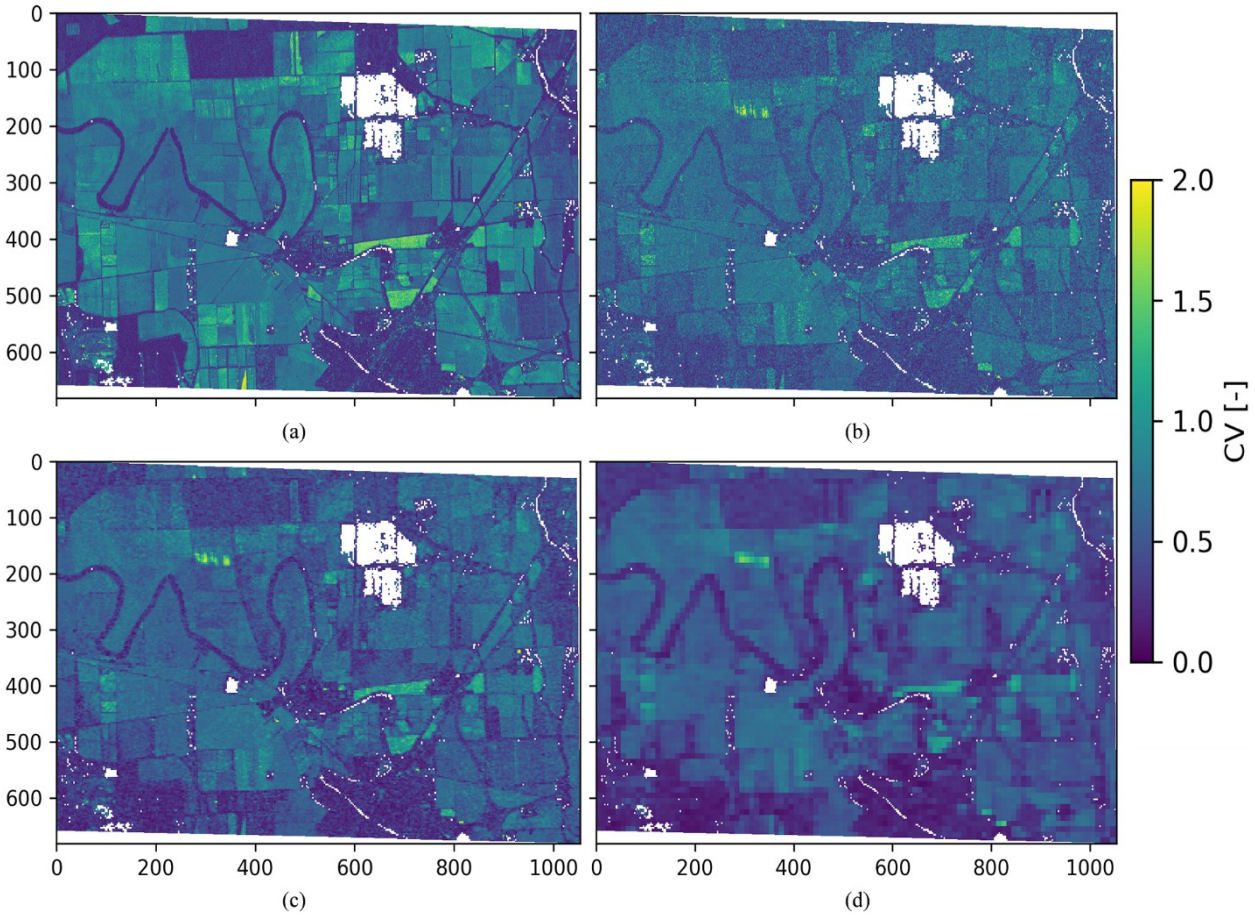

Figure 4. CV values for (a) UAVSAR 10 m x 10 m and (b)-(d) simulated NISAR data interpolated to 10, 30 and 100 m grids. Some CDL land covers (e.g. aquaculture) were masked according to Table 1 and are shown in white.

In general, features are more washed out for the simulated NISAR data, especially at 10 m (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. Although CV values decrease at the 30 and 100 m postings, contrast between low and high CV areas are much improved over the 10 m posting. This should lead to relatively easier crop/non-crop delineation when applying a single CV_{thr} value over the scene.

ROC curves also provide valuable insight on the relative performance of crop/non-crop classifications using the different input datasets (Figure 5). ROC curves of the UAVSAR (Figure 5a) and 10 m simulated NISAR classifications (Figure 5b) respectively have the greatest and least separation from the line of no discrimination (the 1:1 line). 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 m 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.

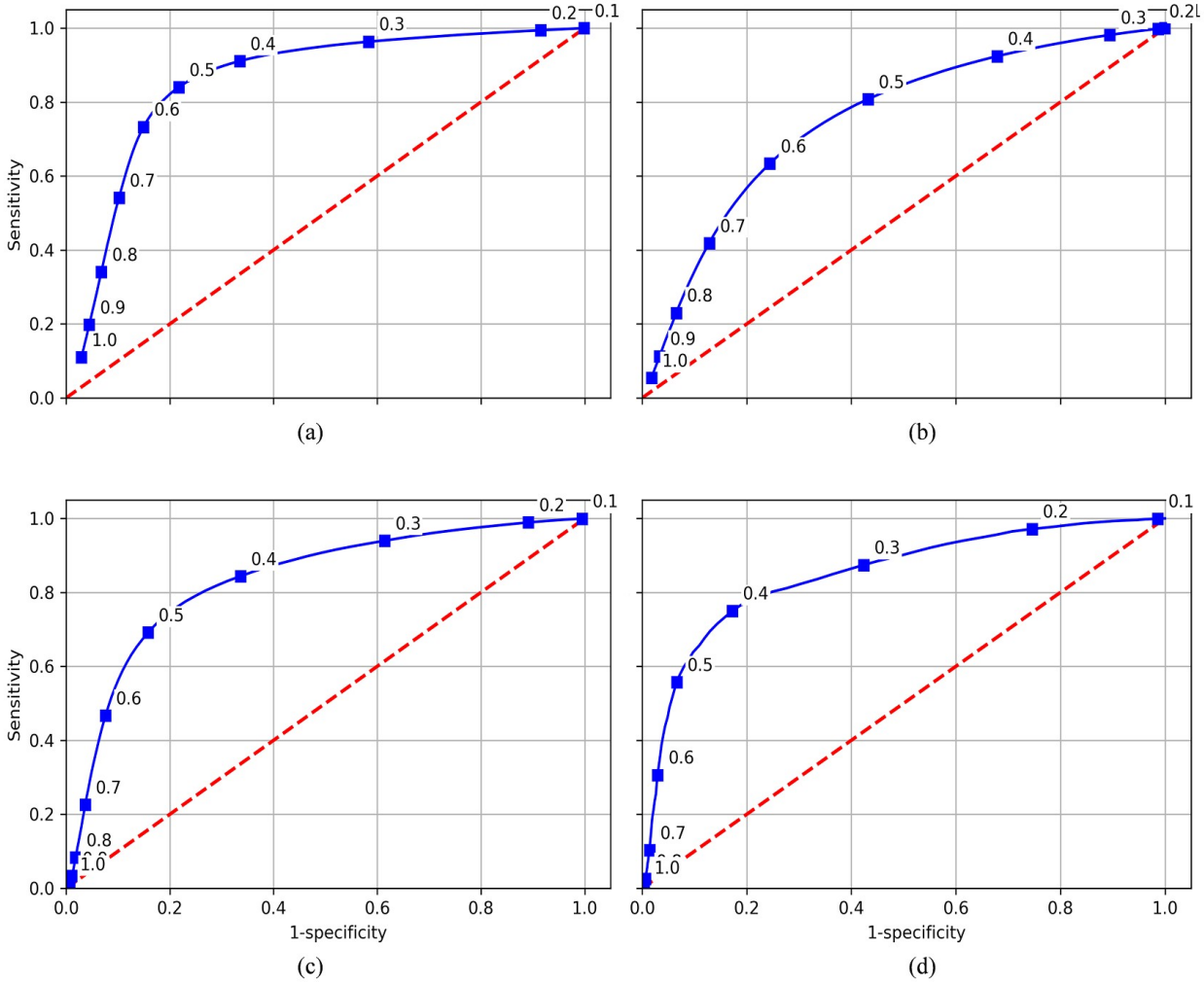

Figure 5. Receiver operating characteristic curves for (a) UAVSAR at 10 m x 10 m and (b)-(d) simulated NISAR at 10 m x 10 m, 30 m x 30 m and 100 m x 100 m spatial resolution. The marked points on the curve indicate the Sensitivity and 1-Specificity values for the annotated value of the crop/non-crop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

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 datasets, 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 performs the best, 10 m simulated NISAR the worst,

and that the 30 m 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.

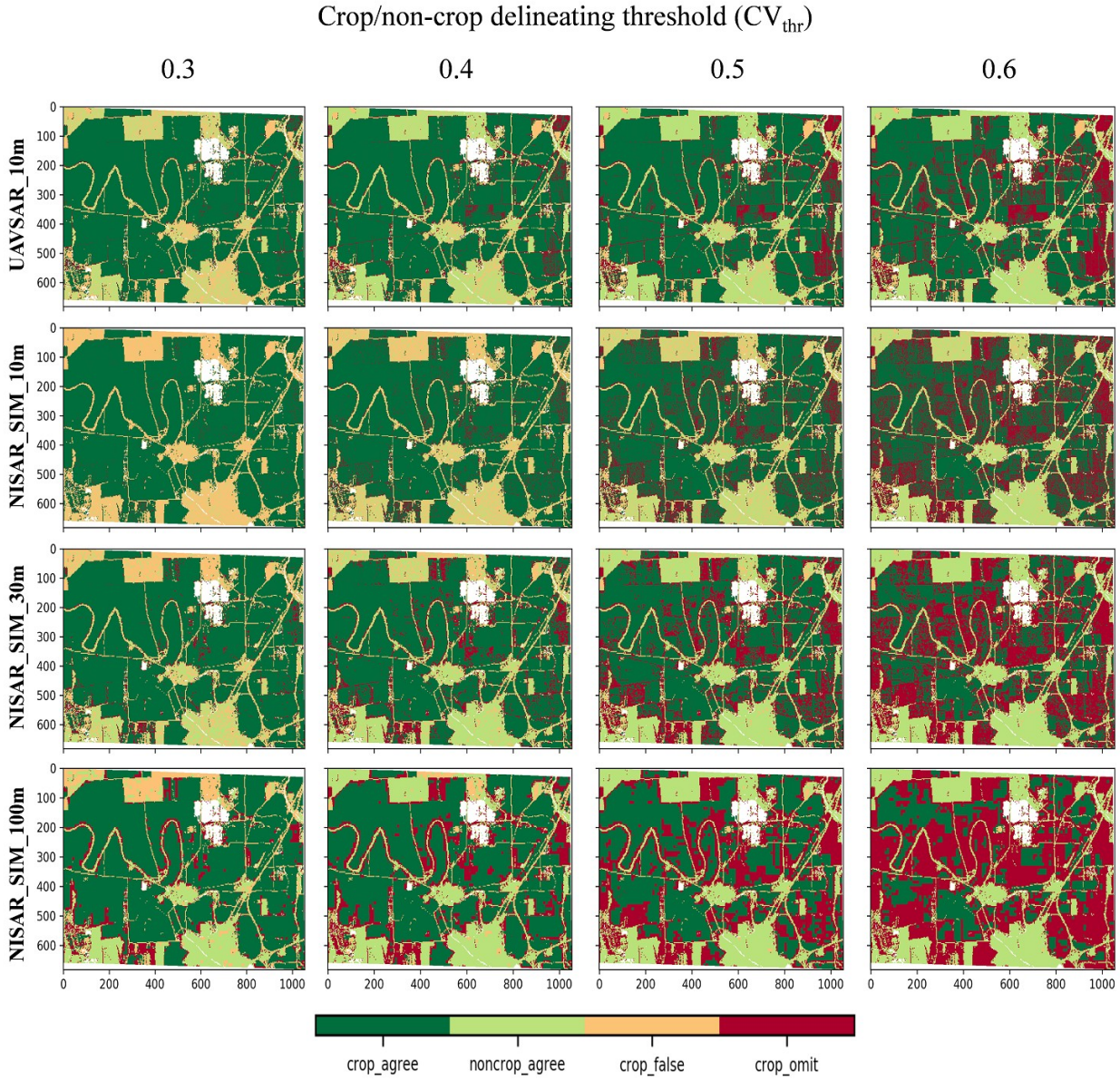

Figure 6. Crop/non-crop confusion matrix results for the UAVSAR (10 m x 10 m) and simulated NISAR (10 m x 10 m, 30 m x 30 m and 100 m x 100 m) products (rows) as function of crop/non-crop delineating threshold (CV_{thr}, columns). Results show that to maintain the best classification accuracy with spatial resolution, it is necessary to employ smaller CV_{thr} values as the spatial resolutions become coarser.

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 non-crop 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 non-crop classes: at a CV_{thr} of 0.57, 48972 out of 166978 non-crop grids (29%) are misclassified as crop, while 154095 out of 501378 crop grids (31%) are misclassified as non-crop. 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.

Table 4. The unbracketed values indicate the optimal CV_{thr} for each accuracy metric. The bracketed values show the metric's maximum value, occurring at the given CV_{thr}. N refers to the simulated NISAR data at 10, 30 and 100 m grids.

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)

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 towards 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 betweeen them.

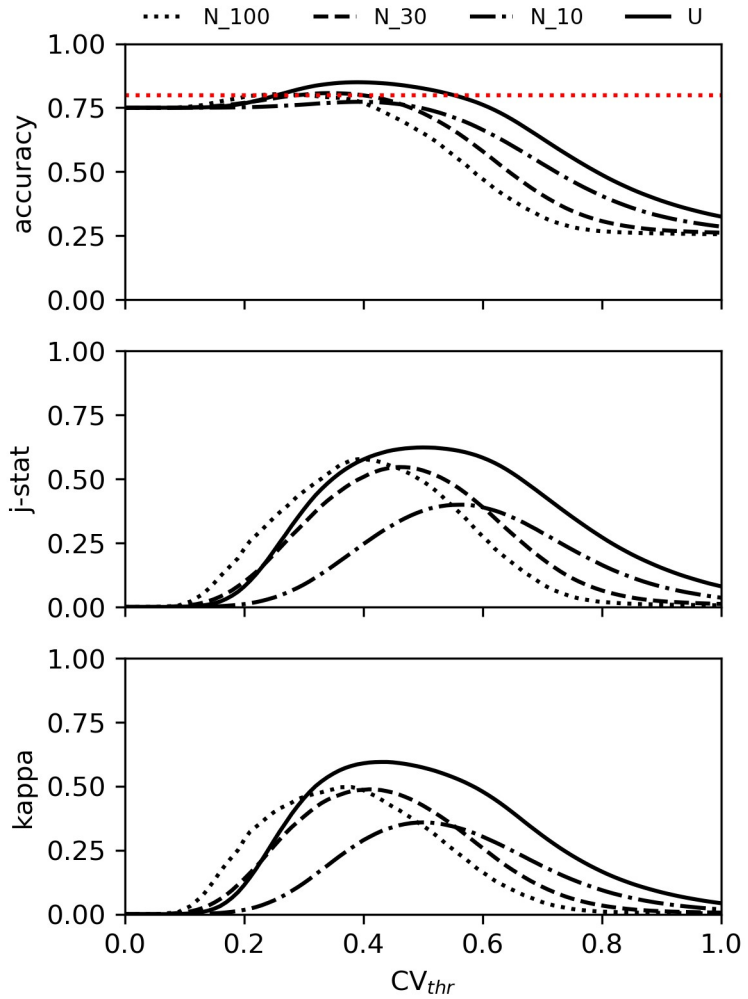

Figure 7. Plot of three performance metrics (Accuracy, J-statistic and Cohen's Kappa) versus crop/non-crop delineating threshold for UAVSAR at 10 m x 10 m ('U') and simulated NISAR at 10 m x 10 m ('N_10'), 30 m x 30 m ('N_30') and 100 m x 100 m ('N_100') spatial resolution.

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. We also note that there appears to be a relatively large range 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 there 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 one hectare scale, the simulated NISAR products meet these goals.

Rose et al. (2020) applied the same CV approach but using Sentinel-1 data posted at 150 m x 150 m and year round comparisons for 2017 (Rose et al., 2020). Optimal CV_{thr} values were computed for one hundred 1° by 1° areas over agricultural regions throughout the Contiguous United States. The study reported on a strong geographic dependence on optimal CV_{thr} values, with values falling closer to 0.5 in the central US and closer to 0.3 near coastal regions. While

that study did not include any region in Mississippi, the nearest 1° by 1° study areas in Arkansas and Georgia also indicated optimal CV_{thr} values falling somewhere between a 0.2 to 0.4 range. Thus, although there are many fundamental differences between this and Rose et al. (2020) (e.g. grid spacing and frequency), results reported here are consistent with the regionally dependent optimal CV_{thr} values reported in Rose et al. (2020). While further study is needed to confirm the consistency between these results more thoroughly, there appears to be some potential of using the Sentinel-based CV_{thr} values as initial estimates when using NISAR'data for computing cropland area.

While our findings are generally consistent with our initial hypotheses, we note that classification performance only slightly decreased between using 30 m (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.

5 Conclusions

This main goal of this work was to evaluate the proposed algorithm for generating crop/non-crop classifications using NISAR data. Crop/non-crop 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/non-crop classifications are to be evaluated (100 m). Using UAVSAR at 10 m resolution, we achieved crop/non-crop 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 m x 6.2 m MLCs on 10 m x 10 m grids, resulting in misclassifications. And while results were not substantially better using 30 m (81%) or 100 m (80%) grids, we found that a fairly large range of CV_{thr} values (0.2 to 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 m 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. Because this study's results yielded an optimal CV_{thr} values similar to what was expected according to Rose et al. (2020), it may be possible to use that study's CV_{thr} values to inform on more suitable thresholds when computing cropland area from NISAR data.

Author Contributions

S.K. and P.S. designed the study. S.K. and S.R. obtained and processed the datasets. S.K. analyzed and interpreted the results. S.K. wrote the manuscript with support of S.R., M.C., N.T., X.H and P.S.

Acknowledgements

This work was supported by NASA grant 80NSSC19K1497. The authors declare no conflicts of interest. The authors thank the anonymous reviewers for their helpful comments.

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 labelled simulated NISAR data and another as UAVSAR data. To find the datasets 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.

References

Becker-Reshef, I; Barker, B.; Humber, M.; Puricelli, E.; Sanchez, A.; Sahajpal, R.; McGaughey, K.; Justice, C.; Baruth, B.; Bingfang, W.; Prakash, A.; Abdolreza, A.; Jarvis, I., "The GEOGLAM Crop Monitor for AMIS: Assessing Crop Conditions in the Context of Global Markets," Global Food Security, vol. 23, pp. 173-181, Dec. 2019

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. *International Geoscience and Remote Sensing Symposium* (pp. 8641–8644).

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.

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. *IEEE International Geoscience and Remote Sensing Symposium*.

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, Heather, & 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

Rose, S., Kraatz, S., Siqueira, P., & Kellndorfer, J. (2020). Evaluating NISAR's cropland area algorithm over the conterminous United States using Sentinel-1 data: performance of globally fixed vs. regionally optimized crop/non-crop threshold. *Remote Sensing*.

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. In *2006 IEEE Conference on Radar* (pp. 8-pp). 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. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 5342–5345). 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.