

Evaluating NISAR's cropland mapping algorithm over the conterminous United States using Sentinel-1 data



Shannon Rose ^a, Simon Kraatz ^{a,*}, Josef Kellndorfer ^b, Michael H. Cosh ^c, Nathan Torbick ^d, Xiaodong Huang ^d, Paul Siqueira ^a

^a Department of Electrical and Computer Engineering, 113 Knowles Engineering Building, 151 Holdsworth Way, University of Massachusetts, Amherst, MA 01003, USA

^b Earth Big Data LLC, Woods Hole, MA 02543, USA

^c USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705, USA

^d Applied Geosolutions, Durham, NH 03824, USA

ARTICLE INFO

Editor: Dr. Marie Weiss.

Keywords:

SAR
Coefficient of variation
Time series
Sentinel-1
Crop classification
Agriculture

ABSTRACT

Accurate knowledge of the distribution, breadth and change in agricultural activity is important to food security and the related trade and policy mechanisms. Routine observations afforded by spaceborne Synthetic Aperture Radar (SAR) allows for high-fidelity monitoring of agricultural parameters at the field scale. Here we evaluate the approach to be used for generating NASA's upcoming NASA ISRO SAR (NISAR) mission's L-band cropland product using Sentinel-1C-band data. This study uses all ascending Sentinel-1A/B data collected over the conterminous United States in 2017 to compute the coefficient of variation (CV) at 150 m × 150 m resolution and evaluates the overall accuracy (OA) of CV-based crop/non-crop classifications at 100 one-by-one degree tiles. We calculate accuracies using two approaches: (a) using a literature-recommended constant CV threshold of 0.5 ($CV_{thr,0.5}$) and (b) determining optimal CV thresholds for every tile using Youden's *J* statistic (YJS), $CV_{thr,YJS}$. These accuracy comparisons are important for determining (1) the viability of using a computationally inexpensive and straightforward approach for cropland classification over large spatial scales/diverse land covers (i.e., can accuracies $\geq 80\%$ be routinely achieved?), (2) how closely $OA_{0.5}$ compares to the performance ceiling (OA_{YJS}). This information will help determine whether approach (a) is appropriate and how much potential room of improvement there could be in modifying it. Results for $OA_{0.5}$ and OA_{YJS} are 81.5% and 86.8%, respectively. A breakdown by census geographic region, showed that $OA_{0.5}$ (OA_{YJS}) exceeded 80% (90%) in the South and Midwest, but was only 76.1% (73.5%) in the West. The improvement in OA_{YJS} mainly stems from tiles with >40% crop prevalence having about 10% greater OA values. To better examine the potential of the approach for land cover classification, results of approach (b) were also stratified by crop. Approach (b) accurately detected most non-crop classes as non-crop (>80%), but with low OA_{YJS} values for grasslands/pasture, especially in the West. CV values for crop were distinct from non-crop indicating that the approach is suitable for crop/non-crop classifications. Because results CV values have substantial overlap within crop/non-crop classes, indicating the approach is poorly suited for land cover classifications. We also detected a strong geographic dependence of $CV_{thr,YJS}$: values ranged from about 0.2 at the coasts and gradually increase to about 0.6 in the Central United States, most often falling close to 0.3 and 0.5.

1. Introduction

Cropland area is an Essential Agricultural Variable and priority metric identified by the Group of Earth Observations Global Agricultural Monitoring initiative (Becker-Reshef et al., 2019). Identification of actively used cropland is particularly important because it is a key indicator of agricultural productivity and thus important to food security

(Yin et al., 2020). This application can greatly benefit from using dense time series, because cropland may experience substantial change at sub-weekly time scales due to agricultural practices such as tilling, sowing, growing and harvesting. Access to operational and open access moderate spatial resolution SAR and optical observations have provided observations that are useful in characterizing active cropland over large areas at seasonal scales. Both types of observations allow for spatial

* Corresponding author.

E-mail address: skraatz@umass.edu (S. Kraatz).

resolutions are on the order of tens of meters and therefore provide information at the field scale.

Dense time series can be derived from both optical and SAR data, but there are some fundamental differences and limitations. For example, the availability of optical imagery is limited by atmospheric effects (i.e., clouds, fog and smoke) and solar illumination. And while SAR data are relatively less impacted by atmospheric effects, it may require careful pre-processing to account for speckle noise, viewing geometry (i.e., incidence angle) and terrain. Today, both SAR (e.g. Sentinel-1) and optical data sources (e.g. MODIS, VIIRS, Sentinel-2) are increasingly available in global mapping modes with high spatial and temporal resolution to facilitate the generation of detailed land cover and land cover change data products (Abdiyan et al., 2016; Belgiu and Csillik, 2018; Friedl et al., 2002; Justice et al., 2013; Teluguntla et al., 2017). Not only are either of the approaches viable – they are also synergistic with one another, often resulting in improved classifications when combined (e.g., Fontanelli et al., 2014; Shang et al., 2008).

Optical approaches of cropland mapping had already been extended to large or global scales through several efforts. USDA's Cropland Data Layer (CDL) provides annual information on agriculture over the contiguous United States (CONUS) at 30 m resolution and also includes detailed breakdowns of crop and non-crop land cover types (Boryan et al., 2011). Other well-known efforts of mapping cropland are global products such as the MODIS-based Global Cropland Extent product and the 30 m Landsat-based Global Food Security-support Analysis Data (GFSAD) (Pittman et al., 2010; Teluguntla et al., 2017). On the other hand, studies of agriculture relying on SAR data only tend to be relatively more limited in spatial extent, often corresponding to regions that are several tens by tens of kilometers in extent (Huang et al., 2021; Kraatz et al., 2021b; Nguyen and Wagner, 2017; Torbick et al., 2017; Whelen and Siqueira, 2018a). Global SAR-based cropland maps are expected to be of value because SAR data have some strengths over optical approaches: it can provide maps based on data that are consistently spaced in time (i.e., not as impacted by clouds), measurements correspond to a different physical quantity (e.g., radar cross section vs. reflectance) and because optical and SAR data can be combined to improve cropland mapping and crop classifications. For example, cropland maps would be useful for screening out non-crop data ahead of training machine learning models for crop classification (e.g., Huang et al., 2021, 2019).

Here we revisit a straightforward and computationally inexpensive approach that had already been successfully used over large regions (consisting many agricultural regions in CONUS), achieving 75% OA (Whelen and Siqueira, 2018b). The temporal CV is calculated using a time-series stack of co-registered SAR imagery (in linear power units) by dividing the standard deviation (calculated over time) over the mean (calculated over time), for every pixel. The classification of individual pixels as crop or non-crop then depends on whether the CV exceeds a certain threshold value (CV_{thr}) or not, respectively. The challenge then lies in deciding on the 'best' value of CV_{thr} for achieving the best OA, or at least one that exceeds some required value (e.g., 80%). This is relatively easy when optimizations are made for individual study areas having small extents (e.g., Kraatz et al., 2021a; Whelen and Siqueira, 2017) but presents a more challenging problem when only a single CV_{thr} value is to be used over many regions, CONUS or globally. It is desirable to use ultimately only one or perhaps a few different CV_{thr} values for the global classifications to limit computational cost. Owing to its simplicity, potential of applicability over large regions (Whelen and Siqueira, 2018b) and low computational cost, the temporal CV approach will also be used for the NISAR Level-2 Cropland Area product.

A review of past work using the temporal CV approach shows that it usually achieves accuracies between 75% to 85% (Huang et al., 2021; Kraatz et al., 2021b, 2021a; Whelen and Siqueira, 2017). An evaluation of crop/non-crop classification accuracy using airborne AgriSAR L- and C- band data indicated that L-band classifications performed best when using cross-polarized data (87%) and that they were substantially better

than classifications made at C-band (78%) (Whelen and Siqueira, 2017). Using the same approach, Huang et al. (2021), also obtained relatively better results when cross-polarized data were used. Only one prior study studied this approach at a large spatial scale, covering many important agricultural areas in CONUS, using 217 approximately 0.5 by 0.5 degree tiles spread over 11 regions (Whelen and Siqueira, 2018b). They tested three approaches for determining CV_{thr} values: a histogram-based approach, a receiver operating characteristic (ROC) curve approach and a constant threshold approach with $CV_{thr} = 0.5$ ($CV_{thr,0.5}$) and found that these approaches yielded comparable accuracies: accuracies by region varied between 66% to 81% and when averaging across the regions and by approach, the OAs were 74.2%, 75.2% and 75.4%, respectively. Thus, $CV_{thr,0.5}$ performed even slightly better than either of the more computationally expensive optimization approaches.

However, that study also has some important limitations. One major limitation was that it relied on temporally coarse (at best, every 46 days) L-band data. For each of the 217 tiles, only 6–14 images spread over a four growing seasons (2007–2010) were available. Thus, it is likely that the reported CV_{thr} values in that study (0.48 averaged over the 11 regions) have a relatively large uncertainty and should be re-calculated when temporally dense L-band data become available (e.g., NISAR). This potential concern is supported by results of other L-band studies using the same approach, but also using temporally dense time series obtained during the summer growing season, at few sites in the Southeast and Canada (Huang et al., 2021; Kraatz et al., 2021a, 2021b). Those studies for the most part reported smaller optimal CV_{thr} values in a 0.2 to 0.4 range. Thus, it will be important to re-calculate the optimal CV_{thr} when temporally dense, global L-band time series become available. Another important limitation was that the study did not describe in detail how the optimal CV_{thr} values were determined from the ROC or histogram approach. This presents an important issue because regional and global calibration and validation efforts require a systematic means of estimating optimal CV_{thr} values, so that the optimizations are consistent, the processes can be automated and reproduced. Also, that study did not report optimal CV_{thr} values for each of the 217 tiles, but the 11 regions. This may be insufficient to show how optimal CV_{thr} values vary spatially or how OAs for a constant CV_{thr} value compare the optimized results at each tile. Also, while it is plausible that the two optimization processes (histogram and ROC) yield comparable accuracy, it is somewhat surprising that $CV_{thr,0.5}$ appeared to perform slightly better than either of the optimizations. This may again have to do with only having a temporally coarse time series available, and the fact that it spanned four growing seasons.

Currently there are no L-band SAR data that meet the desired criteria of being temporally dense, freely available and also cover large regions for testing the NISAR's cropland algorithm. Alternatively, the cropland algorithm may also be evaluated using freely available Sentinel-1 data. Sentinel-1 has dense time series (every 6 to 12 days), global coverage and sufficient spatial resolution to produce ~1 ha products like those expected from NISAR. Thus, Sentinel-1 possibly is the best available alternative for testing some of NISAR's science algorithms. In anticipation of NISAR, and in light of abovementioned knowledge gaps, further testing and development of the temporal CV approach is needed, especially in light of: (1) a mission accuracy requirement of 80%, (2) prior work being limited in spatial extent and/or the temporal density, (3) prior work lacking automated or reproducible means of determining optimal CV_{thr} values and reporting these only over individual or few regions at a time.

Using all available ascending Sentinel-1 data (Sentinel 1A and 1B) collected in 2017, this work achieves a dense time series corresponding to an effective revisit time of between 6 and 12 days. This work only uses cross-polarized ('VH') data because prior work indicated relatively better OAs at L-band, and also to limit amount of data to be processed (Huang et al., 2021; McNairn et al., 2009b; Whelen and Siqueira, 2017). As in Whelen and Siqueira (2018b), this study also aims to provide an estimate of the maximum attainable OAs of the temporal CV approach

(OA_{YJS}), and compare results when using $CV_{thr,0.5}$ ($OA_{0.5}$) against it. We use an automated optimization process (YJS) to determine the best performing CV_{thr} value at each tile referred to as $CV_{thr,YJS}$ (Youden, 1950). Because the optimization process is also computationally expensive, and because it is also desirable to be able to prioritize some areas over others in calibrating and validating this approach, data are further subdivided into 100 one-by-one degree tiles. Each tile has a $150\text{ m} \times 150\text{ m}$ grid spacing and is located in an important agricultural region within CONUS (see Section 3.4). OAs are determined by comparing results to a reference dataset, the CDL. Because the CDL includes many different crop and non-crop categories, each category is assigned as crop or non-crop (or masked). These re-classifications are performed in a consistent manner with NASA's Sustainable Development Indicator Group's (SDIG) definition of Agricultural land¹ (see Section 3.1). Because of the robustness of the temporal CV approach, and good results obtained in prior work, we anticipate $OA_{0.5}$ values between 75% to 85%. We also expect there to be some geographic dependence in the optimal CV_{thr} values, due to some crops and non-crops and agricultural practices being more or less prevalent in different regions (e.g., the 'Corn Belt', irrigation). For the same reasons, we also expect that optimal CV_{thr} values at nearby tiles would generally tend to be similar to one another.

2. Data

2.1. Sentinel data and processing

This study used C-band imagery from the European Space Agency² Sentinel-1 satellites that is freely available from the Alaska Satellite Facility.³ Sentinel-1 currently consists of two satellites, Sentinel-1A and Sentinel-1B. Sentinel 1 satellites have a 180 degrees orbital phasing difference and provide nearly full coverage over land every 12 days when the two satellites work in unison, a six-day temporal resolution is achieved for certain regions (Torres et al., 2012). The platforms are sun-synchronous with an orbit height of 693 km and an incidence angle ranging between 20° and 45° at time of collection. At capture, Sentinel-1 imagery has a resolution of $20\text{ m} \times 22\text{ m}$, which is resampled for distribution to a $10\text{ m} \times 10\text{ m}$ resolution data product.

This study employed all VH polarized 2017 (January through December) ascending Level-1 GRD SAR imagery from both Sentinel-1A and Sentinel-1B satellites. Ascending data were used because it covers CONUS in its entirety, whereas descending data only approximately covers the westernmost third (<https://sentinel.esa.int/web/sentinel/missons/sentinel-1/observation-scenario>). Data were processed using Gamma remote sensing software running on Amazon Web Services (Werner et al., 2000). Sentinel-1 data were radiometrically and terrain corrected using the approach described in Small (2011), speckle filtered using a multi-temporal filter (Quegan and Yu, 2001), projected to the γ^0 scattering plane and multi-looked to a $150\text{ m} \times 150\text{ m}$ resolution (Werner et al., 2000). At this spatial resolution, each approximately one-by-one degree tile consists of 720×720 grid points. CV and crop/non-crop classifications were then made according to Section 3.2.

2.2. Cropland data layer (CDL)

The CDL was used as a baseline comparison. The CDL is generated annually, is distributed by the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) and encompasses 106 different crops (Boryan et al., 2011). The CDL is available at a 30 m spatial resolution and the 2017 layer was developed using imagery from

the Landsat 8 OLI/TIRS, the Disaster Monitoring Constellation DEIMOS-1 and UK2, ISRO ResourceSat-2 LISS-3 and Sentinel-2 satellites.⁴ The strength and emphasis of the CDL are crop-specific land cover categories⁴, but it also incorporates non-crop land covers from sources such as the United States Geographical Survey's National Land Cover Database (NLCD). Nationally, the CDL attains state-level classifications with OAs ranging from 68.0% to 92.5%, with an average OA of 81.5%⁴. USDA NASS uses a decision tree classifier constructed through the Rulequest See5.0 software⁴. The approach necessitates that data are split into two categories: training data for creating the decision tree, and validation data for calculating OA of the CDL. Then the MRLC NLCD Mapping Tool is used to apply the decision-tree to the validation data using via ERDAS Imagine⁴.

3. Methods

3.1. Definition of cropland and masking approach

To make comparisons with the binary crop/non-crop results of the temporal CV method, it was necessary to reclassify the CDL layers into a crop/non-crop dataset. To add a degree of robustness and remove effects of edge pixels between different land-cover types within the CDL, pixels that were not surrounded by their same land cover type and smaller than two hectares in size were masked out to remove errors inherently present in the CDL, similar to Whelen and Siqueira (2018a). Portions of the CDL that were classified as "Open Water" were also masked. Water-bodies have low signal-to-noise ratio because of low backscatter cross section (specular scattering), which results in large CV values (false classification as crop).

For determining which CDL classes will be included as cropland, non-cropland or masked we employ the same land use and land cover classification system as Anderson et al. (1976) and SDIG. In this land cover classification system agricultural lands are categorized as (1) confined feeding operations, (2) cropland and pasture, (3) orchards and vineyards and (4) other. Since Anderson et al. (1976) and the SDIG definitions, many studies have shown that it is difficult to distinguish between grassland, pasture and other land covers using SAR only (Fisette et al., 2014; Lark et al., 2017; Lindsay et al., 2019; McNairn et al., 2009a; Waske and Braun, 2009). The CDL is a more recent remote-sensing-based dataset, and also reflects this difficulty in grouping grassland/pasture into the same categorization code. Thus, viewed in terms of SDIG and classification capabilities through remote sensing, there appears to be a gray area with regards to whether this CDL class might be better categorized as cropland or non-crop. This distinction can have substantial impact on reported performance metrics because this CDL class is highly prevalent throughout CONUS and our tiles (Section 4.2.3). Therefore, this study will examine both cases.

The temporal CV approach is expected to perform reasonably well overall (e.g., 70–90% accuracy range) and for many land covers (Kraatz et al., 2021b). But over other land covers such as those consisting of trees (orchards and some perennials) or bare ground (fallow) throughout the year, SAR backscatter values are expected to have relatively smaller CV values. This is a reasonable expectation and fundamentally along the same line of thinking provided in Goetz et al. (2006). That study focused on mapping forest disturbance and recovery in space and time, using the same temporal CV approach, but with Normalized Difference Vegetation Index values instead of radar cross section. The study selected areas having low CV values "... to represent unburned areas in order to identify areas that were least likely to be influenced by land use, sub-pixel fires and other types of disturbance (e.g., insects), or interannual influences of climate" (Goetz et al., 2006). Because here, it is necessary for CV values to exceed a threshold to be classified as crop, it is should be

¹ Sustainable Development Indicator Group (SDIG): https://www.hq.nasa.gov/iwgsdi/Agricultural_Land.html.

² Copernicus Sentinel Data 2017 processed by ESA.

³ Alaska Satellite Facility Data Search available at <https://search.asf.alaska.edu/>.

⁴ USDA NASS CDL FAQ: https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php#Section3_2.0.

expected to be mutually exclusive for those trees corresponding to crops or orchards to be correctly classified as crops, and tree crops were masked.

Summarizing, this work's definition of cropland is consistent with the SDIG: (1) we consider the same land cover classes as crop – including fallow/idle fields; (2) the SDIG also distinguishes between pasture and cropland (in nomenclature), and we furthermore compute results when it is considered a crop ([Section 4.3.2](#)) and non-crop; (3) the SDIG has orchards in a separate category: between either masking or classifying them as a non-crop, we opted to mask them; (4) the SDIG doesn't mention anything specific regarding perennials and we only mask some of them. We masked the categorization codes listed under the block "CROPS 66–80" in the CDL, corresponding to orchard type crops: cherries, peaches, apples, grapes, Christmas trees, other tree crops, citrus, pecans, almonds, walnuts, pears. We masked some 'perennial' crops corresponding to the block "CROPS 195–255", but only if they corresponded to crops grown on trees: pistachios, prunes, olives, oranges, avocados, pomegranates, nectarines, plums, and apricots. Overall, out of the total of 51,840,000 possible pixels (the 100 tiles each consisting of 720 by 720 pixels), 48.7% were considered valid for making classifications. The large proportion of pixels being masked can mainly be attributed to the two hectare mask implemented as described above.

3.2. Crop/non-crop classification

Crop/non-crop classification are based on the temporal CV defined as

$$CV_{pq}(t) = \frac{\sigma_{pq}}{\mu_{pq}} \quad (1)$$

where σ is the standard deviation and μ is the mean of the backscatter in polarization pq (here, VH) and linear power units over time. CV is a unitless quantitative statistical measurement of variation. The CV was calculated on a pixel-by-pixel basis for the sigma naught derived backscatter values in the time series images. The calculated CV raster image captures the amount of variation that exists through the time series in each individual pixel, with relatively higher and lower CV values expected to correspond to crops and non-crop, respectively ([Whelen and Siqueira, 2017](#)). To generate a crop/non-crop classification from the calculated CV statistic data layer it is necessary to determine a CV_{thr} where any pixel value greater than the CV_{thr} is classified as crop and any pixel value less than the CV_{thr} is classified as non-crop:

$$CV_{pixel} \left\{ \begin{array}{l} < CV_{thr}, \text{non-crop} \\ \geq CV_{thr}, \text{crop} \end{array} \right. \quad (2)$$

3.3. Overall accuracy

The confusion matrix ([Table 1](#)) and Eq. ([3](#)) are used to calculate the OA for each tile. [Section 4.2.1](#) reports OA values that are aggregated by regional subset, calculated from the arithmetic mean and standard deviations of all tiles in the subset. [Section 4.2.3](#) reports OA values by crop and regional subset but calculated from the sum totals of all pixels. Results are calculated using all valid pixels in each of the 100 tiles, with the CDL as reference.

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

Table 1
Confusion matrix.

Model (CV_{thr} value)	Observed (the CDL)	
	Crop	Non-Crop
Crop	TP	FP
Non-Crop	FN	TN

where TP, FN and TN respectively are the True Positive, False Negative and True Negative counts.

It is important to note that OAs were calculated the same way, irrespective of whether $CV_{thr,0.5}$ or $CV_{thr,YJS}$ ([Section 3.4](#)) is used. $CV_{thr,YJS}$ can be considered a kind of model training and validation, but here – in order to obtain an estimate of the performance ceiling (optimization), the same data were used for training and validation, on purpose. Thus, $CV_{thr,YJS}$ and associated performance metrics (i.e., OA_{YJS}) should not be considered the model, nor should they be considered representative of normally attainable performance by the temporal CV approach. OA_{YJS} should rather be considered an estimate of the performance ceiling of the temporal CV approach, at each tile as well as overall, because $CV_{thr,YJS}$ is optimally estimated at each of the 100 tiles ([Section 3.5](#)). But it is important to note that $CV_{thr,0.5}$ had not been informed on by the optimization procedure – this value is used as a starting point because [Whelen and Siqueira \(2018b\)](#) indicated that it performed best, albeit using temporally coarse L-band data. Thus, for $CV_{thr,0.5}$ and associated performance metrics (i.e., $OA_{0.5}$) there is no model calibration step, and no overlap between calibration and validation data. Then, the capability of the computationally inexpensive approach ($OA_{0.5}$) is assessed by comparing results to OA_{YJS} : if $OA_{0.5}$ falls reasonably close to OA_{YJS} and also exceeds the desired 80% OA level, it is an acceptable model. This also allows for an accurate assessment regarding the cost-benefit in further refining the model, i.e. making it more accurate at the expense of making it more complicated and computationally expensive.

3.4. Optimizing CV_{thr} values to estimate the performance ceiling

The base model uses a $CV_{thr} = 0.5$ in Eq. ([2](#)) for every pixel in each tile, because prior work showed that doing so was as accurate as when using CV_{thr} values that were optimized at each tile ([Whelen and Siqueira, 2018b](#)). Knowledge of the performance ceiling, optimal CV_{thr} values and their spatial distribution is useful for model calibration and validation (see [Section 3.5](#) and [Whelen and Siqueira, 2018b](#)). This study estimates the performance ceiling using Youden's J statistic ([Youden, 1950](#)), which is suitable for a binary classification (more in [Section 4.3.3](#)). YJS optimization is an iterative approach requiring computing the True Positive Rate (TPR) and False Positive Rate (FPR) for a range of CV_{thr} values. TPR (or sensitivity) and FPR (or 1 – specificity) are calculated as

$$TPR = \frac{TP}{TP + FN} \quad (4)$$

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

The ROC curve is the result of plotting of TPR versus FPR values given a range of CV_{thr} , for example between 0.00 and 1.00 in 0.01 intervals ([Fig. 1](#)). The confusion matrix (including TPR and FPR) are calculated using all of the valid pixels at each of the 100 tiles with the CDL as reference. Each point on the ROC curve is based on results of the confusion matrix (i.e., the True Positive Rate versus False Positive Rate) for a given CV_{thr} value. $CV_{thr,YJS}$ then is equal to the CV_{thr} value that corresponds to the point on the curve yielding the largest vertical distance (Eq. ([6](#))) between the ROC curve and the line of no discrimination ([Habibzadeh et al., 2016](#)). Instead of reporting all of the results (the 101 results for TPR, FPR, OA and YJS) at each of the 100 tiles, this work only reports the optimal value of OA and CV_{thr} at each tile (i.e., OA_{YJS} and $CV_{thr,YJS}$). Also, because OA_{YJS} is the OA value that corresponds to the CV_{thr} value that yielded maximum YJS value, OA_{YJS} does not necessarily correspond to the highest obtainable accuracy. Since using all available data yields better estimates of the performance ceiling, OA_{YJS} and $CV_{thr,YJS}$ are calculated using all valid pixels rather than a subset (e.g., training, validation).

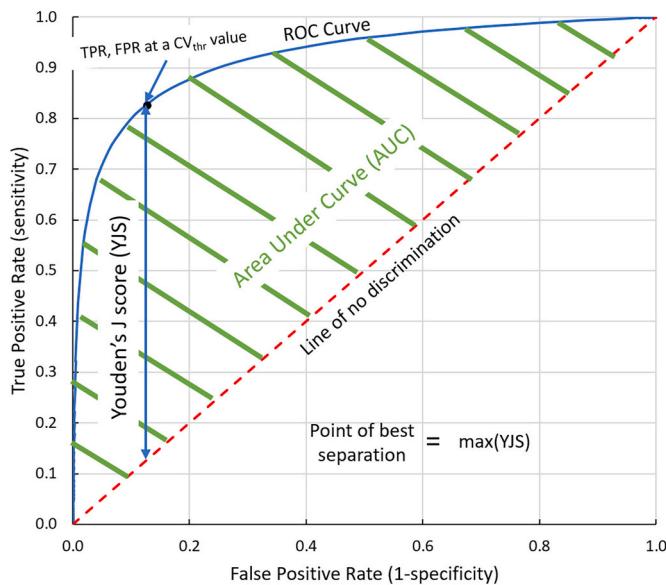


Fig. 1. A ROC curve is constructed from the results obtained when using different crop/non-crop delineating thresholds, i.e., the computed for each threshold True Positive Rate versus False Positive Rate. The optimal CV_{thr} value is that corresponding to the point on the curve yielding the largest vertical distance between the ROC curve and the line of no discrimination.

$$YJS = TPR - FPR \quad (6)$$

We also evaluate the performance of the YJS approach for crop and non-crop classifications in view of a secondary metric common to ROC-based approaches, the area under the ROC curve (AUC). AUC can be viewed as a measure of expected classification performance over a range of cost parameters and data points (Flach et al., 2011). Thus, it can be a useful quantity for showing how the expected classification performance of the CV approach varies over different tiles. The AUC ranges between 0 and 1 and characterizes the discrimination power between the two separate classes. The higher an AUC value the better the model is at separating the two classes. An AUC value of 0.5 signifies there is no distinguishability between the classes (Fan et al., 2006).

3.5. Use of ROIs to assist in model calibration and validation

CV values themselves are more or less straightforward to compute (barring all the details regarding applying the appropriate SAR processing steps corrections on a large dataset). The difficulty lies in how to convert those values to yield an accurate cropland product at scale ($\geq 80\%$). The easiest case would be to use a single CV_{thr} value for delineating crop and non-crop everywhere. But unless an estimate for maximum OA is available, one would not know well the CV_{thr} value performed or how much room for improvement remains. This is not only important information when aiming to meet an OA requirement (e.g., NISAR's 80% for cropland), but also to better weigh tradeoffs between potential accuracy gain versus resources invested.

Estimates of a ceiling for OA using given approach and dataset requires some manner of optimization which can be computationally expensive. Computational cost can be reduced via approaches that limit the number of pixels that are evaluated, by limiting analyses to specific regions of interests (ROI) (e.g., as in Whelen and Siqueira (2018b)). ROIs are an attractive option, because they are intuitive and they allow for CV_{thr} value(s) to be selected such that good performance in the ROIs is prioritized. The OA ceiling is then estimated from the OA_{YJS} values at each ROI. This approach also allows identification of how $CV_{thr,YJS}$ values are distributed in space in order to: (1) estimate the upper bounds of accuracy and show how closely the computationally inexpensive approach compares to it; (2) inform on alternative optimization

methods, such as the use of multiple, regional thresholds. Furthermore, there will not be as strong a need in pursuing (2), if a single CV_{thr} value (i.e., 0.5) is able to result in OAs $\geq 80\%$.

Therefore, we limited our analyses to the 100 one-by-one degree tiles shown in Fig. 2a. As described above, the $CV_{thr,YJS}$ values are computed at each tile, yielding data on the geographic distribution and spread of $CV_{thr,YJS}$ and OA_{YJS} . As we are interested in evaluating this approach for correctly identifying cropland over a wide range of environments where accurate data are available (i.e., the CDL), analyses focused only on known cropland regions throughout CONUS. Locations of the tiles were selected based on a combination of the major Farm Resource Regions identified by the USDA (Heimlich, 2000), the locations of the 18 Long-Term Agroecosystem Research sites (Spiegal et al., 2018) and locations of dense agricultural land identifiable in the CDL. The USDA Farm Resource Regions are derived from four data sources: USDA Farm Production Regions, NASS Crop Reporting Districts, USDA Land Resource Regions and an analysis of U.S. farm characteristics at the county-level (Heimlich, 2000). These tile placements result in 25 different states being sampled, with data in each of the four statistical census geographic regions (Bureau, 2010).

Tile placements resulted in two, 55, 22 and 21 tiles being placed in the Northeast, Midwest, South and West (Fig. 2a). The Northeast is omitted from the regional analyses, due to there being only two tiles. Figs. 2a and b also give information on the number of dates used for computing the CV. Because Sentinel overlaps some parts of a tile more often than others, we estimated the number of different dates used according to those occurrences where most of a tile was covered by an image. Fig. 2b clearly indicates that there are two peaks, located at about 30 and 50 dates used (average: 41). This corresponds to respective revisit times of about 12 and 7 days, and is consistent with using all available ascending data from Sentinel 1A and 1B. Higher values may be found at specific locations because ESA's acquisition scenarios⁵.

Fig. 2c shows the histogram of the crop prevalence (bin size 0.05) for the 100 tiles studied. It is important to account for crop prevalence, because (1) areas entirely consisting of crop or having no crop at all yield are trivial cases for determining optimal CV_{thr} values and (2) the YJS optimization performance has some dependence on the prevalence of the item to be optimized (here, the crop percentage of a tile) (Smits, 2010). Fig. 2c shows that the selected tiles cover a full range of possible crop prevalence, and that a little less than half the tiles (41) had crop prevalence ≥ 0.5 .

3.6. Sub-setting of CDL classes for conducting analyses by crop

Prior work, but only over one agricultural area, indicated that the temporal CV approach is poorly suited for classifying specific land cover types (Kraatz et al., 2021b). It is valuable to include analyses by crop, because the study by Kraatz et al. (2021b) was limited in spatial extent, and other studies did not stratify by land cover. This study stratifies CV values by CDL class, but only focusses on the most prevalent ones. OA values are also stratified, but only for OA_{YJS} , to more clearly examine the approach's capabilities and limitations, even when the optimized results are used. First, at each of the 100 tiles, statistics are first tallied over all the CDL classes accounting for over 2% of the valid pixels (those that had not been masked). Because this step is performed on each tile, it is expected that many of the remaining CDL classes still may only have a marginal contribution to OA values at the regional or national scale. Also, OA values for CDL classes having very low prevalence at these scales may not be as reliable compared to others, because their values were calculated from comparatively smaller data points. Second, to ensure a focus only on those CDL classes that have relatively greater impact to OA, a secondary screening was applied requiring CDL class to

⁵ Sentinel-1 Acquisition Scenarios: <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/observation-scenario>.

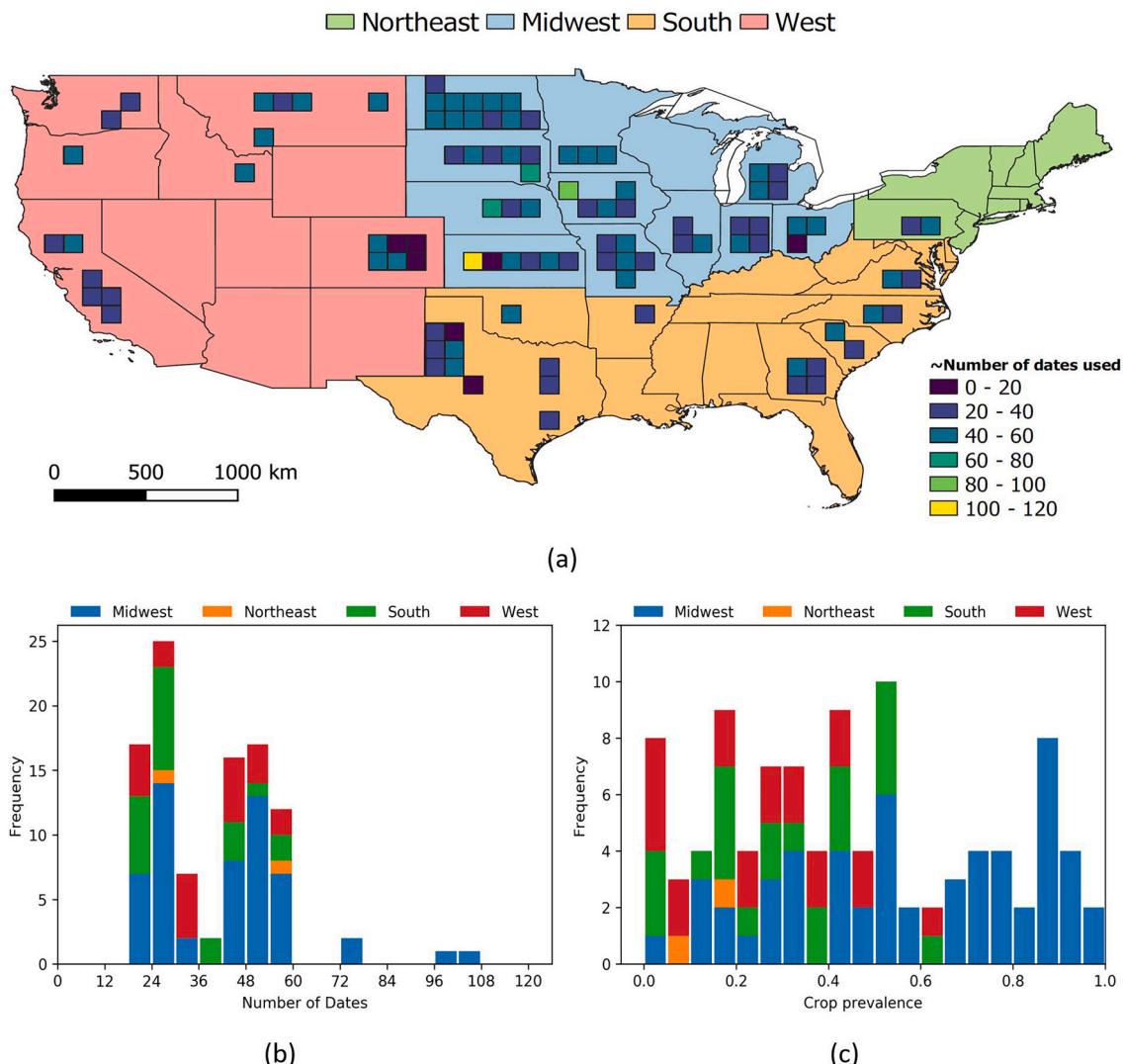


Fig. 2. (a) Map showing the locations of the 100 tiles used across the United States and the approximate number of Sentinel 1 observations used for computing CV values (counting those observation covering >50% of a tile). The background shows the census geographical regions. (b) The histogram corresponding to (a), colored by census geographical region. (c) A histogram of crop prevalence for the tiles shown in (a), colored by census geographical region. Crop prevalence was determined using the CDL.

have a prevalence $\geq 0.5\%$ at the national scale.

4. Results and discussion

4.1. Comparison of CV values with CDL over CONUS

Fig. 3 shows that the temporal CV approach generally works as intended for identifying croplands when using Sentinel-1C-band data: there is reasonably good agreement between cropland determined via the CDL (Fig. 3a) and the locations where CV values are elevated (Fig. 3b). Furthermore, all the major cropland regions also tend to have the relatively greatest CV values: California (Central Valley), Eastern Washington and the Corn Belt in the Central United States. Visual comparisons of the datasets yield rough estimates of CV values greater than 0.5 and less than 0.3 appearing to be good indicators of cropland and non-crop areas, respectively. Thus, there is a relatively large range of CV (from about 0.3 to 0.5) for which the CDL indicates that both crops and non-crops are present. Furthermore, a large proportion of pixels correspond to this range of CV values. This hints at the potential challenges in applying a single CV_{thr} values for the binary classification over large regions such as CONUS. For example: (1) large areas in western

states (e.g., Nevada) have CV values near 0.5, but the CDL indicating few if any crops; (2) CV values in the Northeast fall into a narrow 0.2 and 0.3 range, making it difficult to ascertain what the cropland extents are, but they can be readily identified when using the CDL.

4.2. Accuracy metrics, CV_{thr} values and crop prevalence

Applying the temporal CV approach for crop/non-crop classification to Sentinel 1C-band data resulted in $OA_{0.5}$ and OA_{YJS} values of 81.5% and 86.8%, respectively (Table 2). OA improved the most from the YJS optimization in the Midwest (10%), to a lesser degree in the South (0.9%) and decreased in the West (-2.5%). Crop prevalence varied substantially over individual tiles and regions, ranging from 30% in the West and South to nearly 60% in the Midwest. The Midwest also has about 50% greater variation in crop prevalence than other regions (Table 2). The optimizations yielded much greater improvements in the Midwest, indicating that YJS optimization may be biased towards performing better at tiles having greater crop prevalence (see Section 4.3.3).

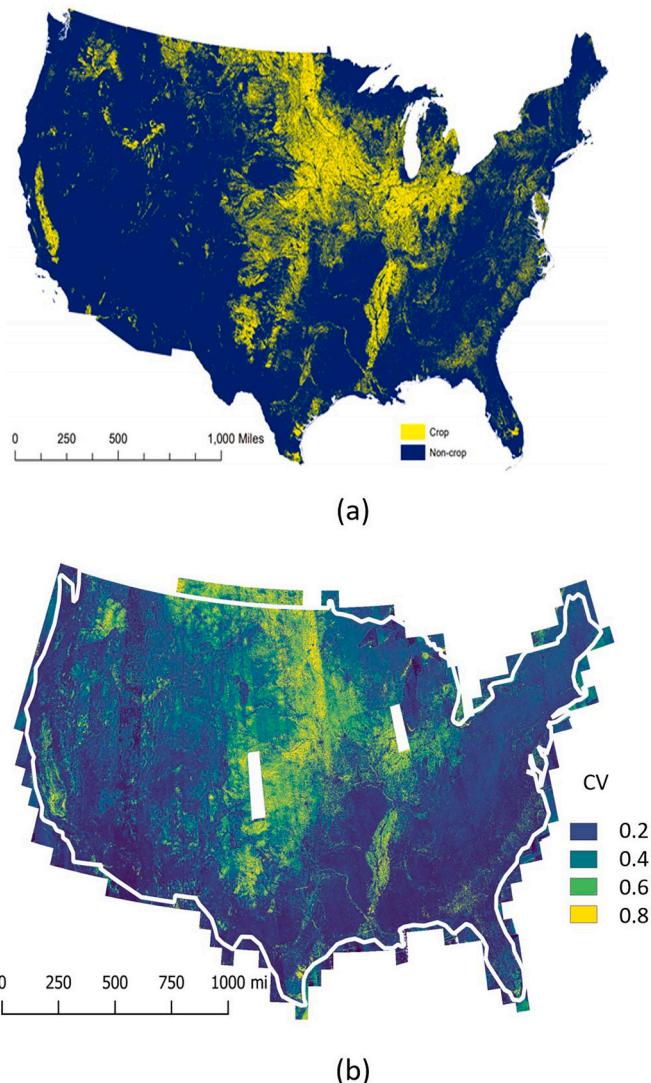


Fig. 3. (a) Crop and non-crop classification as per the CDL and (b) CV values computed from VH polarized, ascending Sentinel-1 SAR data. White colors indicate regions that are masked or where no data were collected.

Table 2

Mean and standard deviation of OA_{0.5}, OA_{YJS}, CV_{thr_YJS} and crop prevalence, by census geographical region. The regional statistics for the Northeast are omitted, as there were only two tiles.

Region	OA _{0.5}	OA _{YJS}	CV _{thr_YJS}	Prevalence
Midwest	80.2 ± 13.4	90.2 ± 6.4	0.42 ± 0.12	58.7 ± 27.0
South	89.1 ± 7.4	90.0 ± 6.8	0.34 ± 0.12	30.3 ± 18.2
West	76.1 ± 13.7	73.6 ± 14.4	0.40 ± 0.11	25.0 ± 17.5
Overall	81.5 ± 13.0	86.8 ± 11.1	0.40 ± 0.12	44.4 ± 28.1

4.2.1. Overall accuracy (OA)

Performance may vary greatly by tile (Fig. 4a). Overall, OA_{YJS} values ranged from 32.3% (a tile in California) through 97.8% (a tile in Illinois), and OA_{0.5} values ranged from 35.1% (a tile in Ohio) through 99.7% (a tile in Missouri).

Using CV_{thr_YJS} values instead of CV_{thr_0.5} led to improved, similar and worse OAs at 56, 35 and 9 tiles, respectively (Fig. 4b). Unlike the study by Whelen and Siqueira (2018b), we note a clear performance decrease when CV_{thr_0.5} is used instead of an optimized approach. Improvements predominantly occurred in the Midwest (27 tiles). Optimized thresholds also led to more consistent crop/non-crop

classifications in the Midwest: the standard deviation of OA halved (Table 2). OA_{YJS} performed poorly in the West, with 6 tiles having lower OAs and only two improving compared to OA_{0.5}. OA_{YJS} also performed better (5 tiles) than OA_{0.5} in the South, compared to two with worse results. Fig. 5 shows how performance shifted under the two classification approaches, indicating that results for OA_{YJS} are better, overall.

Compared to Whelen and Siqueira (2018b), this study obtained much better results for both the optimized (86.8% vs. 75.2%) and fixed threshold (81.5% vs. 75.4%). Recent work using the same approach and dense (about every two weeks) L-band UAVSAR data showed that close to 80% accuracy could be obtained using time series as short as about 7 images when (Huang et al., 2021; Kraatz et al., 2021a). While more research is needed to identify the length of the time series needed to reliably obtain OAs $\geq 80\%$, results indicate that the length and density of the time series appears to be a limiting factor, and that this impacted the evaluations of Whelen and Siqueira (2018b). We also note that results of other recent studies can help explain how good results could be obtained even when only using a single constant threshold applied to all tiles. For example, studies by Kraatz et al. (2021a, 2021b, 2020) show the ROC curves, OA and YJS values as function of CV_{thr}, but only over a few ROIs. But consistent across those results was that when YJS scores are relatively greater, so are OAs (also see Section 4.3.3) and also that near-optimal YJS and OA values occurred over a fairly wide range of CV_{thr} values. For example, if CV_{thr_YJS} is 0.3, CV_{thr} values of 0.2 and 0.4 usually also yielded close to optimal OAs (within $\sim 5\%$). Besides, Whelen and Siqueira (2018b) had already shown that a single CV_{thr} value can perform reasonably well over CONUS.

4.2.2. Optimal CV_{thr} values

Fig. 6a shows that tiles located near the coast tend to have considerably smaller CV_{thr_YJS} values (about 0.3) than the Central United States (about 0.5). CV_{thr_YJS} values have a wide spread, between 0.19 and 0.65. A histogram of CV_{thr_YJS} values shows two peaks at about 0.3 and 0.5 (Fig. 6b). Thus, using a combination of CV_{thr_0.5} plus CV_{thr_0.3} may further improve results i.e., CV_{thr_0.5} for between about 110° West to 95° longitude, and CV_{thr_0.3} elsewhere.

Fig. 6a also shows a clustering of similar CV_{thr} values, consistent with nearby regions sharing similar characteristics (e.g., climates, crops and management practices). Differences in the crop type, prevalence, climate (e.g. continental vs. coastal) and management practices could explain the relatively greater CV_{thr} values found in the Central United States compared to near the coasts. Results obtained by Whelen and Siqueira (2018b) also show a similar spatial stratification of CV_{thr} values: although values only ranged between about 0.4 to 0.6, all the values in the 0.4 to 0.5 range corresponded to areas closer to the coasts and all CV_{thr} values between 0.5 and 0.6.

4.2.3. Classification performance by CDL class

After the first data reduction step, consisting of only keeping those CDL classes to having $\geq 2\%$ prevalence in each of the 100 tiles, 24 crop and 9 non-crop classes remain. Altogether they amount to 96.1% of all valid pixels, indicating that OA metrics computed for these CDL classes accurately represent a national scale analysis. However, most of these CDL classes only have marginal prevalence ($\leq 0.5\%$) across all 100 tiles, leaving only seven crop and five non-crop classes, altogether amounting to 93% of all valid pixels (Fig. 7a). Fig. 7b shows the CDL occurrence by tile for those classes exceeding the 2% prevalence threshold in a tile, for the same CDL classes as shown in Fig. 7a.

Results shown in Figs. 7a and b are mostly consistent with one another, indicating that our study mainly encompassed corn (13.9%), soybeans (14.7%) and grassland/pasture (29.9%). Other crop and non-crop classes had prevalences below 5% and 10%, respectively. The 93% cumulative sum is composed of 40.3% crop and 52.7% non-crop CDL classes, amounting to a relative crop and non-crop breakdown of 43.4% to 56.6%, which is very close to the overall mean value of 44.4% to 55.6% shown in Table 2. For reference, USDA NASS 2012 estimates of

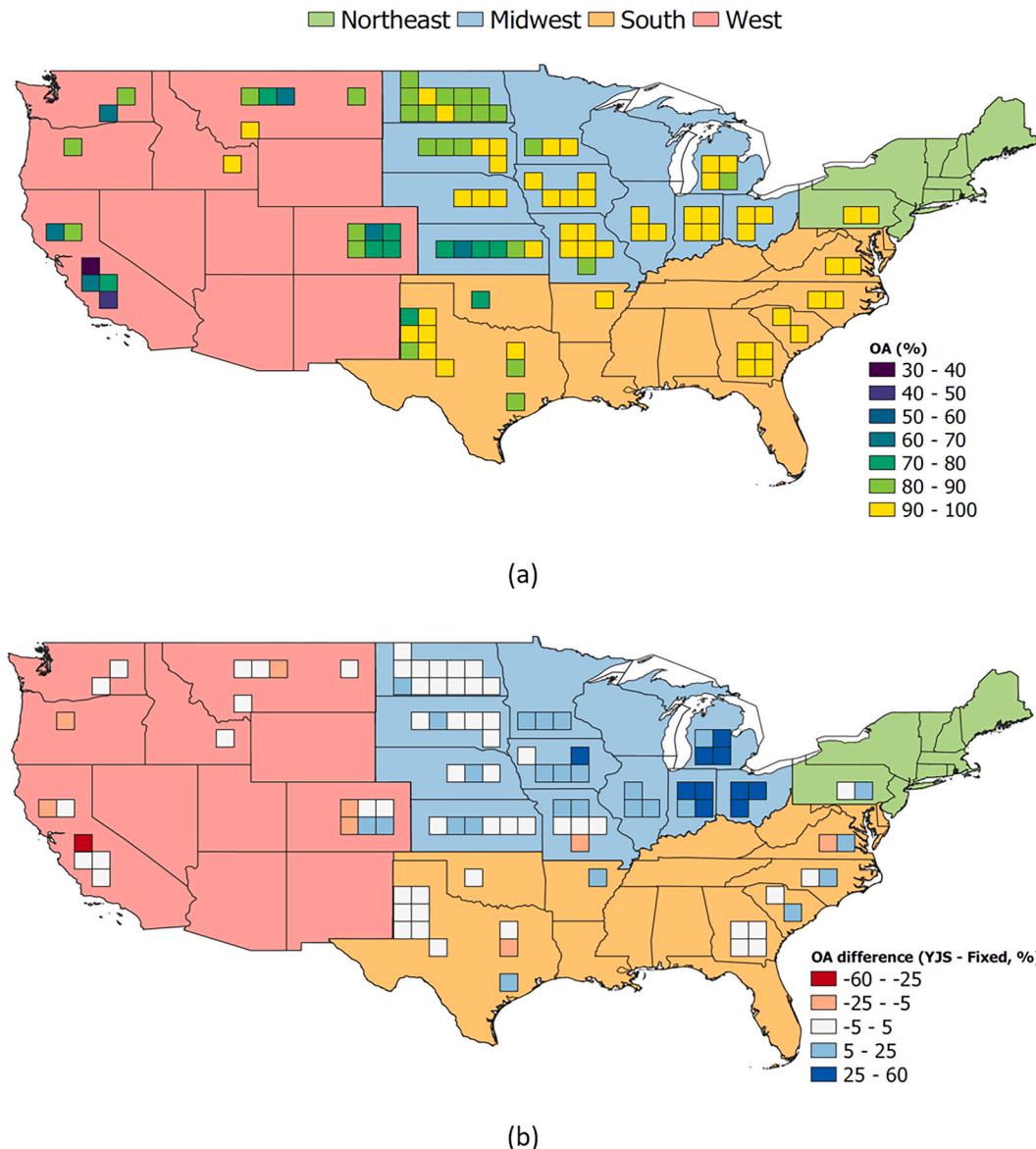


Fig. 4. Spatial distribution of (a) OA_{YJS} and (b) the OA_{YJS} – OA_{0.5}.

major land usage classes over CONUS, as percentage of total land area, are: cropland 18%, fallow at 2.1%, cropland used for pasture plus grassland pasture and range at 35.3% (Bigelow and Borchers, 2017). It also lists total agricultural land (cropland, grazing land and special use land) at 62.6%. Whereas our sampling approach (see section 3.4) more predominantly focusses on cropland used for crops (44.4%, Table 2) and relatively less on grazing land (grassland/pasture) with 29.9%. In comparison, the proportion of land considered agricultural land in this work is about 10 percentage points greater than that what is found in CONUS (74.3% vs. 62.6%).

We also evaluated CV values by CDL class and region by averaging the median, first (Q1) and third (Q3) quartile CV values for the same crop over all the tiles in the region (Fig. 8). The difference between Q3 and Q1 (Q3–Q1) is referred to as the ‘box’ or the interquartile range (IQR). Fig. 8 shows that there is considerable overlap among the crop and non-crop classes in all regions, which indicates that it would be difficult to apply the temporal CV approach to reliably classify individual CDL classes. However, there is a reasonably good separation between the crop and non-crop classes in that non-crop classes tend to clearly have relatively lower CV values, by about 0.3. These two points are consistent with results of a study that applied the CV approach to

UAVSAR L-band and Sentinel C-band data over an agricultural site in Canada (Kraatz et al., 2021b). CV values of each crop are fairly consistent across the different regions (median CV values within 0.1). Notable exceptions are fallow/idle cropland and soybeans which have CV values that are smaller by about 0.2 in the South. And grassland/pasture has relatively greater CV (by about 0.15) and wider IQR in the West. This will likely produce relatively higher rates of false classifications of grassland/pasture as a crop in the West compared to the other regions (see Section 4.3.2).

Fig. 9 shows the accuracy breakdown by crop and region, corresponding to the performance ceiling OA_{YJS}. Non-crop classes, except for grassland/pasture in the West, generally perform well ($\geq 80\%$). The relatively poorer accuracy in grassland/pasture in the West can be attributed to the relatively greater CV values and IQR (Fig. 8). Because, grassland/pasture has high prevalence (29.9%) and is a ubiquitous land cover (79 of the 98 tiles of the three regions, Fig. 7b), OA strongly depends on how accurately grassland/pasture is classified. The relatively lower OA values found in the West (Table 2) can be attributed to the poorer results obtained over grasslands/pasture there (see Section 4.3.2). Many of the land cover classes have fairly consistent OA’s across the three regions (Fig. 9). CV values can directly indicate which classes

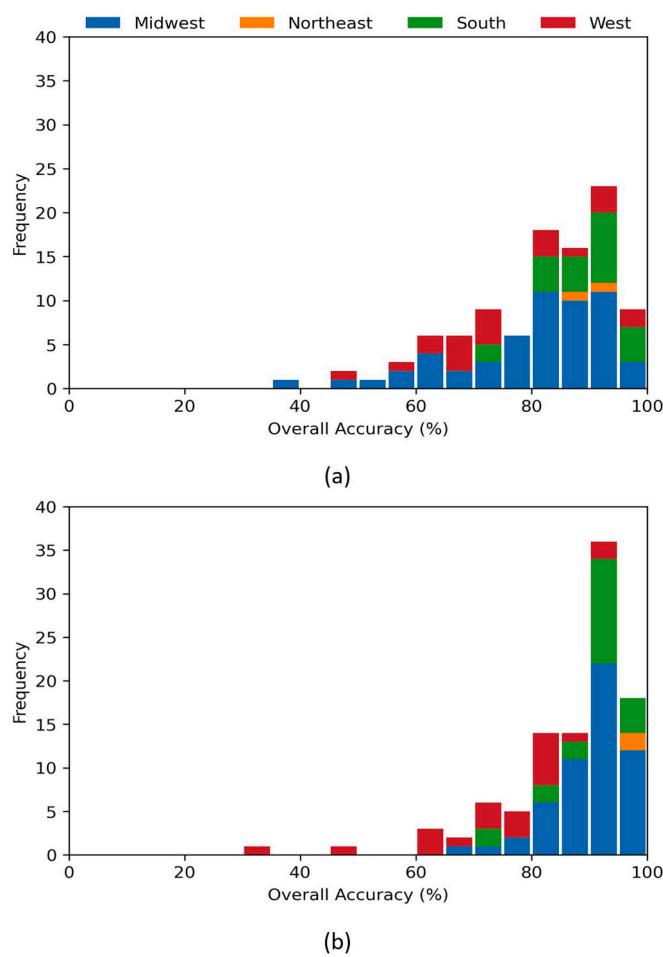


Fig. 5. OA histograms for (a) $\text{CV}_{\text{thr},0.5}$ and (b) $\text{CV}_{\text{thr},YJS}$ values using a bin size of 5%.

would classify accurately or poorly (Fig. 8). For example, corn, soybeans, cotton have the highest OAs, and this can be attributed to these crops also having highest CV values (Fig. 8). On the other hand, alfalfa, fallow/idle cropland, spring and winter wheat have relatively smaller CV values that resembling those of non-crop classes (Fig. 8), explaining their relatively lower OA values. And winter wheat classifies about 25% better in the West than other regions, owing to it having a median CV value that is 0.1 greater than other regions (Fig. 8).

Low OA_{YJS} values for alfalfa can be explained by alfalfa being grown differently than the other classified crops: alfalfa is grown and harvested several times per year, meaning that the growth stages portrayed by alfalfa will be different from the other crops that are harvested once a season (Fick and Mueller, 1989). The CV calculated for alfalfa in a time series analysis is more similar to bare soil than cropland due to frequent harvesting (Ferrazzoli et al., 1997; Moran et al., 2002). SAR approaches therefore have difficulty in distinguishing alfalfa from other crops, usually only obtaining relatively low classification accuracy (Arias et al., 2020; Hong et al., 2014; McNairn and Brisco, 2004). In various land cover/land use data products, alfalfa is not included in classification as a crop and is commonly misclassified in the CDL. The NLCD includes alfalfa as a cultivated crop in classification; however, alfalfa better fits the definition of a pasture/hay land cover defined by the NLCD. CDL products from 2008 to 2011 were only able to achieve 50% accuracy rate for alfalfa (Johnson, 2013), so errors may still be present in 2017 CDL alfalfa classifications. This inconsistency over alfalfa's classification and misclassification errors likely contribute the low OA when classified as a crop using the CV statistic for classification. Ultimately, it is important to note that while alfalfa was present in a fair number of tiles

(Fig. 7b, $N = 18$), it did not have high prevalence at national scale (0.7%) and poor classifications of this crop would not greatly impact the overall results.

4.3. Limitations

4.3.1. Evaluating performance by crop (or CDL class)

Results of Section 4.2.3 support the idea that the temporal CV approach is not suitable for identifying individual CDL classes. A fundamental limitation of the temporal CV approach is that even those areas that are clearly not agricultural in nature could readily be classified as crop, as long the surface routinely changes in a detectable way at C-band. Conversely, this approach may also underestimate agricultural areas due to CV values of some crop classes being similar to dominant non-crop classes. One potential reason is related to the composition of CDL classes within a ROI (here, a tile) (Kraatz et al., 2021b). For example, alfalfa has a median CV value close to that of the most prevalent non-crop class (grasslands) and consistently classifies poorly (Fig. 7). Also, if the biomass of the crop is relatively low compared to the wavelength used, such as when computing CV based on L-band over soybeans (Kraatz et al., 2021b), crop areas may be underestimated. Likewise, if an agricultural field is fallow, it may also be relatively more difficult to correctly detect it as cropland as shown by this class having one of the lowest accuracy values (Fig. 9). Except for the South, the median CV values for fallow/idle cropland were comparable to other crops (Fig. 8). However, OA_{YJS} values were poor, owing to this CDL class having some of the greatest IQR values and whiskers, irrespective of region (Fig. 8). For these reasons, the CV approach is clearly too simple for distinguishing between different crop and non-crop land covers and should not be expected to be capable of making land cover classifications beyond crop and non-crop.

4.3.2. Errors due to considering grassland/pasture as crop or non-crop

Grassland/pasture because it is a dominant non-crop CDL class in the census geographic regions (Fig. 7) and its classification may have a large impact on OAs. The CV approach could not accurately classify this land cover as non-crop, especially in the West (Fig. 9). Also, OA_{YJS} values in the West were found to perform particularly poorly - especially particular at several tiles in California (Fig. 4). A closer look at one California tile clearly shows that the overall classification performance is strongly impacted by whether grassland/pasture is considered a non-crop (Fig. 10a) or crop (Fig. 10b). When grassland/pasture was considered a crop, OA_{YJS} improved from 70.4% to 95.3% (Figs. 10 and 11).

Owing to the very large improvement noted the tile examined in Fig. 10, we further explored the impact of considering the CDL class "grassland/pasture" as a crop and repeated the OA_{YJS} calculations over all tiles with this assumption (Fig. 11). Results show that considering the grassland/pasture CDL class as either a non-crop or crop produces differences exceeding 5% at 47 of the 100 tiles: OA_{YJS} improved (decreased) at 18 (29) tiles. OA_{YJS} increased from 60.3% to 85.5% in California but decreased from 86.8% to 80.6% when considering all the tiles. Notably, changes in classification performance were highly localized: most improvements (11 out of 18 tiles) were limited to California and Kansas; and most of the decreases (16 out of 29) occurred in the Dakotas.

There are several potential explanations as to why considering grassland/pasture as crop or non-crop may result in the large, highly localized OA differences as shown in Fig. 11. One explanation is that the land included in grassland/pasture class in the CDL is different regionally in the United States. CDL classes containing grass, such as hay or grassland/pasture classes, are the weakest component of the CDL product, which has difficulty distinguishing between different grassland types (Lark et al., 2017). It is probable that different land cover types are placed in the same CDL classes, especially the non-crop classes which are classified with less concern over accuracy than the crop classes (Lark et al., 2017). Another possibility are regional differences in land use for

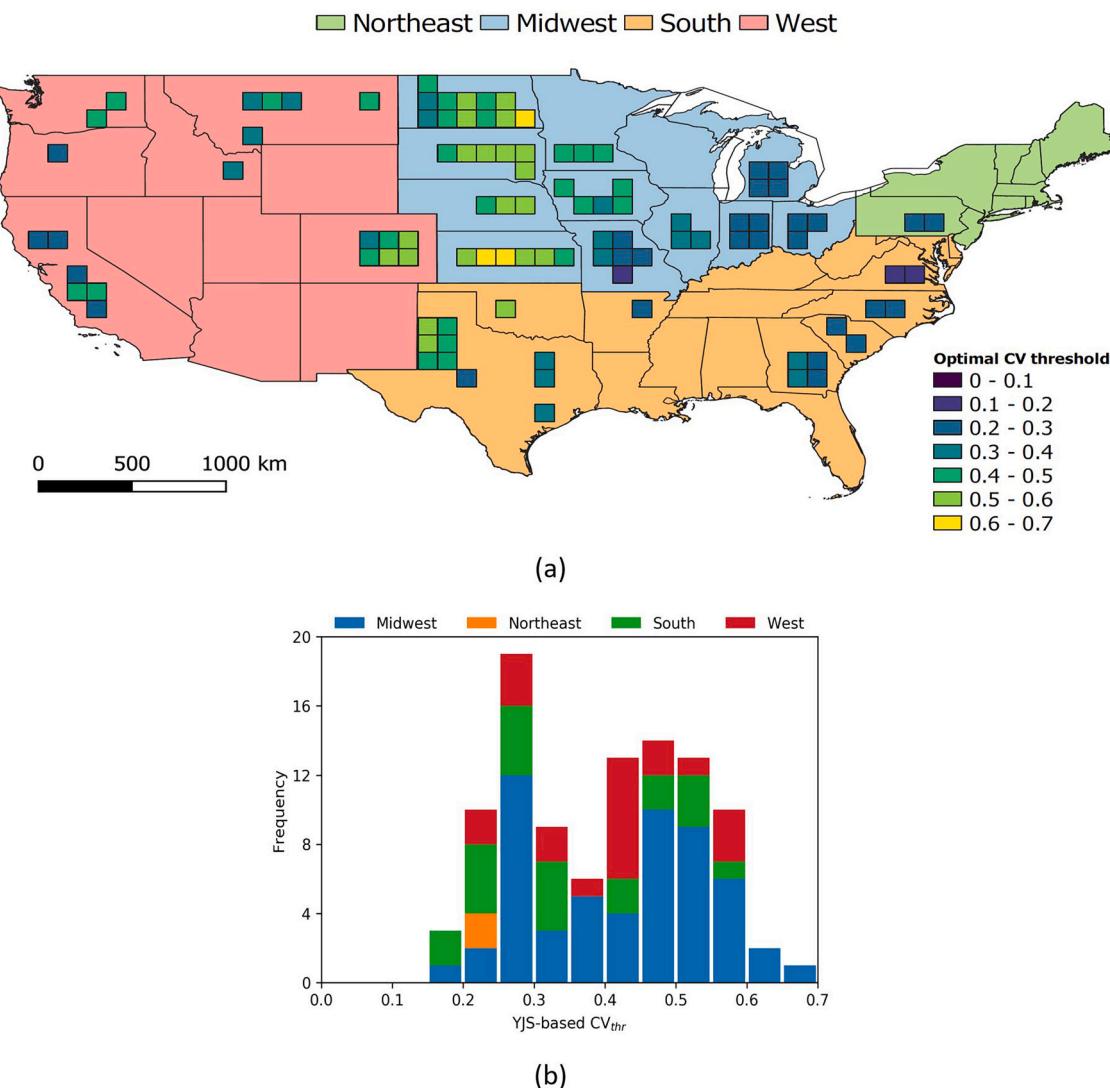


Fig. 6. (a) Map and (b) histogram (0.05 bins) of CV_{thr_YJS} values.

lands classified by the CDL as grassland/pasture that are primarily used for feeding livestock (Butler et al., 2003). In 2017 for example, North Dakota had 44% more beef cattle than California (984,687 as compared to 682,372) according to USDA NASS Quick Stats data, available at <https://quickstats.nass.usda.gov/> (accessed 2 February, 2021). Due to the relatively greater amount of grazing beef cattle, there could have been a relatively more stunted growth of grasslands/pasture in North Dakota as compared to California. A final possibility is that a severe drought that occurred during May to July 2017, severely impacting North Dakota and several other central states. This could also have affected the growth of vegetation at the CDL grassland/pasture pixels (Svoboda et al., 2002). A drought would limit the rate of change (growth) over grasslands/pasture leading to relatively small CV values at this CDL class, potentially explaining why this class classifies better as a non-crop than a crop in the Central United States.

4.3.3. Limitations of YJS for optimizing CV_{thr} values

Section 4.2.1. showed that OA_{YJS} values were similar (within 5%) or better (>5%) than OA_{0.5} at 56 and 35 of the tiles, respectively. The optimization often did not result in substantially better OAs (>5%). And, OA_{YJS} was substantially smaller than OA_{0.5} at 9 tiles, indicating that the YJS optimization has some manner of limitation. Also, as noted in Section 4.2.1, YJS optimization performance can be biased (Smits, 2010), here towards yielding greater improvements at tiles having

greater crop prevalence. This section takes a closer look at these two potential limitations of using YJS optimizations for the binary crop and non-crop classifications.

To verify that the YJS optimization process functions as intended, we plot the relation between the ROC analysis metrics (YJS, AOC) and OA_{YJS} (Fig. 12). YJS and AOC are directly proportional to OA for the binary crop classifications, indicating that the YJS optimization has been implemented properly and works as intended. Fig. 12 shows that good classification accuracies (>80%) are obtained when AUC and YJS are greater than about 0.85 and 0.65, respectively. Four out of the nine cases where OA_{0.5} performed $\geq 5\%$ better than OA_{YJS} occurred even when both AUC and YJS exceeded 0.85 and 0.65, respectively. It is also notable that these tiles corresponded to the lowest crop prevalence bin (0% to 5%, Fig. 2c) and that they are located in each of the different census geographic regions: Virginia (1.8% prevalence), Missouri (0.2% prevalence), Oregon (1.5% prevalence) and Colorado (0.5% prevalence).

Thus, CV_{thr_YJS} values may result in sub-optimal OAs especially at locations of low crop prevalence. Here, the relatively poorer performance of CV_{thr_YJS} may be partially explained due to YJS assigning different cost ratios for classification, depending on the prevalence of the class (Smits, 2010). Whereas the OA metric computations gives equal weight to false positives and false negatives, YJS assigns different weights to each of them, depending on the direction of the imbalance

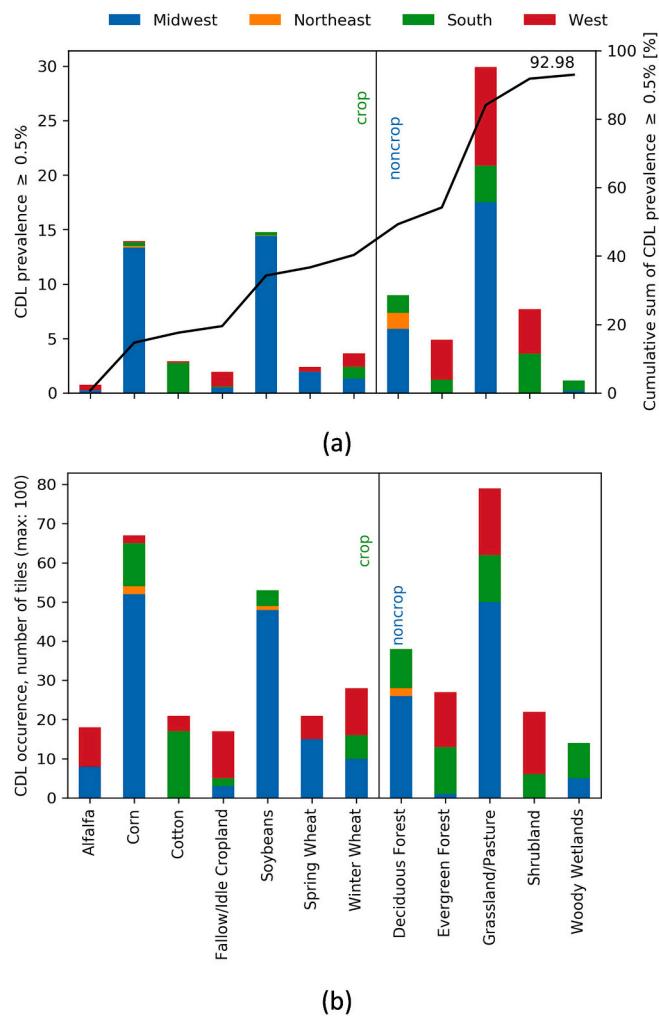


Fig. 7. (a) Prevalence by CDL class and region, after imposing a national scale $\geq 0.5\%$ prevalence threshold. Twelve CDL classes account for 93% of the non-masked data within the 100 tiles that were studied. (b) The number of tiles in which each CDL class was found (meaning it had a $\geq 2\%$ prevalence within tile).

(Smits, 2010). This more clearly shown in Fig. 13, showing that $CV_{thr, 0.5}$ was about as likely to yield better OA values as $CV_{thr, YJS}$ in tiles having lower crop prevalence. Fig. 13 and Table 3 show that OA_{YJS} was

substantially better ($>5\%$) than $OA_{0.5}$ at crop prevalence $\geq 40\%$, but otherwise comparable. Thus, ROIs having $\geq 40\%$ crop cover benefitted greatly from the YJS optimization.

4.3.4. Other limitations

The study has several other limitations. For example, this study did not use an actual ground truth dataset as reference and instead relied on using the USDA CDL layer as reference. Therefore, the reported accuracy metrics only inform on the correspondence of our results with respect to the CDL rather than a ground truth. It was important to use the CDL because it is produced at high spatial resolution, it covers a large region, and it also includes detailed information on non-crop land covers. But since the CDL spatial extent is limited to CONUS, studies that include other regions will need to be assessed using other datasets that might not be as accurate, updated as frequently or stratified into as many land-cover types. A further limitation is that the YJS optimization is computationally more expensive: for example, the fixed threshold approach conducts the classification only once, but the YJS optimization requires many more computations (here, 100) and a reference dataset (here, the CDL) to assess performance metrics for each of the computations, in order to robustly estimate the optimal CV_{thr} value. Thus, we employed an ROI-based sampling strategy (Section 3.5). The limitation here is that because of the ROI-based sampling and screening of data by land cover type and a two hectare mask (Section 3.1), performance metrics were not reported using all available data, and thus OAs over CONUS may still differ somewhat from what was reported here. Another potential concern is that the time series to compute CV values were much greater at some tiles than others (Fig. 2). The basic concept between the temporal CV approach for crop/non-crop classification is that performance will improve as longer time series are considered, which was demonstrated by Huang et al. (2021), albeit with time series limited to about 7 observations (Huang et al., 2021). Thus, it was important for this study to use all the available data at each tile, to obtain the most robust estimates. This resulted in a high variation of the number of available observations between the tiles in this study, ranging between 19 and 107 in this study (Fig. 2a and b). The sampling differences should have some impact on the analysis. But even the least sampled tile used 19 independent observations, which in light of findings by Huang et al. (2021) appears to be more than sufficient for making robust evaluations.

5. Conclusion

This paper demonstrates that using the temporal CV method applied to C-band SAR data can accurately map cropland over large regions, encompassing many different crops, climates and agricultural management practices. The temporal CV method is straightforward, has low

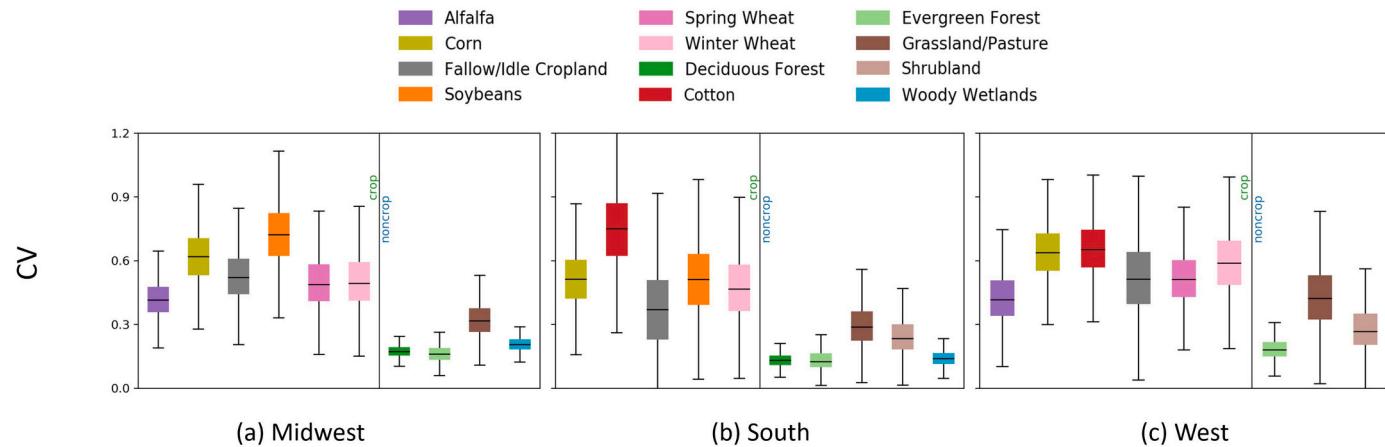


Fig. 8. Boxplots based on the median, first (Q1) and third (Q3) quartiles of CV values of each crop when averaging over the all CDL tiles within each region having more than 10 tiles. The whiskers extend to 1.5 times the interquartile range (Q3-Q1) beyond the upper (Q3) and lower (Q1) edges of the box.

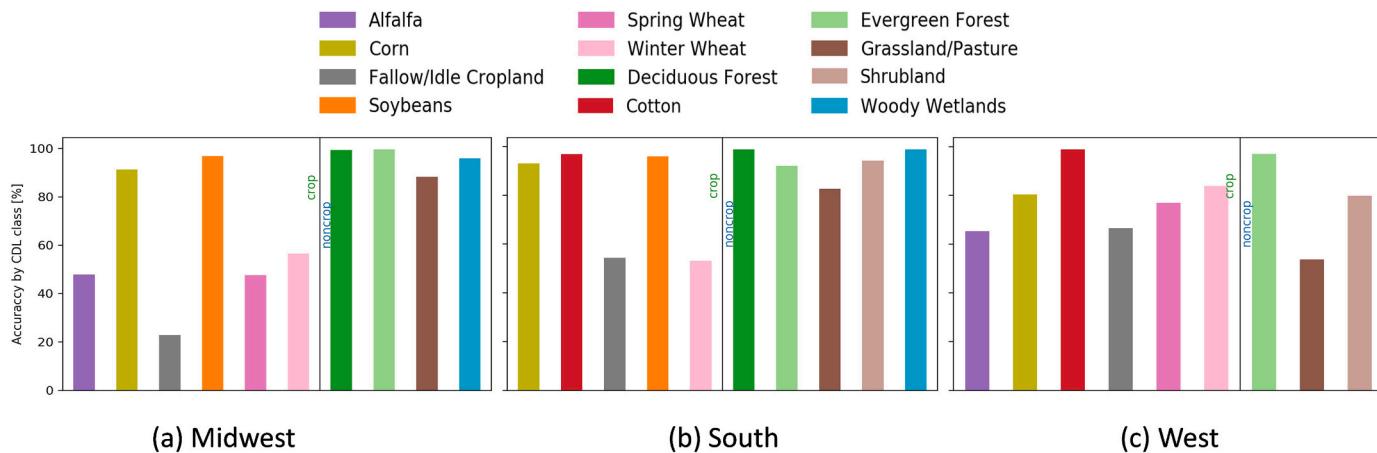


Fig. 9. Accuracies based on optimal thresholds ($CV_{thr,YJS}$) for the most prevalent CDL classes.

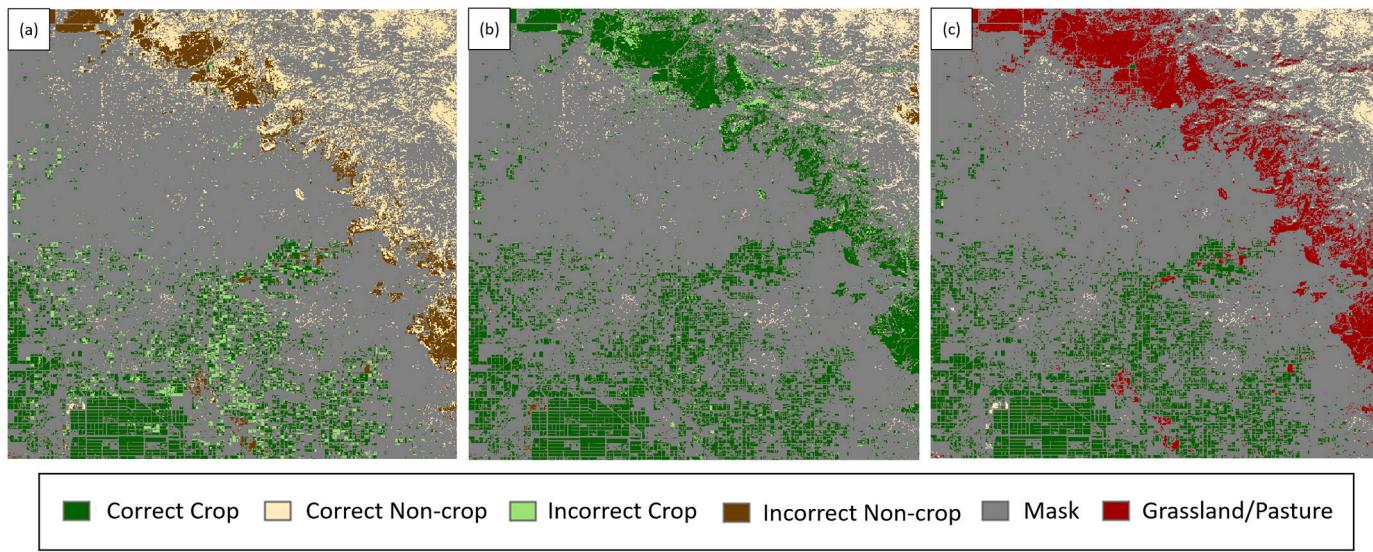


Fig. 10. Classification performance of the northeastern most CA tile ($OA_{YJS} = 70.4\%$ and $CV_{thr} = 0.43$) when: (a) grassland is considered a non-crop which leads to many incorrect crop and non-crop classifications; (b) grassland is considered a crop yielding improved classifications and a substantially larger OA_{YJS} ; (c) Shows the CDL-based crop, non-crop and grassland/pasture classes, respectively colored in green, beige and red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

computational cost and requires no training data. The national scale classification evaluated the use of CV for binary crop/non-crop classification at 100 1 by 1 degree tiles of Sentinel-1A and 1B data (all ascending VH) throughout the conterminous United States, using (1) a constant threshold set at 0.5 for all tiles and (2) using tile-specific CV_{thr} obtained from optimization. OA_{YJS} and $OA_{0.5}$ respectively are, 86.8% and 81.5%, indicating that the computationally inexpensive approach can yield OA values close to the estimated performance ceiling (OA_{YJS}), and that it can exceed a desired OA of 80%. Because $OA_{0.5}$ has $\geq 80\%$ accuracy and is reasonably close to OA_{YJS} , we showed that a single CV_{thr} value can be used for obtaining accurate cropland maps and that it may not be necessary to spend additional effort on exploring other CV_{thr} values or more complicated stratification approaches (i.e., multiple, regional CV_{thr} values). $OA_{0.5}$ (OA_{YJS}) was also stratified by census geographical region, averaging 76.1% (73.5%) in the West and exceeding 80% (90%) in the South and Midwest. We also examined CV values by crop and region and found that the approach is well-suited for making crop and non-crop classifications, but not for identifying individual land cover types. We also showed how the consideration of grassland/pasture as crop or non-crop had a clear impact on OA_{YJS} values (exceeding $\pm 5\%$) at about half the tiles studied, and especially

the West. Grassland/pasture classified best across the United States as a non-crop but results in California greatly improved (by 25%) when grassland/pasture was considered a crop. This study also identified a strong geographic dependence of the optimal CV_{thr} values. CV_{thr} values varied from about 0.2 near the coasts to over 0.6 in the Central United States. CV_{thr} Histograms of showed that values of most tiles fall close to either 0.3 or 0.5. This suggests that crop/non-crop classifications in the United States may be further improved by implementing additional, but regionally specific constant thresholds. There may also be other constant thresholds yielding slightly better results than $CV_{thr} 0.5$. But in light of $OA_{0.5}$ already exceeding 80% and only being about 5% lower than the performance ceiling, $CV_{thr} 0.5$ is a suitable candidate for cropland mapping using Sentinel-1 data in CONUS. Future work should focus on testing this approach over other regions and with respect to other available reference datasets. It could focus more closely on how crop/non-crop classification exactly improves with the density of the time series, to better understand the tradeoff between using more data and OAs, and to assess whether $CV_{thr} 0.5$ performs reasonably well elsewhere. Also, of potential interest is to determine how OAs and CV_{thr} values vary depending on bands used (i.e., L vs. C).

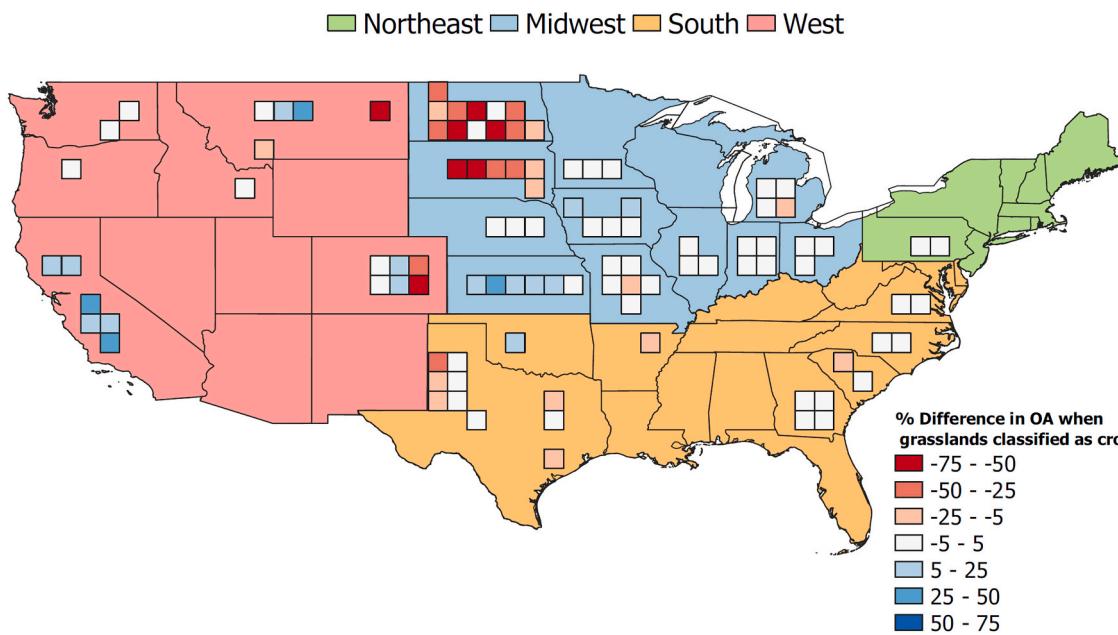


Fig. 11. OA of each tile when the CDL-class grassland/pasture was included as crop in the classification.

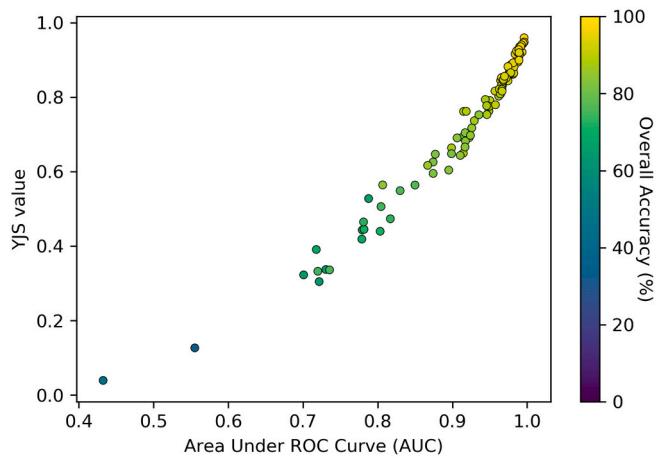


Fig. 12. Comparison between AUC, YJS values and OA (Colour) values.

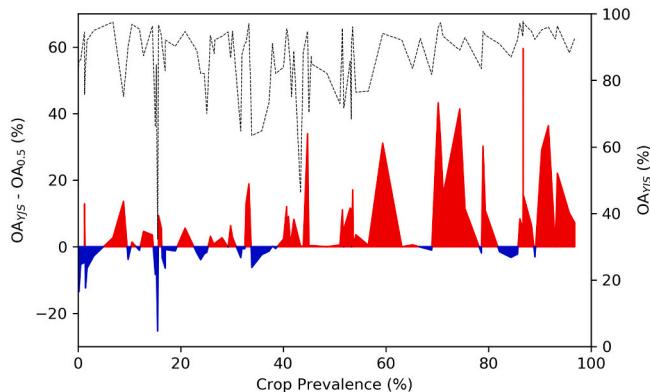


Fig. 13. Difference between OA_{YJS} and OA_{0.5}, stratified by crop prevalence. Red (Blue) colored areas indicate where the YJS CV_{thr} (constant CV_{thr}) performed better. The dashed line represents OA_{YJS}. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

The average OA difference between using the YJS-based CV_{thr} (OA_{YJS}) and CV_{thr} = 0.5 (OA_{0.5}) for the different percent crop quintiles (crop prevalence).

		Percent Crop Quintiles				
Quintile	Quintile Avg. (OA _{YJS} -OA _{0.5})	0-20	20-40	40-60	60-80	80-100
	-2.5%	1.4%	6.3%	15.2%	13.0%	

Author responsibilities

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

Declaration of Competing Interest

None.

Acknowledgements

NASA Headquarters supported this research with funding from the NASA Science Team for the NASA ISRO Synthetic Aperture Radar (NISAR) Mission (NNX16AK59G, 80NSSC19K1497). We also acknowledge funding by USDA ARS grant 58-8042-8-072. The authors would like to thank the European Space Agency and the Alaska Satellite Facility for providing the Sentinel-1 data. We would like to thank the three anonymous reviewers that provided us with detailed and helpful comments on this work.

References

- Abdikan, S., Sanli, F.B., Ustuner, M., Calò, F., 2016. Land cover mapping using sentinel-1 SAR data. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* 41, 757.
- Anderson, J.R., Hardy, E.E., Roach, J.T., Witmer, R.E., 1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. Professional Paper. <https://doi.org/10.3133/pp964>.
- Arias, M., Campo-Bescós, M.Á., Álvarez-Mozos, J., 2020. Crop classification based on temporal signatures of sentinel-1 observations over Navarre Province, Spain. *Remote Sens.* 12, 278.
- Becker-Reshef, I., Barker, B., Humber, M., Puricelli, E., Sanchez, A., Sahajpal, R., McGaughey, K., Justice, C., Baruth, B., Wu, B., 2019. The GEOGLAM crop monitor

- for AMIS: assessing crop conditions in the context of global markets. *Glob. Food Sec.* 23, 173–181.
- Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* 204, 509–523.
- Bigelow, D., Borchers, A., 2017. Major Uses of Land in the United States, p. 2012.
- 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 Int.* 26, 341–358. <https://doi.org/10.1080/10106049.2011.562309>.
- Bureau, U.S.C, 2010. Census Regions and Divisions of the United States. US Census Bur. Website.
- Butler, L., Cropper, J., Johnson, R., Norman, A., Peacock, G., Shaver, P., Spaeth, K., 2003. National Range and Pasture Handbook. USDA Natl. Resour. Conserv. Serv., Washington, DC, USA, p. 214.
- Fan, J., Upadhye, S., Worster, A., 2006. Understanding receiver operating characteristic (ROC) curves. *Can. J. Emerg. Med.* 8, 19–20.
- 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 Trans. Geosci. Remote Sens.* 35, 5–17.
- Fick, G.W., Mueller, S.C., 1989. Alfalfa: Quality, Maturity, and Mean Stage of Development.
- Fisette, T., Davidson, A., Daneshfar, B., Rollin, P., Aly, Z., Campbell, L., 2014. Annual space-based crop inventory for Canada: 2009–2014. In: 2014 IEEE Geoscience and Remote Sensing Symposium. IEEE, pp. 5095–5098.
- Flach, P.A., Hernández-Orallo, J., Ramírez, C.F., 2011. A coherent interpretation of AUC as a measure of aggregated classification performance. *ICML*.
- Fontanelli, G., Crema, A., Azar, R., Stroppiana, D., Villa, P., Boschetti, M., 2014. Agricultural crop mapping using optical and SAR multi-temporal seasonal data: a case study in Lombardy region, Italy. In: 2014 IEEE Geoscience and Remote Sensing Symposium. IEEE, pp. 1489–1492.
- Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sens. Environ.* 83, 287–302.
- Goetz, S.J., Fiske, G.J., Bunn, A.G., 2006. Using satellite time-series data sets to analyze fire disturbance and forest recovery across Canada. *Remote Sens. Environ.* 101, 352–365.
- Habibzadeh, F., Habibzadeh, P., Yadollahie, M., 2016. On determining the most appropriate test cut-off value: the case of tests with continuous results. *Biochem. Med. Biochem. Med.* 26, 297–307.
- Heimlich, R.E., 2000. Farm Resource Regions. United States Department of Agriculture, Economic Research Service.
- Hong, G., Zhang, A., Zhou, F., Brisco, B., 2014. Integration of optical and synthetic aperture radar (SAR) images to differentiate grassland and alfalfa in Prairie area. *Int. J. Appl. Earth Obs. Geoinf.* 28, 12–19.
- Huang, X., Liao, C., Xing, M., Ziniti, B., Wang, J., Shang, J., Liu, J., Dong, T., Xie, Q., Torbick, N., 2019. A multi-temporal binary-tree classification using polarimetric RADARSAT-2 imagery. *Remote Sens. Environ.* 235, 111478.
- Huang, X., Reba, M., Coffin, A., Runkle, B.R., Huang, Y., Chapman, B.D., Ziniti, B., Skakun, S., Kraatz, S., Siqueira, P., Torbick, N., 2021. Cropland mapping with L-band UAVSAR and development of NISAR products. *Remote Sens. Environ.* <https://doi.org/10.1016/j.rse.2020.111280>.
- Johnson, D.M., 2013. A 2010 map estimate of annually tilled cropland within the conterminous United States. *Agric. Syst.* 114, 95–105.
- Justice, C.O., Román, M.O., Csiszar, I., Vermote, E.F., Wolfe, R.E., Hook, S.J., Friedl, M., Wang, Z., Schaaf, C.B., Miura, T., 2013. Land and cryosphere products from Suomi NPP VIIRS: overview and status. *J. Geophys. Res. Atmos.* 118, 9753–9765.
- Kraatz, S., Cosh, M., Torbick, N., Huang, X., Siqueira, P., 2020. Evaluating the robustness of NISAR's cropland algorithm to time of observation, observing mode and dithering. In: AGU Fall Meeting Abstracts.
- Kraatz, S., Rose, S., Cosh, M., Torbick, N., Huang, X., Siqueira, P., 2021a. Performance evaluation of UAVSAR and simulated NISAR data for crop/non-crop classification over Stoneville, MS. *Earth Sp. Sci.* 8 <https://doi.org/10.1029/2020EA001363>.
- Kraatz, S., Torbick, N., Jiao, X., Huang, X., Dingle Robertson, L., Davidson, A., McNairn, H., Cosh, M.H., Siqueira, P., 2021b. Comparison between dense L-band and C-band Synthetic Aperture Radar (SAR) time series for crop area mapping over a NISAR calibration-validation site. *Agronomy* 11. <https://doi.org/10.3390/agronomy11020273>.
- Lark, T.J., Mueller, R.M., Johnson, D.M., Gibbs, H.K., 2017. Measuring land-use and land-cover change using the US department of agriculture's cropland data layer: cautions and recommendations. *Int. J. Appl. Earth Obs. Geoinf.* 62, 224–235.
- Lindsay, E.J., King, D.J., Davidson, A.M., Daneshfar, B., 2019. Canadian prairie rangeland and seeded forage classification using multiseason Landsat 8 and Summer RADARSAT-2. *Rangel. Ecol. Manag.* 72, 92–102.
- McNairn, H., Brisco, B., 2004. The application of C-band polarimetric SAR for agriculture: a review. *Can. J. Remote. Sens.* 30, 525–542.
- McNairn, H., Champagne, C., Shang, J., Holmstrom, D., Reichert, G., 2009a. Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories. *ISPRS J. Photogramm. Remote Sens.* 64, 434–449.
- McNairn, H., Shang, J., Jiao, X., Champagne, C., 2009b. The contribution of ALOS PALSAR multipolarization and polarimetric data to crop classification. *IEEE Trans. Geosci. Remote Sens.* 47, 3981–3992.
- Moran, M.S., Hymer, D.C., Qi, J., Kerr, Y., 2002. Comparison of ERS-2 SAR and Landsat TM imagery for monitoring agricultural crop and soil conditions. *Remote Sens. Environ.* 79, 243–252.
- Nguyen, D.B., Wagner, W., 2017. European rice cropland mapping with Sentinel-1 data: the Mediterranean region case study. *Water* 9, 392.
- Pittman, K., Hansen, M.C., Becker-Reshef, I., Potapov, P.V., Justice, C.O., 2010. Estimating global cropland extent with multi-year MODIS data. *Remote Sens.* 2, 1844–1863.
- Quegan, S., Yu, J.J., 2001. Filtering of multichannel SAR images. *IEEE Trans. Geosci. Remote Sens.* 39, 2373–2379.
- Shang, J., McNairn, H., Champagne, C., Jiao, X., 2008. Contribution of multi-frequency, multi-sensor, and multi-temporal radar data to operational annual crop mapping. In: IGARSS 2008–2008 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 3–378.
- Small, D., 2011. Flattening gamma: radiometric terrain correction for SAR imagery. *IEEE Trans. Geosci. Remote Sens.* 49, 3081–3093.
- Smits, N., 2010. A note on Youden's Jand its cost ratio. *BMC Med. Res. Methodol.* 10, 89.
- Spiegel, S., Bestelmeyer, B.T., Archer, D.W., Augustine, D.J., Boughton, E.H., Boughton, R.K., Cavigelli, M.A., Clark, P.E., Derner, J.D., Duncan, E.W., 2018. Evaluating strategies for sustainable intensification of US agriculture through the Long-Term Agroecosystem Research network. *Environ. Res. Lett.* 13, 34031.
- Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., 2002. The drought monitor. *Bull. Am. Meteorol. Soc.* 83, 1181–1190.
- Teluguntla, P., Thenkabail, P.S., Xiong, J., Gumma, M.K., Congalton, R.G., Oliphant, J.A., Sankey, T., Poehnelt, J., Yadav, K., Massey, R., 2017. NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-Support Analysis Data (GFSAD) Cropland Extent 2015 Australia, New Zealand, China, Mongolia 30 M V001.
- Torbick, N., Chowdhury, D., Salas, W., Qi, J., 2017. Monitoring rice agriculture across Myanmar using time series Sentinel-1 assisted by Landsat-8 and PALSAR-2. *Remote Sens.* 9, 119.
- Torres, R., Snoej, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Flouri, N., Brown, M., 2012. GMES Sentinel-1 mission. *Remote Sens. Environ.* 120, 9–24.
- Waske, B., Braun, M., 2009. Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS J. Photogramm. Remote Sens.* 64, 450–457.
- Werner, C., Wegmüller, U., Strozzi, T., Wiesmann, A., 2000. Gamma SAR and interferometric processing software. In: Proceedings of the Ers-Envisat Symposium, Gothenburg, Sweden. Citeseer, p. 1620.
- 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). IEEE, pp. 5342–5345.
- Whelen, T., Siqueira, P., 2018a. Time-series classification of Sentinel-1 agricultural data over North Dakota. *Remote Sens. Lett.* 9, 411–420.
- Whelen, T., Siqueira, P., 2018b. Coefficient of variation for use in crop area classification across multiple climates. *Int. J. Appl. Earth Obs. Geoinf.* 67, 114–122.
- Yin, H., Brandão Jr., A., Buchner, J., Helmers, D., Iuliano, B.G., Kimambo, N.E., Lewińska, K.E., Razenková, E., Rizayeva, A., Rogova, N., 2020. Monitoring cropland abandonment with Landsat time series. *Remote Sens. Environ.* 246, 111873.
- Youden, W.J., 1950. Index for rating diagnostic tests. *Cancer* 3, 32–35.