ISPRS Journal of Photogrammetry and Remote Sensing 212 (2024) 164–180

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Effect of intra-year Landsat scene availability in land cover land use classification in the conterminous United States using deep neural networks

Giorgos Mountrakis *, Shahriar S. Heydari

Department of Environmental Resources Engineering, State University of New York, College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210, United States

ARTICLE INFO

Keywords:
Deep learning
Time-series
Classification
Landsat
United States

ABSTRACT

The Landsat archive having consistent revisit times, near global extent and extensive multi-decadal temporal coverage offers a unique opportunity for land cover land use product generation. Along with this vast volume of freely available data, new classification methods based on deep learning have improved modeling capabilities. This manuscript investigates the effect of intra-annual Landsat scene availability in the accuracy of land cover land use classification in the conterminous United States. More specifically, we seek to quantify the effect of: i) increased monthly scene availability, and ii) specific months that may result in higher classification accuracy across different classes. Identifying specific months with comparable classification accuracy to the entire time series could offer significant computational gains for large-scale mapping. Our experiment incorporated deep learning classifiers and a wide range of reference data across the continental United States. Results were contrasted between five large U.S. climatic regions to further differentiate this intra-annual effect. Our findings indicate that the total number of months can have a highly variable effect in the classification accuracy ranging from minor (a few percentage points in terms of class F1 accuracy) to extremely beneficial (approaching 50% F1 improvement moving from four to twelve month observations). The benefit of increased month observations varied among climatic regions and classes: when all climate regions were combined, the grass/shrub and cultivated classes improved their F1 accuracy up to 30%, while the water class saw the least improvement of about 5%, partially due to its limited room for improvement. The effect of specific month combinations was also examined, where the total number of months was kept constant and the included months varied. The difference between the best month combination and the median combination value was estimated to be as high as about 30% for the four monthly observations scenario and the grass/shrub class. Further validation of the month selection importance comes from an example implementation scenario where F1 improvements can be as high as 10%. Our work demonstrated that month selection may offer such benefits that in some classes and climatic regions this time selection optimization is an inevitable choice due to large accuracy improvements. Also, the potential data reduction with targeted month selection would be particularly appealing to large-scale classification tasks. Due to the large extent of the climatic regions further studies are needed to quantify a more localized effect along with explanation of potential drivers.

1. Introduction

Land Cover Land Use (LCLU) classification using remote sensing observations has been a popular and well-studied research topic over the past decades. Currently, multiple remote sensing platforms exist that provide global or near-global earth surface observations. This provides a plethora of possibilities to combine data from different dimensions (spatial, temporal, spectral) and acquisition technologies (optical, radar, lidar). However, when multi-decadal studies are the focus, Landsat is the only mission series that provides consistent, continuous global coverage of the earth surface for almost five decades. Justifiably so, Landsat is known as the workhorse satellite series when it comes to land imaging. Example Landsat-based products include USGS's national land cover database (NLCD) maps created for selected years from 2001 till 2019 (USGS, 2018) plus a legacy product for 1992 (Dewitz, 2000). Other continental or global examples include Europa's Corine land cover maps for the years 1990 and 2000 (EC JRC, n.d.), and GlobeLand30 global land cover map for the years 2000 and 2010 (J. Chen et al., 2017). A

* Corresponding author.
E-mail address: gm@esf.edu (G. Mountrakis).

https://doi.org/10.1016/j.isprsjprs.2024.04.027
Received 15 January 2024; Received in revised form 25 March 2024; Accepted 26 April 2024

comparison of products over the continental U.S. is available in Wang and Mountrakis (2023).

One of Landsat's benefits is the near bi-weekly revisit time from a single sensor or more frequent for years with overlapping operational sensors. This high temporal frequency offers additional modeling capabilities. For example, Julien, Sobrino, and Jiménez-Muñoz (2011) used multi-temporal Landsat data to generate a model for land cover dynamics and derived parameters for classification. Zhu and Woodcock (2014) presented their CCDC model, which takes into account dense Landsat observations. Lymburner et al. (2013) and Shen and Evans (2020) combined different Landsat mission data to improve their vegetation spectral signature model, and Singh and Gray (2020) took advantage of multi-temporal observations and spectral unmixing to resolve endmembers. Y. Chen et al. (2022) used stacks of Landsat imagery for cloud removal.

The above examples provide a sample of the application breadth for the use of dense Landsat data in modeling land cover dynamics. Since our work focuses on direct use of dense multi-temporal Landsat information in classification tasks we focus further on these approaches. Traditionally, direct use of dense multitemporal information was based on calculating spectral-temporal metrics (STMs) to reduce a variable-length time series to a fixed number of statistical metrics used as classification inputs. This can be done on original bands data or derived spectral indices such as NDVI, see examples by Bauer, Yuan, and Sawaya (2004), Zheng, Campbell, and De Beurs (2012), or McInerney et al. (2019). Another approach reporting gains is to concatenate a vector of a number of observations (or derived features) and use a multitemporal sequence to conduct the classification A common practice in this case was the combination of Leaf-on / Leaf-off or wet / dry season observations to enhance classification performance. Examples include forest classification by Wolter et al. (1995), wetland classification by Frohn et al. (2012), or general land use classification by Kantakumar and Neelamsetti (2015).

In recent years, our capabilities to incorporate more temporally dense Landsat information in the classification have increased. Firstly, the Landsat archive has been reprocessed and access to the full temporal span of Landsat is provided in a consistent manner through USGS Landsat 4–8 Analysis Data Ready (ARD) product line (Dwyer et al., 2018). The ARD product not only uses enhanced processing on acquired Landsat scenes to increase their radiometric and geometric accuracy, but also includes reprojection and combination of multiple overlapping scenes within a unified square tiling system. Preprocessing tasks and classification activities have moved to the cloud (e.g., using Google Earth Engine and Google Colab) thus democratizing access and processing capabilities. Secondly, with the proliferation of deep learning methods, classification algorithms can more effectively take advantage of temporally dense information (see a review of deep learning applications in remote sensing in Heydari and Mountrakis, 2019). Examples of time-dimension processing deep networks are gaining momentum in remote sensing classification applications, starting from memory-based network (e.g., employing Long Short Term Memory cells), attention-based or transformer mechanisms. Memory-based networks look at the entire time-series instead of focusing on just one observation and try to extract useful features. They first appeared in the land cover classification literature in 2017 (for example see Rußwurm and Körner, 2017). Attention-based mechanisms go one step further and try to detect samples with higher importance within the sequence and focus on them. The transformers utilize attention layers in an encoder-decoder architecture. These advanced models were developed in other deep learning fields and have recently transitioned in remote sensing applications (for example see Jannat and Willis, 2022).

Due to variability in geographic availability and quality of Landsat observations through time, the effect of the observations number and timing (especially for calculating spectral-temporal metrics (STMs)) has been of interest to many researchers. One of the early studies is by Guerschman et al. (2003), which used multitemporal Landsat data in Argentina and combined two, three, or four observations (one from each season) showing better results when using multiple dates. Rufin et al. (2015) calculated the drop in Amazonian deforestation mapping accuracy by randomly dropping a different number of yearly observations when computing the STMs used for classification. Hansen et al. (2016) mapped tree-height distribution in Africa using Landsat data and correlated mapping accuracy to Landsat observation counts. (Griffiths et al., (2019) processed over a year of harmonized Sentinel/Landsat data over Germany to produce 10-day, monthly, and seasonal composites for land cover classification and found the accuracy to be highest for 10-day composites and lowest for seasonal composites. X. Zhang et al. (2019) used an ensemble of RF classifiers to classify each Landsat observation to a land cover type and aggregated the results by a second stage RF classifier to map the entire China. Accuracy increased by the number of observations to a maximum of 21 observations and then decreased, but no month or seasonal analysis was performed. Recently, Frantz et al. (2023) looked at the availability of Landsat observations globally for different world climate zones and estimated the number of observations/gap between the observations required to get a reliable value of an STM to use in further classifications.

Season-based observation binning and mixing was also considered. For example, Liu et al. (2015) found that vegetation/open land cover classification in their study area in Burkina Faso is more accurate when using dry vs wet season imagery, but they did not find significant improvement in concurrent dry and wet imagery usage. Senf et al. (2015) considered a 5-class land use (agriculture/natural land) classification for an area in Portugal based on six Landsat images, each acquired in a different month in 2011. Best single month and two-month (March/September) combinations were identified along with their implications for different land cover types. Li et al. (2017) tackled 11-class global land cover mapping and used a specific number of samples from each month to contrast accuracy between of a single month, seasonal, and complete samples. Specific months and seasons with better performance were identified, but optimal multi-month combination was not studied. Karakizi et al. (2018) considered 4 datasets composed of 11 to 18 scenes with different scene cloud cover percentage (not sorted in time or season) in 2016 for an area in Greece, and generated detailed land cover maps based on each of 4 scenario datasets and calculated change in accuracy metrics by including more scenes. Zhang et al. (2019) analyzed the importance of different bands for each of their five Landsat image scenes when used individually or together, and found two specific bands and two specific dates to be more important than the others. At a continental scale, Pflugmacher et al. (2019) used some different sets of STMs including both yearly and seasonally (over 4 different ranges of months) to classify land covers over Europe.

However, most of the past studies were either working on derived statistics and not full time series, or did not look at the effect of including specific month observations and their combination on classification accuracy. Particularly, elaboration on the impacts of month selection on specific land covers and advising on the optimum configuration for different land covers is rarely studied. The scope of some researches was global and some was local, but there was no study to analyze the importance of monthly observations over a large continental area in southern or northern hemisphere such as conterminous United States. Here, we fill this gap and explicitly study the effect of intra-annual Landsat scene availability in the accuracy of land cover land use classification. More specifically, we seek to answer these research questions:

- What is the accuracy benefit of increased monthly scene availability?
- Which calendar months result in higher classification accuracy across different classes?
- Can the intra-month accuracy benefit be quantified across different classes and continental US regions?

The rest of the manuscript contains a description of our reference dataset in section 2, section 3 presents the methods used to conduct the

experiment, followed by results in section 4 and discussion in section 5.

2. Study area and reference dataset

The study area covers 84 blocks of size 10 km x 10 km (333x333 pixels with spatial resolution of 30 m), selected from an original dataset of 2717 land cover blocks that were produced for USGS Land Cover Trends Project (https://www.usgs.gov/centers/wgsc/science/land-cover-trends). Due to the extensive manual labeling conducted to create the reference dataset 84 representative blocks were selected from the 2717 available. The 84 sample blocks were selected in such a way that each one of them fall into one of 84 level III EPA ecoregions (see https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states). The land cover labels were re-inspected and validated in a previous study over years of 2005–2019. For more information on block processing and land cover assignment and validation see Mountrakis and Heydari (2023). For the purpose of this research, all samples were aggregated in climate regions to access classification accuracy across different climatic conditions using the Köppen-Geiger climatic classification scheme 1 km maps (Beck et al., 2018). Considering the limited amount of our model validation points in each individual Köppen-Geiger climatic class, we further aggregated these classes into five general climatic regions (Table 1). The resulting climatic regions (CR) map is shown in Fig. 1.

In this manuscript pretrained deep learning classifiers are used from our priori work (Mountrakis and Heydari, 2023). These classifiers were trained on a large 20 M sample dataset. The validation dataset, which was spatially independent from the training dataset, contained 2,297,679 samples. This was the combined result of samples from each of the 84 blocks and represents a single Landsat cell. Each sample is a sequence of all available Landsat Surface Reflectance (SR) Tier-1 observations for a given year and topographic data, accessed and processed via Google Earth Engine platform. Landsat SR Tier-1 data is already processed for atmospheric errors but we also used the *radsat_qa* and *pixel_qa* quality bits to further remove saturated pixels and cloud/cloud shadow cases. Landsat observations were taken from all available Landsat 5, 7, and 8 missions but limited to six bands of Blue, Green, Red, NIR, SWIR1, and SWIR2. Pixel topographic variables including elevation, slope, and aspect are also calculated from the Shuttle Radar Topography Mission (SRTM) digital elevation data and added to Landsat data. Furthermore, eight different spectral indices were added targeting vegetation, water, built-up area, bare soil, and soil wetness along with 32 spatial features from Gray-Level Co-occurrence Matrix (GLCM) analysis (see Mountrakis and Heydari (2023) for a detailed list).The length of the sequence (i.e., the number of Landsat scenes for that year for a given sample) can vary significantly depending on cloud coverage and Landsat sensor availability. To conduct our multi-temporal analysis only samples that contained at least one valid observation for each month of the sequence's year were selected. In cases where multiple observations were available within a given month only one observation from that month was selected randomly. After this filtering, the reduced dataset contained 540,078 samples, each containing the same sequence length (12 months) with only a single observation per month.

Each sample had one of seven land use classes associated with it: Water, Impervious, Grass/shrub, Forest, Bare, Cultivated, and Wetland. Land use class definitions are given in appendix A, and these definitions were used in model building carried out in our previous work (Mountrakis and Heydari, 2023). Distribution of these classes varied between climatic regions (CRs). To make results across climatic regions comparable, the reduced dataset is randomly sampled aiming to extract approximately 1000 samples per land use class per climatic region to have the total number of samples for each class equal to 5,000 (and total number of samples close to 35,000 for the seven classes) as summarized in Table 2. The bare class was considered only for the Arid climatic region in further processing due to insufficient samples in other CRs. For the other four climatic regions the missing share from the bare class was filled by samples from other classes to keep the total samples per class across all CRs equal at 5000. Therefore, our final dataset consists of 35,000 samples well distributed across climatic regions and classes and containing full 12-month complete sequences. The uneven bare class sample distribution may have a negative impact on the performance of other classes, especially those that are spectrally similar (e.g. impervious). However, overall our analysis is considered appropriate.

Table 1
Aggregated five climatic regions in our study.

#	Description	Corresponding Köppen-Geiger class code
1	Temperate, no dry season	1,2,3,14,15,16
2	Arid	4,5,6,7
3	Temperate, dry summer	8,9,10
4	Cold, dry/hot summer	17,18,19,21,22,25
5	Cold, no dry season	26,27,29

3. Methods

3.1. Classifiers

Predicted maps from pre-trained classifiers were used to study the effect of scene availability in our experiments. These deep learning classifiers were trained for LCLU classification using a large number of samples from the conterminous United States in our prior research as presented in appendix B but not detailed here because their development is not part of the methods in this paper. Shortly, the architecture consisted of a two-layer convolutional neural network to extract spatial features from input data, followed by a recurrent neural network to extract features related to the temporal dimension. A multilayer neural network processed the static invariant data (such as topography) and another multilayer network processed the combination of the aforementioned outputs and generated the final features to be used to assign the final class label. It is important to note that the combination of convolutional and recurrent layers allowed all models to capture simultaneously both the spatial and temporal dimensions of spectral variability.

To increase the confidence of the obtained results 40 different deep learning classifiers were selected. These 40 models all have the same general structure presented before, but with different numbers of CNN filters, LSTM cells, MLP neurons, or other optimization parameters (the details are presented in appendix B). Our prior simulations showed very similar classification accuracy across the 40 models, thus ensuring that accuracy variability would be driven by input scene availability, not classifier limitations.

3.2. Month combinations

Based on the selected 35,000 full 12-month samples discussed in section 2, we generated specific month combinations for a given number of months for each sample and created a combined months dataset. This combined month dataset is tested by the classifiers introduced in section 3.1, and the classifiers predictions are compared to reference values. For example, when four-month combinations were examined, there were 495 unique four-month combinations (e.g., Feb + Apr + Sep + Oct) possible from the 12-month pool (Jan through Dec). The number of combinations was 924, 495, and 66 for six, eight, and ten months, respectively, while the twelve months had a single combination. Therefore, for each sample the total combinations considered was 1981 leading to a total pool of 69,335,000 simulations for the 35,000 samples.

4. Results

Each of the 40 different deep learning classifiers was simulated 69 M times. F1 class accuracy metrics were obtained and organized in climatic regions for reporting and discussion.

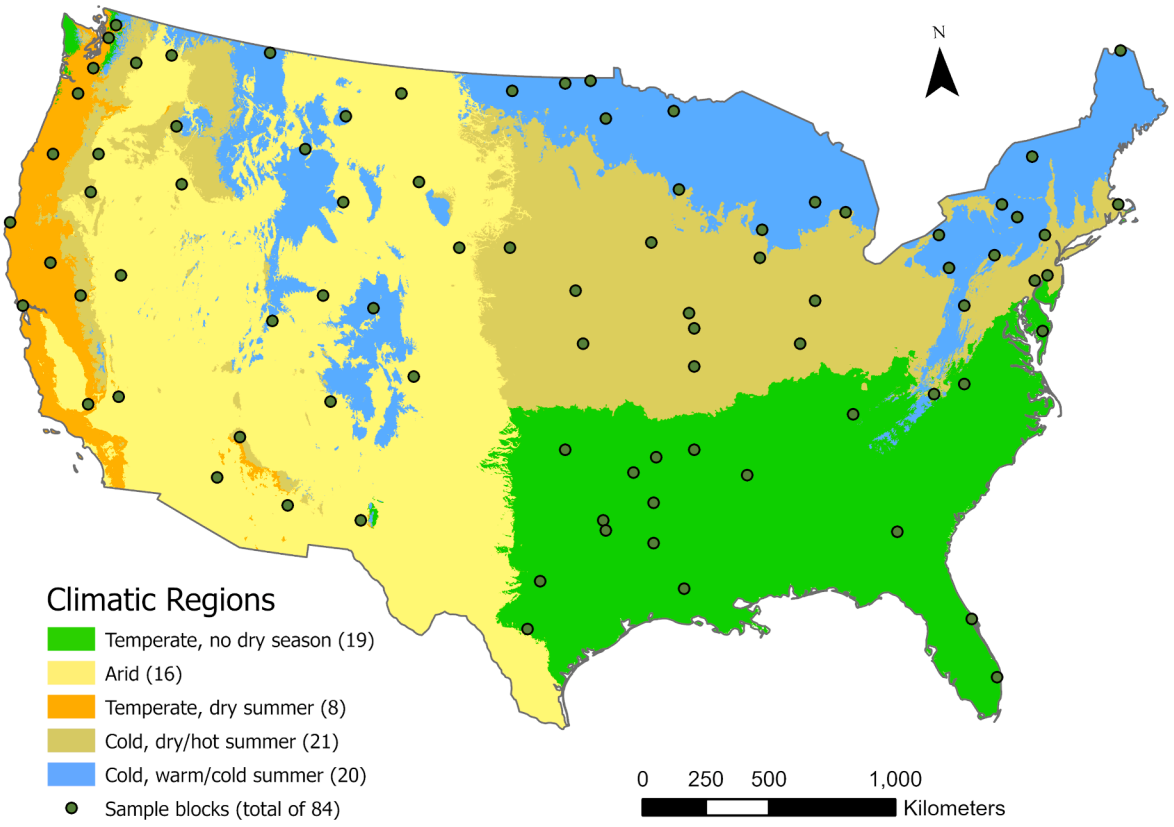

Fig. 1. Climatic regions and sampling blocks.

Table 2
Number of complete 12-month samples for each class and climatic region.

	Support (number of complete 12-month samples)						
Climatic Region	Water	Impervious	Grass/shrub	Forest	Bare	Cultivated	Wetland
Temperate, no dry season	1000	1000	1000	1000	72	1000	1273
Arid	1000	1000	1000	1000	4579	1000	1000
Temperate, dry summer	1000	1000	1000	1000	40	1000	1000
Cold, dry/hot summer	1311	1000	1000	1000	289	1000	727
Cold, no dry season	689	1000	1000	1000	20	1000	1000
Total (35,000):	5000	5000	5000	5000	5000	5000	5000

4.1. Accuracy effect of total month scene availability per class and climatic region

The first objective was the examination of the effect of total scene availability. Results were grouped separately for each climatic region and class. Fig. 2 presents F1 class accuracy for different total month number combinations, from 4 to 12. Using less than 4 months was not considered as it would not be taking advantage of the multitemporal intrinsic value of the Landsat archive. First, each of the 40 classifiers was simulated on all possible month combinations for a given total month number (e.g. 495 combinations for N4) for each sample that falls in that climatic region and class (e.g. 1000 samples for the water class in the arid region). The class F1 accuracy was calculated for each of the possible month combinations (e.g. 495 combinations for N4) and the median F1 value was extracted. Each of the 40 classifiers generated a single median F1 accuracy for a given class in a specific climatic region and for a specific total month combination. The variability of these 40 classifiers is presented as the boxplot range. The extraction of the median and not the best F1 accuracy for each possible month combination (e.g. median, not best, of the 495 combinations for N4) results in the boxplots depicting the average, not optimal, performance of each total month combination. Therefore, results should be interpreted in a relative comparative manner, not in isolation as the best performance for each region/class. Further analysis of the specific month combinations for a given total month number is presented in the next section.

The effect of increasing temporal information varies by region and class:

- The water class benefits the least from multitemporal information. It is well-documented that water is fairly straightforward to extract using Landsat's infrared bands.
- The impervious class improvements vary from considerable (~10 %) for the two temperate CRs and the cold, dry/hot summer CR, to substantial (~20 %) in the cold, no dry season CR to immense (~35 %) for the arid CR.
- The forest class gains are considerable (5–10 %) for all regions with substantial (~15 %) improvements in the cold, no dry season.
- The wetland class improvements are considerable (5–10 %) for three CRs, with immense improvements for the cold, no dry (~30 %) and arid (~50 %) CRs. For the latter two CRs the accuracy gains from four to six month combinations are particularly notable demonstrating the quick improvements for a lengthier temporal record.

Fig. 2. Variation of F1 class accuracy within 40 models for different CRs and different total month number combinations.

- The grass/shrub and cultivated classes exhibit similar behavior, an expected result due to their spectral mixing, especially for some months of the year. Improvements are interchangeable and substantial across all CRs. Of particular note is the grass/shrub improvement of almost 40 % from 4 months to the full 12 month sequence in the cold/dry region.
- The bare class, that was examined only in the arid CR due to lack of strong presence in the other CRs, also shows substantial improvements, partially attributed to improvements to other spectrally similar classes such as the impervious and cultivated classes.

4.2. Accuracy effect of specific month scene availability per class and climatic region

In this section we seek to identify the specific months that offer the highest accuracy improvements within a given total month scenario. Previously the median F1 value across all possible combinations for each of the possible month combinations was extracted for each class (e.g. median of 495 combinations for N4). Here, we ranked all possible combinations and extracted the highest performing ones for each class using the 90-percentile value of F1 as the inclusion threshold. For example, in the case of N4, the 495 possible four-month combinations

were ranked based on their F1 accuracy and the top 10 % combinations (49) were extracted. The process was repeated for each of the 40 classifiers, and all of selected month combinations were aggregated. Then the frequency of occurrence of each specific month (e.g. Jan) among them was counted, and the result is converted to a percentage to compare between the effect of different months. For example, if a month appears in all selected month combinations, then it's percentage will be 10.

Tables 3, 4 and 5 present the results for four, six and eight month combinations, respectively. The color-coded 12 columns on the right show the relatively presence of a particular month in those top 10 % combinations. These tables organize findings by class and then by climatic region (CR). A reordered version of these tables by individual climatic regions and then classes is provided in appendix C. The column F1(Max) shows the maximum of calculated F1 value in each row, and the column F1(Med-Max) provides the accuracy difference between the median and maximum value calculated for that row. It is an important starting point in the analysis as it quantifies the impact of the month selection. A low difference indicates that month selection is not considerably affecting the results for that particular CR and class, therefore specific month analysis is not needed. The patterns in these tables generally match, particularly 6 and 8 best months as they are probably better indicators than the 4-month table. For example, for Impervious class of Cold, dry/hot summer CR, Aug is an important month for all 4/6/8 months. For the same class in Arid CR, best months for 6- and 8-month scenario match but 4 month scenario is a bit different.

A first general remark is that the highly influential months may vary for a given class from one CR to another. Furthermore, some classes and CRs benefit from specific month combinations while in others the gains are minimal. This information is of considerable value for large scale mapping efforts, for example the NLCD or LCMAP products. Targeting specific months in certain regions instead of using the entire time series could offer significant computational gains with minimal accuracy loss. Another opportunity relates to improved communication of classification limitations. By knowing the incorporated months for each pixel's classification along with the resulting class label using our work as a proof of concept future classification efforts could improve estimations on the associated accuracy of the classification – e.g. were better or worst months used.

While for completeness we included the three tables below, the following discussion is based on the 4-month combination as the specific month benefits are compressed with the inclusion of a higher month total number. More specifically:

- Water: Typically, the easiest classification task, only the arid CR exhibits low accuracy (Max = 92.0 %) and variability between months (Max-Med = 4.3 %) that specific month investigation is warranted. For the arid CR observations from late summer (August and September) are most suitable.
- Impervious: Specific month benefits could be arranged in three groups: 3.4 % for the temperate, no dry CR, ~5% for the three dry CRs and 8.8 % for the cold/no dry CR. In the arid CR, December is showing an unusually strong contribution that is not present in other CRs. The temperate, dry CR also exhibits unique behavior through the importance of the Jan, Feb and Mar months. Late summer/early fall are important in the other CRs.

Table 3
Month relative importance and range of improvement for optimum **4-month** selection.

Class	Climate Zone	Sample#	F1(Med-Max)	F1(Max)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Water	Temperate, no dry season	1000	-0.6%	98.0%	74%	56%	49%	72%	59%	48%	56%	75%	68%	77%	81%	84%
	Arid	1000	-2.3%	94.8%	51%	55%	44%	40%	51%	64%	88%	98%	100%	80%	64%	65%
	Temperate, dry summer	1000	-0.4%	99.5%	43%	54%	52%	71%	82%	68%	79%	84%	96%	63%	52%	56%
	Cold, dry/hot summer	1311	-0.4%	99.4%	57%	42%	61%	49%	72%	67%	79%	76%	81%	79%	73%	63%
	Cold, no dry season	689	-0.9%	96.9%	39%	60%	74%	53%	78%	78%	75%	72%	75%	75%	61%	62%
Impervious	Temperate, no dry season	1000	-1.8%	87.5%	38%	49%	49%	75%	60%	53%	82%	83%	78%	75%	75%	83%
	Arid	1000	-4.0%	72.0%	20%	63%	72%	74%	53%	75%	62%	58%	66%	77%	87%	92%
	Temperate, dry summer	1000	-2.4%	78.4%	32%	51%	84%	84%	64%	56%	74%	88%	86%	60%	63%	58%
	Cold, dry/hot summer	1000	-2.3%	92.6%	22%	28%	49%	64%	62%	79%	85%	87%	83%	86%	83%	70%
	Cold, no dry season	1000	-4.7%	84.2%	17%	45%	48%	46%	93%	98%	82%	81%	85%	75%	60%	69%
Grass/ shrub	Temperate, no dry season	1000	-12.8%	66.4%	16%	34%	51%	55%	54%	75%	84%	79%	80%	88%	92%	92%
	Arid	1000	-3.2%	69.2%	39%	71%	90%	78%	59%	65%	57%	49%	57%	68%	83%	85%
	Temperate, dry summer	1000	-4.8%	78.7%	70%	61%	54%	62%	71%	72%	75%	87%	85%	77%	47%	39%
	Cold, dry/hot summer	1000	-21.5%	82.3%	33%	40%	31%	35%	57%	85%	81%	84%	91%	93%	89%	80%
	Cold, no dry season	1000	-7.3%	81.0%	35%	10%	43%	60%	85%	93%	76%	84%	82%	87%	74%	72%
Forest	Temperate, no dry season	1000	-1.7%	89.6%	34%	71%	59%	80%	43%	77%	88%	89%	61%	66%	63%	69%
	Arid	1000	-4.5%	88.0%	18%	73%	87%	72%	59%	62%	48%	49%	62%	81%	96%	93%
	Temperate, dry summer	1000	-2.2%	86.1%	32%	65%	74%	86%	59%	63%	76%	86%	87%	70%	51%	51%
	Cold, dry/hot summer	1000	-2.4%	92.7%	33%	30%	42%	48%	74%	84%	92%	92%	85%	78%	75%	68%
	Cold, no dry season	1000	-2.9%	89.1%	47%	56%	38%	41%	99%	87%	74%	86%	90%	72%	44%	67%
Bare	Arid	4579	-1.8%	94.7%	6%	42%	72%	87%	76%	81%	75%	68%	74%	70%	68%	82%
Cultivated	Temperate, no dry season	1000	-5.4%	77.5%	29%	31%	31%	54%	67%	79%	83%	80%	74%	90%	91%	92%
	Arid	1000	-4.4%	76.7%	15%	37%	59%	83%	95%	90%	76%	74%	69%	66%	67%	68%
	Temperate, dry summer	1000	-7.5%	62.5%	21%	26%	65%	77%	65%	68%	70%	80%	92%	81%	76%	80%
	Cold, dry/hot summer	1000	-10.7%	89.3%	33%	39%	34%	34%	57%	81%	81%	83%	92%	94%	91%	81%
	Cold, no dry season	1000	-5.3%	88.9%	33%	35%	35%	37%	84%	95%	81%	88%	83%	85%	64%	80%
Wetland	Temperate, no dry season	1273	-2.0%	94.4%	30%	75%	47%	80%	65%	87%	88%	83%	54%	62%	64%	65%
	Arid	1000	-4.5%	89.8%	17%	52%	52%	66%	68%	69%	82%	91%	97%	79%	66%	60%
	Temperate, dry summer	1000	-3.7%	90.7%	22%	51%	78%	94%	75%	65%	84%	93%	83%	62%	56%	36%
	Cold, dry/hot summer	727	-3.7%	88.3%	38%	44%	29%	66%	81%	71%	84%	92%	81%	83%	78%	54%
	Cold, no dry season	1000	-4.5%	85.4%	40%	42%	38%	73%	100%	98%	73%	79%	76%	67%	49%	64%

Table 4
Month relative importance and range of improvement for optimum **6-month** selection.

Class	Climate Zone	Sample#	F1(Med-Max)	F1(Max)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Water	Temperate, no dry season	1000	-0.7%	97.7%	54%	58%	34%	56%	37%	22%	44%	61%	42%	61%	65%	64%
	Arid	1000	-3.3%	93.4%	32%	29%	31%	26%	29%	41%	66%	86%	97%	65%	47%	51%
	Temperate, dry summer	1000	-0.6%	99.5%	17%	32%	38%	57%	71%	55%	64%	65%	91%	41%	31%	38%
	Cold, dry/hot summer	1311	-0.5%	99.3%	34%	24%	54%	36%	53%	60%	60%	60%	65%	63%	52%	38%
	Cold, no dry season	689	-1.5%	96.6%	25%	39%	57%	27%	62%	64%	64%	65%	66%	63%	35%	34%
Impervious	Temperate, no dry season	1000	-2.4%	85.7%	13%	38%	30%	50%	38%	45%	71%	71%	62%	59%	56%	67%
	Arid	1000	-4.9%	62.5%	19%	35%	56%	61%	36%	49%	46%	41%	48%	61%	65%	83%
	Temperate, dry summer	1000	-3.7%	77.1%	46%	53%	79%	60%	34%	33%	58%	70%	65%	37%	38%	27%
	Cold, dry/hot summer	1000	-3.1%	90.7%	6%	14%	42%	58%	49%	58%	66%	73%	67%	67%	65%	37%
	Cold, no dry season	1000	-6.2%	79.8%	12%	29%	35%	22%	75%	90%	72%	68%	66%	51%	34%	45%
Grass/ shrub	Temperate, no dry season	1000	-15.4%	66.0%	3%	17%	28%	32%	40%	60%	68%	65%	59%	69%	79%	78%
	Arid	1000	-4.9%	62.9%	10%	51%	79%	61%	42%	44%	42%	38%	42%	56%	68%	65%
	Temperate, dry summer	1000	-7.0%	76.3%	61%	66%	65%	64%	49%	51%	54%	61%	57%	47%	16%	8%
	Cold, dry/hot summer	1000	-26.0%	78.8%	18%	23%	20%	17%	30%	59%	62%	71%	79%	81%	78%	62%
	Cold, no dry season	1000	-9.6%	76.5%	21%	4%	25%	29%	66%	77%	65%	71%	67%	73%	52%	51%
Forest	Temperate, no dry season	1000	-2.2%	88.5%	11%	65%	53%	58%	46%	63%	67%	76%	44%	44%	36%	38%
	Arid	1000	-8.5%	82.6%	17%	46%	65%	54%	41%	34%	25%	33%	49%	68%	88%	80%
	Temperate, dry summer	1000	-3.0%	84.8%	23%	59%	78%	69%	36%	48%	56%	67%	65%	40%	26%	32%
	Cold, dry/hot summer	1000	-3.3%	91.5%	17%	11%	29%	27%	64%	68%	73%	77%	70%	60%	55%	49%
	Cold, no dry season	1000	-4.2%	86.1%	31%	42%	25%	27%	96%	59%	57%	66%	69%	52%	27%	49%
Bare	Arid	4579	-2.9%	90.5%	9%	29%	58%	70%	55%	66%	55%	37%	48%	50%	52%	72%
Cultivated	Temperate, no dry season	1000	-6.4%	75.8%	7%	17%	22%	35%	56%	70%	65%	65%	49%	68%	73%	74%
	Arid	1000	-7.7%	68.3%	5%	14%	49%	63%	82%	75%	62%	57%	48%	44%	48%	51%
	Temperate, dry summer	1000	-10.0%	59.5%	11%	15%	61%	65%	47%	51%	48%	56%	76%	55%	52%	63%
	Cold, dry/hot summer	1000	-12.8%	89.0%	15%	22%	22%	19%	36%	56%	65%	68%	77%	81%	78%	61%
	Cold, no dry season	1000	-8.7%	86.8%	16%	14%	23%	21%	71%	84%	67%	74%	68%	65%	42%	56%
Wetland	Temperate, no dry season	1273	-3.4%	93.0%	3%	96%	24%	45%	42%	55%	54%	62%	51%	57%	54%	56%
	Arid	1000	-11.9%	84.8%	2%	20%	37%	49%	55%	54%	67%	73%	84%	65%	52%	42%
	Temperate, dry summer	1000	-6.8%	87.5%	6%	31%	68%	80%	51%	47%	69%	79%	66%	43%	39%	20%
	Cold, dry/hot summer	727	-5.5%	85.9%	15%	18%	20%	61%	65%	59%	65%	77%	64%	63%	57%	36%
	Cold, no dry season	1000	-8.7%	82.3%	21%	39%	28%	58%	99%	80%	55%	56%	54%	46%	26%	39%

- Grass/shrub: Consistently the most challenging class also benefits the most from specific month selection. Unfortunately, the least accurate CR (arid with Max = 50.7 %) benefits the least at 6.9 % using Mar and late fall observations. The temperate, dry and cold, no dry CRs see benefits close to 11 %, however with different prevailing months. More distributed with late winter for the former, more concentrated in the summer early fall in the latter. The temperate, no dry CR shows substantial gains of 17.1 % with only imagery from May or later being useful. Lastly, the cold, dry CR shows the most improvement with 30.3 % and highly concentrated month prevalence in the second half of the year and specifically the fall months.
- Forest: Four CRs show improvements in the 3.6 %-5% range with a wide variability in month importance. The most underperforming CR (arid) shows the most improvement with 13.6 % with the unique importance of Oct and Nov and also the highest Dec contribution.
- Bare: With only sufficient samples to study it in the arid CR, December is a critical month offering substantial overall improvements in the 6.1 %.
- Cultivated: Second most challenging class with improvements ranging from 5.9 % to 13.9 % depending on the CR. Mar is influential for the Temperate dry and the arid CRs but not other CRs. Early summer months are important for all CRs except the temperate, dry. Sept is important for the two cold and the temperate, dry CRs.
- Wetland: The high classification accuracy variability between CRs also translates into highly variable benefits for specific month selection. The temperate, no dry season is a notable example of month importance, with Jan and Mar having minimal contribution, while the month in-between, Feb, having a major impact. High importance moves from Mar to Apr to May for the temperate dry to the cold, dry to the cold, no dry, respectively. Late summer months are more pronounced in the arid/dry CRs.

4.3. *Performance variation compared to optimum month combination*

Section 4.1 identified the average benefit of increasing temporal information while section 4.2 identified specific months that were more prevalent in more accurate classifications. In this section, we use the information obtained from our previous results to examine how well it would translate into a practical implementation. From that perspective, a certain month combination should be applicable to all classes for a given CR. Within each CR the importance of specific months may vary from class to class. Subsequently, the "optimal" month combination for a given CR is a balancing act offering the highest average improvement across all classes. The following month selection process was implemented, repeated separately for each CR and each number of months scenario. The average F1 was calculated across all classes for each of the 40 models and for each of the possible month combinations (e.g. 495 combinations for N4). Then the model average F1 was calculated across the 40 models. The month combination with the highest model average F1 was selected.

To quantify the potential benefits of the "optimal" month combination, for each model per class F1 accuracy was compared between the optimal (result of above calculations) and median ranked month combination (result of section 4.1 calculations). Results are reported in Fig. 3.

The improvements offered by the optimal month combination are

Table 5
Month relative importance and range of improvement for optimum **8-month** selection.

Class	Climate Zone	Sample#	F1(Med-Max)	F1(Max)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Water	Temperate, no dry season	1000	-1.2%	97.0%	22%	43%	11%	58%	21%	9%	31%	44%	27%	46%	43%	45%
	Arid	1000	-4.3%	92.0%	22%	16%	17%	9%	9%	16%	48%	69%	82%	47%	31%	33%
	Temperate, dry summer	1000	-1.0%	99.4%	1%	9%	17%	36%	58%	47%	45%	51%	74%	25%	17%	19%
	Cold, dry/hot summer	1311	-0.6%	99.1%	10%	10%	33%	35%	42%	40%	48%	45%	56%	42%	24%	16%
	Cold, no dry season	689	-3.8%	95.8%	5%	9%	21%	22%	59%	47%	64%	55%	50%	33%	13%	21%
Impervious	Temperate, no dry season	1000	-3.4%	82.9%	2%	28%	20%	20%	26%	27%	48%	61%	49%	40%	35%	44%
	Arid	1000	-4.5%	46.2%	20%	16%	42%	36%	16%	22%	30%	26%	35%	49%	37%	72%
	Temperate, dry summer	1000	-5.7%	75.7%	53%	49%	56%	24%	3%	10%	38%	50%	48%	29%	25%	13%
	Cold, dry/hot summer	1000	-4.4%	88.2%	2%	9%	33%	56%	37%	29%	41%	64%	48%	41%	36%	7%
	Cold, no dry season	1000	-8.8%	74.2%	6%	14%	21%	11%	56%	70%	60%	55%	46%	29%	13%	18%
Grass/ shrub	Temperate, no dry season	1000	-17.1%	61.1%	0%	4%	6%	8%	39%	44%	51%	46%	43%	48%	55%	56%
	Arid	1000	-6.9%	50.7%	3%	22%	57%	41%	27%	21%	25%	34%	40%	49%	49%	32%
	Temperate, dry summer	1000	-10.9%	75.3%	40%	53%	55%	43%	24%	35%	39%	44%	39%	23%	3%	2%
	Cold, dry/hot summer	1000	-30.3%	72.7%	5%	8%	7%	4%	16%	35%	46%	57%	63%	61%	62%	36%
	Cold, no dry season	1000	-11.6%	70.1%	11%	3%	17%	8%	47%	60%	56%	58%	49%	45%	23%	24%
Forest	Temperate, no dry season	1000	-3.6%	86.6%	3%	27%	29%	28%	32%	54%	50%	66%	38%	35%	19%	17%
	Arid	1000	-13.6%	74.1%	30%	34%	43%	33%	21%	8%	6%	11%	21%	58%	72%	63%
	Temperate, dry summer	1000	-4.9%	83.3%	12%	47%	67%	38%	13%	24%	30%	47%	54%	26%	17%	23%
	Cold, dry/hot summer	1000	-4.5%	89.2%	4%	1%	12%	25%	45%	50%	39%	51%	57%	38%	37%	41%
	Cold, no dry season	1000	-5.0%	80.7%	32%	46%	21%	20%	52%	28%	47%	44%	38%	15%	9%	49%
Bare	Arid	4579	-6.1%	81.6%	11%	20%	38%	42%	32%	48%	35%	16%	25%	32%	34%	67%
Cultivated	Temperate, no dry season	1000	-5.9%	71.6%	3%	6%	3%	12%	63%	50%	43%	48%	37%	43%	46%	45%
	Arid	1000	-9.9%	56.1%	2%	13%	39%	45%	60%	54%	54%	41%	27%	20%	23%	23%
	Temperate, dry summer	1000	-10.1%	55.2%	10%	14%	58%	28%	22%	32%	35%	39%	54%	30%	28%	50%
	Cold, dry/hot summer	1000	-13.0%	84.8%	0%	2%	5%	5%	44%	49%	50%	51%	56%	53%	51%	35%
	Cold, no dry season	1000	-13.9%	81.2%	2%	0%	1%	9%	59%	68%	47%	54%	50%	43%	31%	37%
Wetland	Temperate, no dry season	1273	-5.7%	90.7%	3%	91%	3%	8%	28%	29%	32%	43%	42%	47%	37%	37%
	Arid	1000	-22.5%	64.2%	1%	3%	23%	29%	34%	29%	48%	59%	63%	50%	41%	20%
	Temperate, dry summer	1000	-11.0%	82.4%	2%	15%	66%	46%	20%	32%	53%	63%	46%	30%	22%	5%
	Cold, dry/hot summer	727	-8.0%	82.2%	2%	0%	10%	70%	48%	48%	50%	52%	42%	35%	28%	14%
	Cold, no dry season	1000	-21.3%	73.1%	9%	18%	17%	48%	82%	54%	52%	40%	35%	21%	10%	15%

tangible on average in the 5 %-10 % range. This is particularly evident in classes with low starting median accuracy, where the room for improvement is high. Of particular note are the improvements in the grass/shrub class in the temperate, no dry and the cold, dry CRs, the wetland in the arid CR, the agriculture in the cold, dry and the cold, no dry CRs. With a few exceptions, these improvements do not come at the cost of other classes, which also improve but to a lesser magnitude. It can be concluded that month selection optimization should be considered as specific temporal patterns could affect the result.

5. Discussion

Results indicate considerable variability for improvement range across classes and CRs. Large climate regions were selected in this first study to identify if there are differences rather than attribute these differences to specific mechanisms. However, at this coarse level we can still offer general explanations, although future follow-up work is needed at a more localized level.

By examining differences across CRs a general vegetation trend, partially driven by climate, can be seen moving from west to east. Feb/Mar months tend to be more important for the Grass/Shrub and Forest classes in the westernmost temperate, dry CR, followed by the arid CR in Mar/Apr. In the eastern U.S. the trend continues from north to south, with late spring having high importance in the cold, no dry northern CR, followed by summer in the cold, dry CR. Grass is also particularly sensitive to late fall months in the temperate, no dry southeast CR, while those months are also influential in the Arid CR for Forest.

Wetlands also show distinct months for improved accuracy. Looking at the four months table the temperate, no dry identified Feb, the cold, dry CR identified Apr and the cold, no Dry May as the optimal months. For the first southeastern CR looking at the greening median week in Fig. 4 it seems that the optimal month precedes by a few weeks the greening of vegetation. For the two cold CRs, the same pattern is generally observed but also amplified further it follows snow melting effects (see Fig. 5). This is expected as the strongest wetland signature would be when the water level is high but before surface vegetation cover alters the water's spectral signal.

By far the largest crop in the United States is corn followed by soybeans, which are mostly produced in the cold, dry CR (Fig. 6). The effects of the crop calendar are clearly present in the month influence. In particular, while the mid-season is influential it is the harvest period that is more highlighted. Another finding is that the planting season does not seem to play an important role in the classification process. The sudden spectral change during harvest, also considering the behavior of other classes, suggests that early fall is the most critical time period. In the cold, no dry and arid CRs, the May/June planting season tends to offer

Fig. 3. Variation of F1 class accuracy gain within 40 models for different CRs and different total month number combinations (note that Y-axis limits vary for presentation).

better separability of the crop class and varies across the other CRs.

While the impervious class does not change its spectral signature over time, the ability to separate it is dependent on spectral confusion with other classes. A particularly known confusion is between impervious and soil. While our dataset did not have enough samples to study bare lands in multiple ecoregions, we can contrast impervious with the cultivated class. In general, summer months tend to assist the most in impervious classification, which could be attributed to the presence of crop cover of the cultivated areas and potentially the greening of the vegetated classes. A notable exception is Dec, which works well for the Impervious and the Bare class for the arid CR. Finally, the water class is easier to distinguish in Sep for the dry CRs (arid and two CRs with dry summers) but improvements are minor. As for Forest class, some months tend to prevail over others (e.g. Aug for the temperate, no dry and Mar for the temperate, dry CR), however, no conclusive explanations could be identified. Possibly smaller regions, driven by forest species distribution, could provide better insights.

The impact of our decision to include a single observation per month should be acknowledged. The rationale is to avoid bias in the obtained results, which look specifically at each month's importance. In some cases, multiple month observations may further benefit classification accuracy further (e.g. detecting agriculture during the growing season). The goal of this work is to identify prevailing months to support future investigations at a finer temporal resolution (e.g. weekly) conducted in a

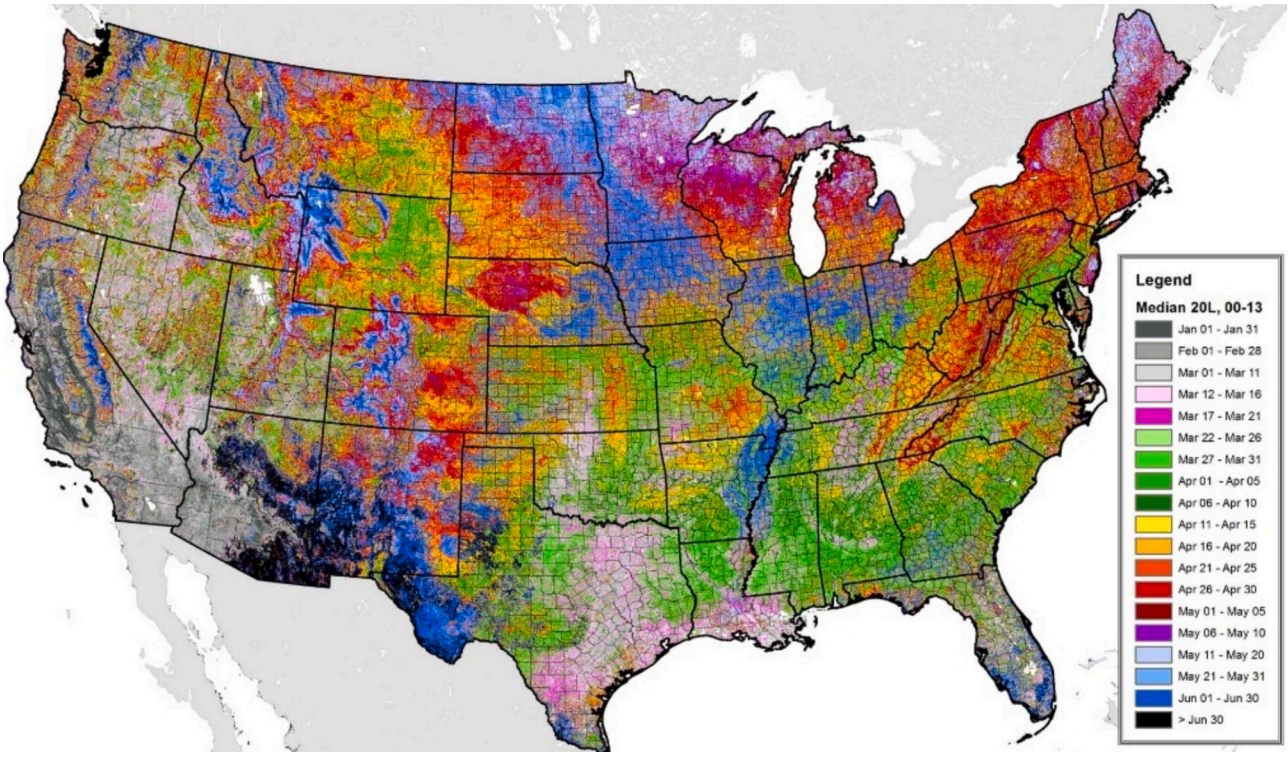

Fig. 4. Median start of greenup date ().
adapted from https://forwarn.forestthreats.org/highlights/917

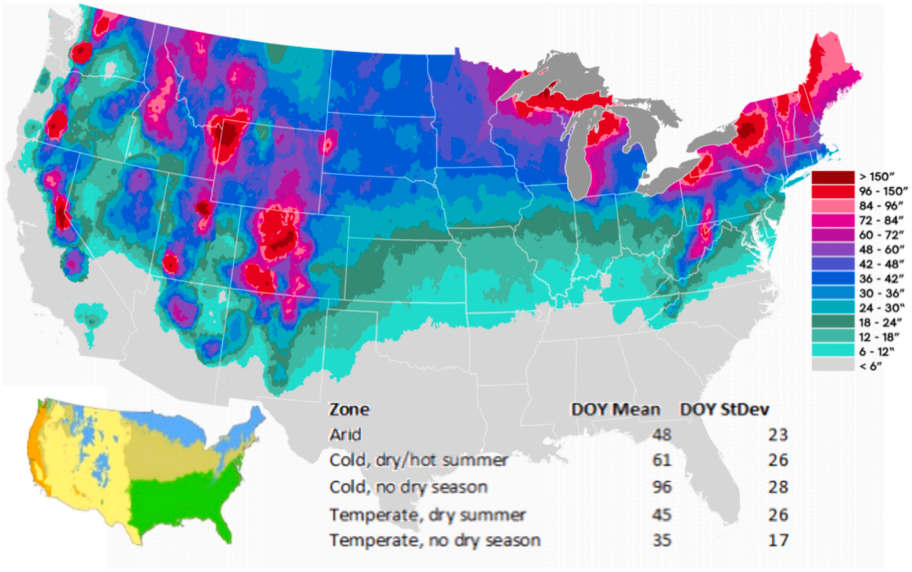

Fig. 5. Average annual snowfall and snowmelt day of the year statistics ().
adapted from https://atlas.niu.edu/klot/snowclimatology/USSnowContours.png

more targeted manner. This is possible as the Landsat archive within the conterminous United States the average annual number of non-cloudy observations varies between 14.23 and 30.07 (Egorov et al., 2019). Additionally, while the sampling design ensured the spatially independence between training and testing samples, the presence of both training and testing samples from each of the 84 blocks is boosting the accuracy of the obtained results, however this is not affecting out conclusions as that bias is present across all months.

Finally, we should clarify that our goal here is to investigate the effect of specific months in the classification, not to produce an optimized classifier. We have selected a reasonable architecture that was well-tested in our prior work to amplify those monthly differences. Our hope is that those targeting mapping activities will embed our findings in their own architectures to optimize month selection based on class and region of interest, which in turn could improve accuracy and computational efficiency.

6. Conclusion

In this research we studied the effect of individual month

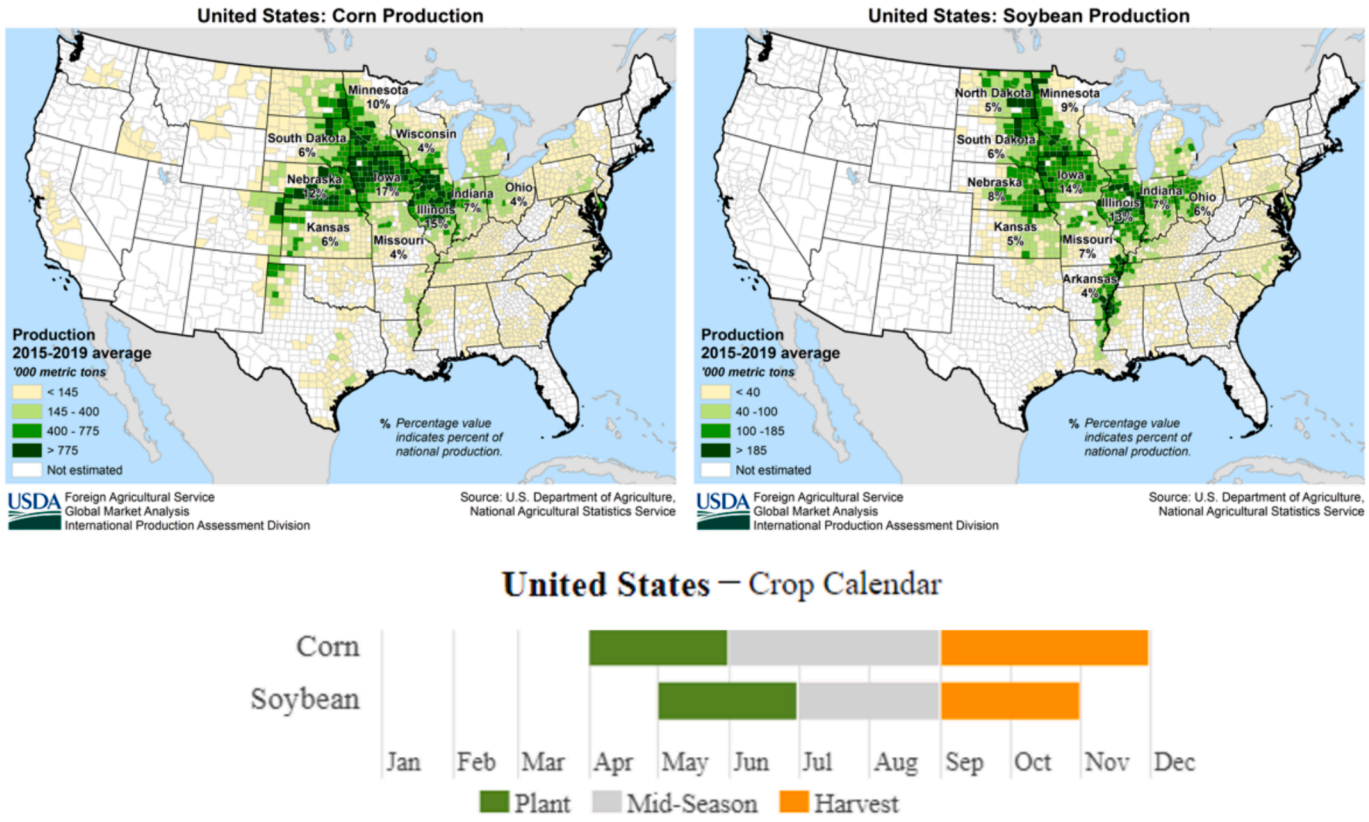

Fig. 6. Corn and soybean map and harvesting calendar ().
adapted from https://ipad.fas.usda.gov/countrysummary/Default.aspx?id=US

observations in Landsat time series for land cover/use classification across five general climatic regions in the conterminous United States. The study showed potentially substantial gains in classification accuracy over different classes and climatic regions by varying the total observation months and the months selected. Benefits can be as substantial as 50 % as shown in Fig. 2 for the grass/shrub class in cold, dry climatic region. Further validation of the month selection importance comes from an example implementation scenario (Fig. 3) where F1 improvements can be as high as 10 %.

Our analysis used one observation per month. New Landsat missions and fusion with harmonized Sentinel observations can reduce revisiting time to a couple of days, thus supporting in the future higher temporal resolution, where for example optimal weeks instead of months could be identified. Furthermore, areas of distinct spectral behavior could be identified (e.g. local climate, specific crops or vegetation) and studied separately to offer an even more targeted approach on temporal information selection. Lastly, while we touched upon explanation on the observed temporal patterns further studies are required to offer more conclusive results.

CRediT authorship contribution statement

Giorgos Mountrakis: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Shahriar S. Heydari:** Investigation, Methodology, Validation, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the NASA's Land Cover Land Use Change Program (grant # NNX15AD42G) and Ecological Forecasting Program (grant # 20-ECOF20-0019) and NASA High-End Computing facility at NASA Ames Research Center, CA. We are grateful for excellent technical and administrative NASA support, in particular the assistance from Blaise Hartman and Samson Cheung among others. We also appreciate the efforts from Kristi Sayler and Mark Drummond (USGS) for providing the Trends data. We thank our reference data generation team of current and former SUNY ESF students: Atef Alla Eddin Amriche, Harrison Goldspiel, Madusha Sammani, Megan Kate Medwid, Rabia Munsaf Khan, and Zhixin Wang.

Appendix A:. Land cover class definitions and class-specific considerations

Water	*Ponds, lakes, rivers, oceans that are persistently filled by water.* For lakes, ponds, and oceans, only the pixels demonstrating a persistent water presence were included. For streams/rivers, a pixel was assigned to the water class when deemed necessary to preserve the spatial continuity (i.e., avoid river breaks), even if the water occupied less pixel area than other classes). Water presence was required to be persistent through time, therefore seasonal water presence (e.g., seasonal streams, limited flooding) did not qualify a pixel for the water class. Presence of algae on the water surface did not disqualify a pixel from water class assignment.

(*continued on next page*)

(*continued*)

Impervious	*Built-up area including houses, factories, barns, parking lots, roads (paved or dirt), railroads.* The developed class was prioritized over all other classes. When 20 % or higher of the pixel area was deemed developed the pixel was assigned to the developed class. Road pixels, even when occupying less than 20 % of pixel area were still assigned as developed when it was necessary to preserve their spatial continuity. Dirt roads were considered falling in the developed class, but irregular/temporary tracks and trails/footpaths were not.
Grass/ Shrub	*Low vegetation that is not cultivated, including natural patches, pasture and grazing land, and man-made patches.* Man-made patches including yard lawns, city parks, golf courses, and soccer fields were assigned to this class. Pasture and Grazing lands that are not intensely cultivated were also included here.
Forest	*Tall vegetation (taller than typical grass/shrubs) that is not intensely cultivated.* All tree types, including forest plantations, were assigned as forest. Tree orchards were not assigned as forest.
Bare	*Soil, rocks, mining land, or land with very limited vegetation.* If vegetation was identified as majority for a pixel area even for a short time period it was labeled as grassland and not barren. Sand dunes and dry sandy areas were assigned in the barren land. This class includes Barren, mechanically disturbed, and nonmechanically disturbed classes of the original 11-level Anderson classification scheme.
Cultivated	*Cultivated areas demonstrating distinct agricultural parcel shapes and tilling lines, including orchards and vineyards.* For designation as cropland, these characteristics were sought for at least 20 % of the pixel area: 1) row pattern of tilling/cultivation, 2) temporal high contrast color transition from green to yellow, 3) regular rectangular shape with clear farm edges.
Wetland	*A typically vegetated area that is periodically saturated or covered with water.* For designation as a wetland (that may have low plants or high trees mixed with water) water should be present mixed with vegetation most of the time. There should be no clear water boundary as the boundary may change every year (unlike a lake or a pond). For woody wetlands, where water was difficult to identify under thick canopy, wetlands were assigned when high vegetation turnover was present. Examination of these challenging pixels during winter months was a critical decision component. Muddy, vegetation-free areas in lake borderlines or seashores were assigned as wetland, not barren land.

Appendix B:. Model characteristics and class accuracy

Classifiers used in this study are taken from a previous research as explained in Mountrakis and Heydari (2023);a summary of that is presented here for completeness. The general schematic of our approach is shown in Figure B-1. To assess the value of temporal and/or spatial information three network architectures were tested. The first deep learner only examines individual pixel temporal information using a Long Short-Term Memory network (T-LSTM). The second network (ST-LSTM) adds expert-selected spatial neighboring features to the T-LSTM. The third network (C-LSTM) adds to the ST-LSTM automatically generated features using a convolutional neural network.

Climate Zone	Class	Sample#	F1(Med-Max)	F1 (Max)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperate, no dry season	Water	1000	-1.2%	97.0%	22%	43%	11%	58%	21%	9%	31%	44%	27%	46%	43%	45%
	Impervious	1000	-3.4%	82.9%	2%	28%	20%	20%	26%	27%	48%	61%	49%	40%	35%	44%
	Grass/shrub	1000	-17.1%	61.1%	0%	4%	6%	8%	39%	44%	51%	46%	43%	48%	55%	56%
	Forest	1000	-3.6%	86.6%	3%	27%	29%	28%	32%	54%	50%	66%	38%	35%	19%	17%
	Bare	Insufficient samples														
	Cultivated	1000	-5.9%	71.6%	3%	6%	3%	12%	63%	50%	43%	48%	37%	43%	46%	45%
	Wetland	1273	-5.7%	90.7%	3%	91%	3%	8%	28%	29%	32%	43%	42%	47%	37%	37%
Arid	Water	1000	-4.3%	92.0%	22%	16%	17%	9%	9%	16%	48%	69%	82%	47%	31%	33%
	Impervious	1000	-4.5%	46.2%	20%	16%	42%	36%	16%	22%	30%	26%	35%	49%	37%	72%
	Grass/shrub	1000	-6.9%	50.7%	3%	22%	57%	41%	27%	21%	25%	34%	40%	49%	49%	32%
	Forest	1000	-13.6%	74.1%	30%	34%	43%	33%	21%	8%	6%	11%	21%	58%	72%	63%
	Bare	4579	-6.1%	81.6%	11%	20%	38%	42%	32%	48%	35%	16%	25%	32%	34%	67%
	Cultivated	1000	-9.9%	56.1%	2%	13%	39%	45%	60%	54%	54%	41%	27%	20%	23%	23%
	Wetland	1000	-22.5%	64.2%	1%	3%	23%	29%	34%	29%	48%	59%	63%	50%	41%	20%
Temperate, dry summer	Water	1000	-1.0%	99.4%	1%	9%	17%	36%	58%	47%	45%	51%	74%	25%	17%	19%
	Impervious	1000	-5.7%	75.7%	53%	49%	56%	24%	3%	10%	38%	50%	48%	29%	25%	13%
	Grass/shrub	1000	-10.9%	75.3%	40%	53%	55%	43%	24%	35%	39%	44%	39%	23%	3%	2%
	Forest	1000	-4.9%	83.3%	12%	47%	67%	38%	13%	24%	30%	47%	54%	26%	17%	23%
	Bare	Insufficient samples														
	Cultivated	1000	-10.1%	55.2%	10%	14%	58%	28%	22%	32%	35%	39%	54%	30%	28%	50%
	Wetland	1000	-11.0%	82.4%	2%	15%	66%	46%	20%	32%	53%	63%	46%	30%	22%	5%
Cold, dry/hot summer	Water	1311	-0.6%	99.1%	10%	10%	33%	35%	42%	40%	48%	45%	56%	42%	24%	16%
	Impervious	1000	-4.4%	88.2%	2%	9%	33%	56%	37%	29%	41%	64%	48%	41%	36%	7%
	Grass/shrub	1000	-30.3%	72.7%	5%	8%	7%	4%	16%	35%	46%	57%	63%	61%	62%	36%
	Forest	1000	-4.5%	89.2%	4%	1%	12%	25%	45%	50%	39%	51%	57%	38%	37%	41%
	Bare	Insufficient samples														
	Cultivated	1000	-13.0%	84.8%	0%	2%	5%	5%	44%	49%	50%	51%	56%	53%	51%	35%
	Wetland	727	-8.0%	82.2%	2%	0%	10%	70%	48%	48%	50%	52%	42%	35%	28%	14%
Cold, no dry season	Water	689	-3.8%	95.8%	5%	9%	21%	22%	59%	47%	64%	55%	50%	33%	13%	21%
	Impervious	1000	-8.8%	74.2%	6%	14%	21%	11%	56%	70%	60%	55%	46%	29%	13%	18%
	Grass/shrub	1000	-11.6%	70.1%	11%	3%	17%	8%	47%	60%	56%	58%	49%	45%	23%	24%
	Forest	1000	-5.0%	80.7%	32%	46%	21%	20%	52%	28%	47%	44%	38%	15%	9%	49%
	Bare	Insufficient samples														
	Cultivated	1000	-13.9%	81.2%	2%	0%	1%	9%	59%	68%	47%	54%	50%	43%	31%	37%
	Wetland	1000	-21.3%	73.1%	9%	18%	17%	48%	82%	54%	52%	40%	35%	21%	10%	15%
All Regions	Water	5000	-1.5%	96.2%	14%	20%	16%	25%	20%	20%	50%	68%	80%	43%	22%	23%
	Impervious	5000	-3.2%	68.4%	8%	13%	40%	18%	35%	27%	58%	52%	52%	36%	27%	35%
	Grass/shrub	5000	-8.2%	59.2%	2%	7%	12%	7%	38%	42%	48%	55%	60%	55%	49%	27%
	Forest	5000	-2.9%	80.2%	6%	32%	38%	37%	33%	23%	22%	37%	56%	32%	32%	52%
	Bare	5000	-6.1%	77.5%	9%	21%	31%	36%	30%	65%	39%	19%	28%	30%	32%	60%
	Cultivated	5000	-7.1%	65.3%	0%	1%	7%	25%	68%	62%	48%	51%	46%	36%	29%	27%
	Wetland	5000	-7.8%	76.3%	0%	11%	26%	56%	55%	34%	53%	56%	46%	34%	21%	8%

Fig. B1. Classifiers system architecture. T-LSTM shows the building blocks of the basic model. The ST-LSTM model adds the ST-LSTM supplement to T-LSTM, and the C-LSTM model adds both ST-LSTM and C-LSMT supplements to the model (figure from (Mountrakis and Heydari, 2023)).

The base data in all models were sequences of Landsat data and static topographic data. Each annual sequence contained several temporal records. Each record represented a specific observation day and sensor. For the T-LSTM the annual Landsat sequence was fed to a recurrent network. Static topographic data were in the form of 1-D vector and for each pixel, there was one vector of static features corresponding to one annual sequence of Landsat-based features. The static features were processed by a standard multilayer neural network. Output features from the recurrent and the standard neural networks were concatenated and fed to a second multilayer network for further processing with the latter producing the assigned land cover to the input pixel.

In the case of the ST-LSTM the only difference was that the recurrent network was also fed with expert-selected patch statistics from neighboring pixels. For the C-LSTM, in addition to the expert-selected spatial features, features selected automatically from a convolutional neural network were also included as inputs for the recurrent network.

It is shown in the reference paper that C-LSTM model provides the best performance, therefore we focused on that model. The classifiers in the prior research were trained using a limited grid search approach to tune the model parameters, and the last simulation runs brought many models all having similar performance (i.e. very little difference in overall accuracy). As explained in the main text of the paper, we chose 40 of the last C-LSTM models to conduct the analysis in this paper as detailed in table B-1. The table shows the architecture, dropout ratio and F1 accuracy of the 40 models used in the analysis. These models have the same layer structure, composed of one convolutional (CNN), one recurrent (LSTM), and two multilayer neural networks (MLPs) cascaded together. For each section, the number of neurons (or filters for CNN) in each layer is mentioned as a comma-separated entry in the corresponding column. Being a complex network with millions of parameters, the dropout method was used to reduce network overfitting and therefore this setting is also included below. Some other minor differences in network learning parameters between different models exist but are not reported here. Each model's average and per-class F1 performance is also given.

Table B1
Structure and performance of 40 candidate models taken from a prior research and used in the research of this paper.

Modelcode	CNNlayers	LSTMlayers	1stMLP	2ndMLP	Dropoutratio*	averageF1	F1water	F1impervious	F1grass	F1forest	F1bare	F1cultivated	F1wetland
G9_12_0	96, 144	340,340,340	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.975	0.969	0.994	0.982	0.983
G9_12_1	96, 168	340,340,340	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.993	0.981	0.974	0.969	0.994	0.982	0.983
G9_13_0	96, 192	360,320,320	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.969	0.993	0.982	0.983
G9_17_0	96, 192	350,320,320	32,32	128,128,128,128	0.3,0.25,0,0	0.982	0.992	0.981	0.974	0.969	0.993	0.982	0.983
G9_17_2	96, 192	400,320,320	32,32	128,128,128,128	0.3,0.25,0,0	0.982	0.993	0.981	0.974	0.969	0.994	0.982	0.983
G9_18_0	96, 144	400,320,320	32,32	128,128,128,128	0.3,0.25,0,0	0.982	0.993	0.981	0.974	0.968	0.994	0.982	0.983
G9_18_1	96, 192	350,320,320	32,32	128,128,128,128	0.3,0.25,0,0	0.982	0.992	0.981	0.975	0.969	0.994	0.982	0.983
G9_19_0	96, 192	400,320,320	32,32	128,128,128,128	0.3,0.25,0,0	0.982	0.992	0.981	0.974	0.969	0.993	0.982	0.983
G9_19_1	96, 144	400,320,320	32,32	128,128,128,128	0.3,0.25,0,0	0.982	0.992	0.981	0.974	0.969	0.993	0.982	0.983
G9_19_2	96, 192	350,320,320	32,32	128,128,128,128	0.3,0.25,0,0	0.982	0.992	0.981	0.973	0.969	0.993	0.982	0.982
G9_20_1	96, 192	400,320,320	32,32	128,128,128,128	0.3,0.25,0,0	0.982	0.992	0.981	0.975	0.969	0.993	0.982	0.983
G9_21_0	96, 192	360,320,320	32,32	256,256	0.3,0.25,0,0	0.981	0.992	0.981	0.973	0.968	0.993	0.981	0.982
G9_21_1	96, 192	340,320,320	32,32	256,256	0.3,0.25,0,0	0.982	0.992	0.981	0.974	0.969	0.993	0.982	0.982
G9_21_2	96, 192	360,320,320	32,32	256,256	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.969	0.993	0.982	0.983
G9_5_0	96, 192	360,320,320	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.969	0.992	0.982	0.983
G9_6_0	128, 96	360,320,320	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.975	0.969	0.993	0.982	0.983
G9_6_1	96, 192	340,340,320	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.973	0.968	0.993	0.982	0.983
G9_6_2	96, 144	340,340,320	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.968	0.993	0.982	0.982
G9_7_0	96, 168	340,340,320	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.969	0.993	0.982	0.983
G9_7_1	128, 96	340,340,320	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.993	0.981	0.974	0.969	0.994	0.982	0.982
G9_7_2	96, 192	340,340,340	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.969	0.993	0.982	0.982
G9_8_0	96, 144	340,340,340	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.969	0.994	0.982	0.983
G9_8_1	96, 168	340,340,340	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.969	0.993	0.982	0.983
G9_8_2	128, 96	340,340,340	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.973	0.969	0.993	0.982	0.983
G9_9_1	96, 144	360,320,320	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.973	0.969	0.993	0.982	0.982
GD_11_1	96, 128	340,340,340	32,32	128,128,128,128	0.3,0.3,0.1,0.1	0.982	0.993	0.981	0.974	0.968	0.993	0.982	0.982
GD_12_1	128, 96	360,320,320	32,32	128,128,128,128	0.3,0.3,0.1,0.1	0.982	0.992	0.981	0.974	0.968	0.993	0.982	0.982
GD_12_2	96, 128	360,320,320	32,32	128,128,128,128	0.3,0.3,0.1,0.1	0.981	0.992	0.981	0.974	0.968	0.993	0.981	0.982
GD_14_1	128, 96	340,340,340	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.993	0.981	0.974	0.969	0.994	0.982	0.983
GD_17_1	96, 128	320,300,280,280	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.975	0.969	0.993	0.982	0.983
GD_18_2	96, 128	360,340,320	32,32	128,128,128,128	0.3,0.3,0.1,0.1	0.981	0.992	0.98	0.973	0.968	0.992	0.981	0.982
GD_21_0	128, 96	340,340,340	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.968	0.993	0.982	0.983
GD_22_0	96, 128	360,340,320	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.968	0.993	0.981	0.983
GD_22_2	96, 128	360,320,320	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.975	0.968	0.993	0.982	0.983
GD_23_0	128, 96	280,280,280,280	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.968	0.993	0.982	0.982
GD_23_2	128, 96	320,300,280,280	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.969	0.994	0.982	0.982
GD_7_1	96, 128	320,300,280,280	32,32	128,128,128,128	0.3,0.3,0.1,0.1	0.982	0.992	0.981	0.974	0.969	0.994	0.982	0.983
GD_7_2	128, 96	340,340,340	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.969	0.993	0.982	0.983
GD_8_0	96, 128	340,340,340	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.992	0.981	0.974	0.969	0.993	0.982	0.983
GD_8_1	128, 96	360,340,320	32,32	128,128,128,128	0.3,0.25,0.05,0.05	0.982	0.993	0.981	0.975	0.969	0.993	0.982	0.984

* The numbers in this column refer to the dropout ratios applied to all layers of each of four sections. The first number applies to CNN layers, second one to LSTM layers, etc.

Appendix C:. Month selection map tables organized by climatic regions

Month selection (heatmap) tables presented in the main text are sorted first by class and then by climate region. Here we reorganize the rows first by climatic region and then by class to allow easier comparisons within CRs.

Table C1
Month relative importance and range of improvement for optimum 4-month selection.

Climate Zone	Class	Sample#	F1(Med-Max)	F1 (Max)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperate, no dry season	Water	1000	-1.2%	97.0%	22%	43%	11%	58%	21%	9%	31%	44%	27%	46%	43%	45%
	Impervious	1000	-3.4%	82.9%	2%	28%	20%	20%	26%	27%	48%	61%	49%	40%	35%	44%
	Grass/shrub	1000	-17.1%	61.1%	0%	4%	6%	8%	39%	44%	51%	46%	43%	48%	55%	56%
	Forest	1000	-3.6%	86.6%	3%	27%	29%	28%	32%	54%	50%	66%	38%	35%	19%	17%
	Bare	Insufficient samples														
	Cultivated	1000	-5.9%	71.6%	3%	6%	3%	12%	63%	50%	43%	48%	37%	43%	46%	45%
	Wetland	1273	-5.7%	90.7%	3%	91%	3%	8%	28%	29%	32%	43%	42%	47%	37%	37%
Arid	Water	1000	-4.3%	92.0%	22%	16%	17%	9%	9%	16%	48%	69%	82%	47%	31%	33%
	Impervious	1000	-4.5%	46.2%	20%	16%	42%	36%	16%	22%	30%	26%	35%	49%	37%	72%
	Grass/shrub	1000	-6.9%	50.7%	3%	22%	57%	41%	27%	21%	25%	34%	40%	49%	49%	32%
	Forest	1000	-13.6%	74.1%	30%	34%	43%	33%	21%	8%	6%	11%	21%	58%	72%	63%
	Bare	4579	-6.1%	81.6%	11%	20%	38%	42%	32%	48%	35%	16%	25%	32%	34%	67%
	Cultivated	1000	-9.9%	56.1%	2%	13%	39%	45%	60%	54%	54%	41%	27%	20%	23%	23%
	Wetland	1000	-22.5%	64.2%	1%	3%	23%	29%	34%	29%	48%	59%	63%	50%	41%	20%
Temperate, dry summer	Water	1000	-1.0%	99.4%	1%	9%	17%	36%	58%	47%	45%	51%	74%	25%	17%	19%
	Impervious	1000	-5.7%	75.7%	53%	49%	56%	24%	3%	10%	38%	50%	48%	29%	25%	13%
	Grass/shrub	1000	-10.9%	75.3%	40%	53%	55%	43%	24%	35%	39%	44%	39%	23%	3%	2%
	Forest	1000	-4.9%	83.3%	12%	47%	67%	38%	13%	24%	30%	47%	54%	26%	17%	23%
	Bare	Insufficient samples														
	Cultivated	1000	-10.1%	55.2%	10%	14%	58%	28%	22%	32%	35%	39%	54%	30%	28%	50%
	Wetland	1000	-11.0%	82.4%	2%	15%	66%	46%	20%	32%	53%	63%	46%	30%	22%	5%
Cold, dry/hot summer	Water	1311	-0.6%	99.1%	10%	10%	33%	35%	42%	40%	48%	45%	56%	42%	24%	16%
	Impervious	1000	-4.4%	88.2%	2%	9%	33%	56%	37%	29%	41%	64%	48%	41%	36%	7%
	Grass/shrub	1000	-30.3%	72.7%	5%	8%	7%	4%	16%	35%	46%	57%	63%	61%	62%	36%
	Forest	1000	-4.5%	89.2%	4%	1%	12%	25%	45%	50%	39%	51%	57%	38%	37%	41%
	Bare	Insufficient samples														
	Cultivated	1000	-13.0%	84.8%	0%	2%	5%	5%	44%	49%	50%	51%	56%	53%	51%	35%
	Wetland	727	-8.0%	82.2%	2%	0%	10%	70%	48%	48%	50%	52%	42%	35%	28%	14%
Cold, no dry season	Water	689	-3.8%	95.8%	5%	9%	21%	22%	59%	47%	64%	55%	50%	33%	13%	21%
	Impervious	1000	-8.8%	74.2%	6%	14%	21%	11%	56%	70%	60%	55%	46%	29%	13%	18%
	Grass/shrub	1000	-11.6%	70.1%	11%	3%	17%	8%	47%	60%	56%	58%	49%	45%	23%	24%
	Forest	1000	-5.0%	80.7%	32%	46%	21%	20%	52%	28%	47%	44%	38%	15%	9%	49%
	Bare	Insufficient samples														
	Cultivated	1000	-13.9%	81.2%	2%	0%	1%	9%	59%	68%	47%	54%	50%	43%	31%	37%
	Wetland	1000	-21.3%	73.1%	9%	18%	17%	48%	82%	54%	52%	40%	35%	21%	10%	15%
All Regions	Water	5000	-1.5%	96.2%	14%	20%	16%	25%	20%	20%	50%	68%	80%	43%	22%	23%
	Impervious	5000	-3.2%	68.4%	8%	13%	40%	18%	35%	27%	58%	52%	52%	36%	27%	35%
	Grass/shrub	5000	-8.2%	59.2%	2%	7%	12%	7%	38%	42%	48%	55%	60%	55%	49%	27%
	Forest	5000	-2.9%	80.2%	6%	32%	38%	37%	33%	23%	22%	37%	56%	32%	32%	52%
	Bare	5000	-6.1%	77.5%	9%	21%	31%	36%	30%	65%	39%	19%	28%	30%	32%	60%
	Cultivated	5000	-7.1%	65.3%	0%	1%	7%	25%	68%	62%	48%	51%	46%	36%	29%	27%
	Wetland	5000	-7.8%	76.3%	0%	11%	26%	56%	55%	34%	53%	56%	46%	34%	21%	8%

Table C2
Month relative importance and range of improvement for optimum 6-month selection.

Climate Zone	Class	Sample#	F1(Med-Max)	F1 (Max)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperate, no dry season	Water	1000	-0.7%	97.7%	54%	58%	34%	56%	37%	22%	44%	61%	42%	61%	65%	64%
	Impervious	1000	-2.4%	85.7%	13%	38%	30%	50%	38%	45%	71%	71%	62%	59%	56%	67%
	Grass/shrub	1000	-15.4%	66.0%	3%	17%	28%	32%	40%	60%	68%	65%	59%	69%	79%	78%
	Forest	1000	-2.2%	88.5%	11%	65%	53%	58%	46%	63%	67%	76%	44%	44%	36%	38%
	Bare	Insufficient samples														
	Cultivated	1000	-6.4%	75.8%	7%	17%	22%	35%	56%	70%	65%	65%	49%	68%	73%	74%
	Wetland	1273	-3.4%	93.0%	3%	96%	24%	45%	42%	55%	54%	62%	51%	57%	54%	56%
Arid	Water	1000	-3.3%	93.4%	32%	29%	31%	26%	29%	41%	66%	86%	97%	65%	47%	51%
	Impervious	1000	-4.9%	62.5%	19%	35%	56%	61%	36%	49%	46%	41%	48%	61%	65%	83%
	Grass/shrub	1000	-4.9%	62.9%	10%	51%	79%	61%	42%	44%	42%	38%	42%	56%	68%	65%
	Forest	1000	-8.5%	82.6%	17%	46%	65%	54%	41%	34%	25%	33%	49%	68%	88%	80%
	Bare	4579	-2.9%	90.5%	9%	29%	58%	70%	55%	66%	55%	37%	48%	50%	52%	72%
	Cultivated	1000	-7.7%	68.3%	5%	14%	49%	63%	82%	75%	62%	57%	48%	44%	48%	51%
	Wetland	1000	-11.9%	84.8%	2%	20%	37%	49%	55%	54%	67%	73%	84%	65%	52%	42%
Temperate, dry summer	Water	1000	-0.6%	99.5%	17%	32%	38%	57%	71%	55%	64%	65%	91%	41%	31%	38%
	Impervious	1000	-3.7%	77.1%	46%	53%	79%	60%	34%	33%	58%	70%	65%	37%	38%	27%
	Grass/shrub	1000	-7.0%	76.3%	61%	66%	65%	64%	49%	51%	54%	61%	57%	47%	16%	8%
	Forest	1000	-3.0%	84.8%	23%	59%	78%	69%	36%	48%	56%	67%	65%	40%	26%	32%
	Bare	Insufficient samples														
	Cultivated	1000	-10.0%	59.5%	11%	15%	61%	65%	47%	51%	48%	56%	76%	55%	52%	63%
	Wetland	1000	-6.8%	87.5%	6%	31%	68%	80%	51%	47%	69%	79%	66%	43%	39%	20%
Cold, dry/hot summer	Water	1311	-0.5%	99.3%	34%	24%	54%	36%	53%	60%	60%	60%	65%	63%	52%	38%
	Impervious	1000	-3.1%	90.7%	6%	14%	42%	58%	49%	58%	66%	73%	67%	67%	65%	37%
	Grass/shrub	1000	-26.0%	78.8%	18%	23%	20%	17%	30%	59%	62%	71%	79%	81%	78%	62%
	Forest	1000	-3.3%	91.5%	17%	11%	29%	27%	64%	68%	73%	77%	70%	60%	55%	49%
	Bare	Insufficient samples														
	Cultivated	1000	-12.8%	89.0%	15%	22%	22%	19%	36%	56%	65%	68%	77%	81%	78%	61%
	Wetland	727	-5.5%	85.9%	15%	18%	20%	61%	65%	59%	65%	77%	64%	63%	57%	36%
Cold, no dry season	Water	689	-1.5%	96.6%	25%	39%	57%	27%	62%	64%	64%	65%	66%	63%	35%	34%
	Impervious	1000	-6.2%	79.8%	12%	29%	35%	22%	75%	90%	72%	68%	66%	51%	34%	45%
	Grass/shrub	1000	-9.6%	76.5%	21%	4%	25%	29%	66%	77%	65%	71%	67%	73%	52%	51%
	Forest	1000	-4.2%	86.1%	31%	42%	25%	27%	96%	59%	57%	66%	69%	52%	27%	49%
	Bare	Insufficient samples														
	Cultivated	1000	-8.7%	86.8%	16%	14%	23%	21%	71%	84%	67%	74%	68%	65%	42%	56%
	Wetland	1000	-8.7%	82.3%	21%	39%	28%	58%	99%	80%	55%	56%	54%	46%	26%	39%
All Regions	Water	5000	-0.8%	96.9%	46%	41%	26%	33%	31%	36%	63%	81%	94%	65%	41%	43%
	Impervious	5000	-2.3%	75.6%	1%	18%	54%	53%	56%	59%	69%	64%	65%	52%	48%	61%
	Grass/shrub	5000	-6.9%	66.5%	4%	23%	32%	25%	52%	61%	61%	67%	72%	78%	72%	53%
	Forest	5000	-2.2%	85.0%	2%	48%	51%	58%	62%	38%	41%	52%	65%	63%	56%	63%
	Bare	5000	-3.2%	86.6%	4%	23%	50%	60%	52%	80%	57%	42%	54%	52%	55%	71%
	Cultivated	5000	-6.0%	73.0%	0%	4%	26%	46%	69%	73%	63%	69%	68%	65%	61%	56%
	Wetland	5000	-4.4%	85.2%	2%	37%	34%	65%	80%	54%	63%	72%	67%	54%	41%	31%

Table C3
Month relative importance and range of improvement for optimum 8-month selection.

Climate Zone	Class	Sample#	F1(Med-Max)	F1 (Max)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperate, no dry season	Water	1000	-0.6%	98.0%	74%	56%	49%	72%	59%	48%	56%	75%	68%	77%	81%	84%
	Impervious	1000	-1.8%	87.5%	38%	49%	49%	75%	60%	53%	82%	83%	78%	75%	75%	83%
	Grass/shrub	1000	-12.8%	66.4%	16%	34%	51%	55%	54%	75%	84%	79%	80%	88%	92%	92%
	Forest	1000	-1.7%	89.6%	34%	71%	59%	80%	43%	77%	88%	89%	61%	66%	63%	69%
	Bare	Insufficient samples														
	Cultivated	1000	-5.4%	77.5%	29%	31%	31%	54%	67%	79%	83%	80%	74%	90%	91%	92%
	Wetland	1273	-2.0%	94.4%	30%	75%	47%	80%	65%	87%	88%	83%	54%	62%	64%	65%
Arid	Water	1000	-2.3%	94.8%	51%	55%	44%	40%	51%	64%	88%	98%	100%	80%	64%	65%
	Impervious	1000	-4.0%	72.0%	20%	63%	72%	74%	53%	75%	62%	58%	66%	77%	87%	92%
	Grass/shrub	1000	-3.2%	69.2%	39%	71%	90%	78%	59%	65%	57%	49%	57%	68%	83%	85%
	Forest	1000	-4.5%	88.0%	18%	73%	87%	72%	59%	62%	48%	49%	62%	81%	96%	93%
	Bare	4579	-1.8%	94.7%	6%	42%	72%	87%	76%	81%	75%	68%	74%	70%	68%	82%
	Cultivated	1000	-4.4%	76.7%	15%	37%	59%	83%	95%	90%	76%	74%	69%	66%	67%	68%
	Wetland	1000	-4.5%	89.8%	17%	52%	52%	66%	68%	69%	82%	91%	97%	79%	66%	60%
Temperate, dry summer	Water	1000	-0.4%	99.5%	43%	54%	52%	71%	82%	68%	79%	84%	96%	63%	52%	56%
	Impervious	1000	-2.4%	78.4%	32%	51%	84%	84%	64%	56%	74%	88%	86%	60%	63%	58%
	Grass/shrub	1000	-4.8%	78.7%	70%	61%	54%	62%	71%	72%	75%	87%	85%	77%	47%	39%
	Forest	1000	-2.2%	86.1%	32%	65%	74%	86%	59%	63%	76%	86%	87%	70%	51%	51%
	Bare	Insufficient samples														
	Cultivated	1000	-7.5%	62.5%	21%	26%	65%	77%	65%	68%	70%	80%	92%	81%	76%	80%
	Wetland	1000	-3.7%	90.7%	22%	51%	78%	94%	75%	65%	84%	93%	83%	62%	56%	36%
Cold, dry/hot summer	Water	1311	-0.4%	99.4%	57%	42%	61%	49%	72%	67%	79%	76%	81%	79%	73%	63%
	Impervious	1000	-2.3%	92.6%	22%	28%	49%	64%	62%	79%	85%	87%	83%	86%	83%	70%
	Grass/shrub	1000	-21.5%	82.3%	33%	40%	31%	35%	57%	85%	81%	84%	91%	93%	89%	80%
	Forest	1000	-2.4%	92.7%	33%	30%	42%	48%	74%	84%	92%	92%	85%	78%	75%	68%
	Bare	Insufficient samples														
	Cultivated	1000	-10.7%	89.3%	33%	39%	34%	34%	57%	81%	81%	83%	92%	94%	91%	81%
	Wetland	727	-3.7%	88.3%	38%	44%	29%	66%	81%	71%	84%	92%	81%	83%	78%	54%
Cold, no dry season	Water	689	-0.9%	96.9%	39%	60%	74%	53%	78%	78%	75%	72%	75%	75%	61%	62%
	Impervious	1000	-4.7%	84.2%	17%	45%	48%	46%	93%	98%	82%	81%	85%	75%	60%	69%
	Grass/shrub	1000	-7.3%	81.0%	35%	10%	43%	60%	85%	93%	76%	84%	82%	87%	74%	72%
	Forest	1000	-2.9%	89.1%	47%	56%	38%	41%	99%	87%	74%	86%	90%	72%	44%	67%
	Bare	Insufficient samples														
	Cultivated	1000	-5.3%	88.9%	33%	35%	35%	37%	84%	95%	81%	88%	83%	85%	64%	80%
	Wetland	1000	-4.5%	85.4%	40%	42%	38%	73%	100%	98%	73%	79%	76%	67%	49%	64%
All Regions	Water	5000	-0.5%	97.5%	65%	58%	40%	46%	50%	57%	83%	93%	99%	80%	63%	66%
	Impervious	5000	-2.1%	80.5%	3%	38%	63%	74%	81%	78%	81%	80%	83%	72%	71%	77%
	Grass/shrub	5000	-5.9%	71.7%	21%	34%	42%	48%	67%	84%	79%	82%	88%	92%	88%	77%
	Forest	5000	-1.4%	88.1%	5%	67%	63%	66%	78%	67%	68%	74%	86%	75%	74%	78%
	Bare	5000	-2.9%	92.3%	1%	35%	61%	78%	75%	86%	79%	75%	81%	76%	75%	79%
	Cultivated	5000	-5.0%	77.5%	14%	32%	37%	57%	75%	85%	80%	86%	88%	86%	82%	77%
	Wetland	5000	-2.4%	89.4%	20%	50%	45%	81%	93%	80%	84%	89%	81%	71%	59%	49%

References

Bauer, M.E., Yuan, F., Sawaya, K.E., 2004. Multi-Temporal Landsat Image Classification and Change Analysis of Land Cover in the Twin Cities (Minnesota) Metropolitan Area. In: Analysis of Multi-Temporal Remote Sensing Images. WORLD SCIENTIFIC, Joint Research Centre, Ispra, Italy, pp. 368–375. https://doi.org/10.1142/9789812702630_0041.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and Future Köppen-Geiger Climate Classification Maps at 1-Km Resolution. Sci. Data 5 (1), 180214. https://doi.org/10.1038/sdata.2018.214.

Chen, J., Cao, X., Peng, S.u., Ren, H., 2017. Analysis and Applications of GlobeLand30: A Review. ISPRS Int. J. Geo Inf. 6 (8), 230. https://doi.org/10.3390/ijgi6080230.

Chen, Y., Weng, Q., Tang, L., Zhang, X., Bilal, M., Li, Q., 2022. Thick Clouds Removing From Multitemporal Landsat Images Using Spatiotemporal Neural Networks. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2020.3043980.

Dewitz, Jon. 2000. "National Land Cover Database (NLCD) 1992 Land Cover Conterminous United States." U.S. Geological Survey. DOI: 10.5066/P92G34R9.

Dwyer, J.L., Roy, D.P., Sauer, B., Jenkerson, C.B., Zhang, H.K., Lymburner, L., 2018. Analysis Ready Data: Enabling Analysis of the Landsat Archive. Remote Sens. (Basel) 10 (9), 1363. https://doi.org/10.3390/rs10091363.

Egorov, A.V., Roy, D.P., Zhang, H.K., Li, Z., Yan, L., Huang, H., 2019. Landsat 4, 5 and 7 (1982 to 2017) Analysis Ready Data (ARD) Observation Coverage over the Conterminous United States and Implications for Terrestrial Monitoring. Remote Sensing 11 (4), 447. https://doi.org/10.3390/rs11040447.

Frantz, D., Rufin, P., Janz, A., Ernst, S., Pflugmacher, D., Schug, F., Hostert, P., 2023. Understanding the Robustness of Spectral-Temporal Metrics across the Global Landsat Archive from 1984 to 2019 – a Quantitative Evaluation. Remote Sens. Environ. 298 (December), 113823 https://doi.org/10.1016/j.rse.2023.113823.

Frohn, R.C., D'Amico, E., Lane, C., Autrey, B., Rhodus, J., Liu, H., 2012. Multi-Temporal Sub-Pixel Landsat ETM+ Classification of Isolated Wetlands in Cuyahoga County, Ohio, USA. Wetlands 32 (2), 289–299. https://doi.org/10.1007/s13157-011-0254-8.

Griffiths, P., Nendel, C., Hostert, P., 2019. Intra-Annual Reflectance Composites from Sentinel-2 and Landsat for National-Scale Crop and Land Cover Mapping. Remote Sens. Environ. 220 (January), 135–151. https://doi.org/10.1016/j.rse.2018.10.031.

Guerschman, J.P., Paruelo, J.M., Di Bella, C., Giallorenzi, M.C., Pacin, F., 2003. Land Cover Classification in the Argentine Pampas Using Multi-Temporal Landsat TM Data. Int. J. Remote Sens. 24 (17), 3381–3402. https://doi.org/10.1080/0143116021000021288.

Hansen, M.C., Potapov, P.V., Goetz, S.J., Turubanova, S., Tyukavina, A., Krylov, A., Kommareddy, A., Egorov, A., 2016. Mapping Tree Height Distributions in Sub-Saharan Africa Using Landsat 7 and 8 Data. Remote Sens. Environ. 185 (November), 221–232. https://doi.org/10.1016/j.rse.2016.02.023.

Heydari, S.S., Mountrakis, G., 2019. Meta-Analysis of Deep Neural Networks in Remote Sensing: A Comparative Study of Mono-Temporal Classification to Support Vector Machines. ISPRS J. Photogramm. Remote Sens. 152 (June), 192–210. https://doi.org/10.1016/j.isprsjprs.2019.04.016.

Jannat, F.-E., Willis, A.R., 2022. "Improving Classification of Remotely Sensed Images with the Swin Transformer." In SoutheastCon 2022, 611–18. Mobile, AL, USA: IEEE. DOI: 10.1109/SoutheastCon48659.2022.9764016.

Julien, Y., Sobrino, J.A., Jiménez-Muñoz, J.-C., 2011. Land Use Classification from Multitemporal Landsat Imagery Using the Yearly Land Cover Dynamics (YLCD) Method. Int. J. Appl. Earth Obs. Geoinf. 13 (5), 711–720. https://doi.org/10.1016/j.jag.2011.05.008.

Kantakumar, L.N., Neelamsetti, P., 2015. Multi-Temporal Land Use Classification Using Hybrid Approach. Egypt. J. Remote Sens. Space Sci. 18 (2), 289–295. https://doi.org/10.1016/j.ejrs.2015.09.003.

Karakizi, C., Karantzalos, K., Vakalopoulou, M., Antoniou, G., 2018. Detailed Land Cover Mapping from Multitemporal Landsat-8 Data of Different Cloud Cover. Remote Sens. (Basel) 10 (8), 1214. https://doi.org/10.3390/rs10081214.

Li, C., Gong, P., Wang, J., Zhu, Z., Biging, G.S., Yuan, C., Tengyun, Hu., et al., 2017. The First All-Season Sample Set for Mapping Global Land Cover with Landsat-8 Data. Science Bulletin 62 (7), 508–515. https://doi.org/10.1016/j.scib.2017.03.011.

Liu, J., J. Heiskanen, E. Aynekulu, and P. K. E. Pellikka. 2015. "Seasonal Variation of Land Cover Classification Accuracy of Landsat 8 Images in Burkina Faso." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-7/W3 (April): 455–60. DOI: 10.5194/isprsarchives-XL-7-W3-455-2015.

Lymburner, L., McIntyre, A., Li, F., Ip, A., Thankappan, M., Sixsmith, J., 2013. In: Creating Multi-Sensor Time Series Using Data from Landsat-5 TM and Landsat-7 ETM+ to Characterise Vegetation Dynamics. IEEE, Melbourne, Australia, pp. 961–993. https://doi.org/10.1109/IGARSS.2013.6721321.

McInerney, D., Kempeneers, P., Marron, M., McRoberts, R.E., 2019. Analysis of Broadleaf Encroachment in Coniferous Forest Plantations Using Multi-Temporal Satellite Imagery. Int. J. Appl. Earth Obs. Geoinf. 78 (June), 130–137. https://doi.org/10.1016/j.jag.2018.12.005.

Mountrakis, G., Heydari, S.S., 2023. Harvesting the Landsat Archive for Land Cover Land Use Classification Using Deep Neural Networks: Comparison with Traditional Classifiers and Multi-Sensor Benefits. ISPRS J. Photogramm. Remote Sens. 200 (June), 106–119. https://doi.org/10.1016/j.isprsjprs.2023.05.005.

Pflugmacher, D., Rabe, A., Peters, M., Hostert, P., 2019. Mapping Pan-European Land Cover Using Landsat Spectral-Temporal Metrics and the European LUCAS Survey. Remote Sens. Environ. 221 (February), 583–595. https://doi.org/10.1016/j.rse.2018.12.001.

Rufin, P., Müller, H., Pflugmacher, D., Hostert, P., 2015. Land Use Intensity Trajectories on Amazonian Pastures Derived from Landsat Time Series. Int. J. Appl. Earth Obs. Geoinf. 41 (September), 1–10. https://doi.org/10.1016/j.jag.2015.04.010.

Rußwurm, M., Körner, M., 2017. "Multi-Temporal Land Cover Classification with Long Short-Term Memory Neural Networks." The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-1/W1 (May): 551–58. DOI: 10.5194/isprs-archives-XLII-1-W1-551-2017.

Senf, C., Leitão, P.J., Pflugmacher, D., Van Der Linden, S., Hostert, P., 2015. Mapping Land Cover in Complex Mediterranean Landscapes Using Landsat: Improved Classification Accuracies from Integrating Multi-Seasonal and Synthetic Imagery. Remote Sens. Environ. 156 (January), 527–536. https://doi.org/10.1016/j.rse.2014.10.018.

Shen, Jianxiu, and Fiona H. Evans. 2020. "Empirical Combination of Landsat 7 and 8 Imagery to Detect the Phenological Changes in Rainfed Cropland Vegetation." In IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, 5258–61. Waikoloa, HI, USA: IEEE. DOI: 10.1109/IGARSS39084.2020.9323444.

Singh, K.K., Gray, J., 2020. Mapping Understory Invasive Plants in Urban Forests with Spectral and Temporal Unmixing of Landsat Imagery. Photogramm. Eng. Remote Sens. 86 (8), 509–518. https://doi.org/10.14358/PERS.86.8.509.

USGS. 2018. "National Land Cover Database." September 11, 2018. https://www.usgs.gov/centers/eros/science/national-land-cover-database.

Wang, Z., Mountrakis, G., 2023. Accuracy Assessment of Eleven Medium Resolution Global and Regional Land Cover Land Use Products: A Case Study over the Conterminous United States. Remote Sens. (Basel) 15 (12), 3186. https://doi.org/10.3390/rs15123186.

Wolter, P.T., Mladenoff, D.J., Host, G.E., Crow, T.R., 1995. Improved Forest Classification in the Northern Lake States Using Multi-Temporal Landsat Imagery. Photogramm. Eng. Remote Sens. 61 (9), 1129–1143.

Zheng, B., Campbell, J.B., De Beurs, K.M., 2012. Remote Sensing of Crop Residue Cover Using Multi-Temporal Landsat Imagery. Remote Sens. Environ. 117 (February), 177–183. https://doi.org/10.1016/j.rse.2011.09.016.

Zhu, Z., Woodcock, C.E., 2014. Continuous Change Detection and Classification of Land Cover Using All Available Landsat Data. Remote Sens. Environ. 144 (March), 152–171. https://doi.org/10.1016/j.rse.2014.01.011.