9-2023

Demonstration of Large Area Land Cover Classification with a One Dimensional Convolutional Neural Network Applied to Single Pixel Temporal Metric Percentiles

Hankui K. Zhang

David P. Roy

Dong Luo

Remote Sensing of Environment 295 (2023) 113653

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Demonstration of large area land cover classification with a one dimensional convolutional neural network applied to single pixel temporal metric percentiles

Hankui K. Zhang [a], David P. Roy [b,*], Dong Luo [a]

[a] Geospatial Sciences Center of Excellence, Department of Geography and Geospatial Sciences, South Dakota State University, Brookings, SD 57007, USA
[b] Department of Geography, Environment, & Spatial Sciences, and the Center for Global Change and Earth Observations, Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

Edited by Dr Marie Weiss

Keywords:
Land cover
Time series
Temporal metric percentiles
Convolutional neural network
Random forest
Deep learning
Large area classification
Landsat

ABSTRACT

Over large areas, land cover classification has conventionally been undertaken using satellite time series. Typically temporal metric percentiles derived from single pixel location time series have been used to take advantage of spectral differences among land cover classes over time and to minimize the impact of missing observations. Deep convolutional neural networks (CNNs) have demonstrated potential for land cover classification of single date images. However, over large areas and using time series their application is complicated because they are sensitive to missing observations and they may misclassify small and spatially fragmented surface features due to their spatial patch-based implementation. This study demonstrates, for the first time, a one-dimensional (1D) CNN single pixel time series land classification approach that uses temporal percentile metrics and that does not have these issues. This is demonstrated for all the Conterminous United States (CONUS) considering two different 1D CNN structures with 5 and 8 layers, respectively. CONUS 30 m land cover classifications were derived using all the available Landsat-5 and -7 imagery over a seven-month growing season in 2011 with 3.3 million 30 m land cover class labelled samples extracted from the contemporaneous CONUS National Land Cover Database (NLCD) 16 class land cover product. The 1D CNNs and, a conventional random forest model, were trained using 10%, 50% and 90% samples, and the classification accuracies were evaluated with an independent 10% proportion. Temporal metrics were classified using 5, 7 and 9 percentiles for each of five Landsat reflective wavelength bands and their eight band ratios. The CONUS and detailed 150 × 150 km classification results demonstrate that the approach is effective at scale and locally. The 1D CNN classification land cover class boundaries were preserved for small axis dimension features, such as roads and rivers, with no stripes or anomalous spatial patterns. The 8-layer 1D CNN provided the highest overall classification accuracies and both the 5-layer and 8-layer 1D CNN architectures provided higher accuracies than the random forest by 1.9% - 2.8% which as all the accuracies were > 83% is a meaningful increase. The CONUS overall classification accuracies increased marginally with the number of percentiles (86.21%, 86.40%, and 86.43% for 5, 7 and 9 percentiles, respectively) using the 8-layer 1D-CNN. Class specific producer and user accuracies were quantified, with lower accuracies for the developed land, crop and pasture/hay classes, but no systematic pattern among classes with respect to the number of temporal percentiles used. Application of the trained model to a different year of CONUS Landsat ARD showed moderately decreased accuracy (80.79% for 7 percentiles) that we illustrate is likely due to different intra-annual surface variations between years. These encouraging results are discussed with recommended research for deep learning using temporal metric percentiles.

1. Introduction

Land cover is a critical descriptor of the Earth's terrestrial surface, and regional to global coverage medium spatial resolution land cover maps are needed to monitor human activity and the state of the land surface (Townshend, 1992). The current state-of-the-practice satellite

* Corresponding author.
E-mail addresses: hankui.zhang@sdstate.edu (H.K. Zhang), roydavi1@msu.edu (D.P. Roy), dong.luo@sdstate.edu (D. Luo).

https://doi.org/10.1016/j.rse.2023.113653
Received 6 October 2022; Received in revised form 18 April 2023; Accepted 24 May 2023
Available online 8 June 2023

land cover mapping approach is to apply a non-parametric supervised classifier, such as random forest, to image time series to take advantage of spectral differences among land cover classes over time (Wulder et al., 2018). Convolutional neural networks (CNN) were first applied to single digital images to classify if they contained a particular class and predominantly used information based on the spatial relationships among pixels (LeCun et al., 1989; Sahiner et al., 1996). Deep convolutional neural networks (hereafter referred to for brevity as CNN) refined the CNN structure with a series of deep learning techniques (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014) and in the last several years have been applied to classify land cover in single date satellite images (Huang et al., 2018; Kellenberger et al., 2018; Mahdianpari et al., 2018; Srivastava et al., 2019; Tong et al., 2020; Yuan et al., 2020; Karra et al., 2021; Brown et al., 2022). In these approaches the CNN is conventionally applied to image patches extracted from single images composed of $n \times n$ pixels and one to several spectral bands. The patch supports the extraction of spatial features that are used to help classify the central patch pixel, and an image is classified by classifying patches extracted systematically across it. Integrating temporal information for CNN based land cover classification is challenging. Patch-based algorithms have been relatively under-explored in this respect and typically have been implemented by application to images acquired over the same location on different dates stacked into a single multi-band image (Karakizi et al., 2018; Kwak et al., 2021; Fazzini et al., 2021) or composited into a single temporal composite (Rosentreter et al., 2020; Chen et al., 2020; Hosseiny et al., 2021). Other architectures such as recurrent neural networks (RNN) that were developed to accommodate multi-temporal information found in one dimensional time series (Rumelhart et al., 1986; Pascanu et al., 2013) have been applied to image patch time series. In this approach a CNN is applied to patches in each image and then the CNN patch results across the time series are combined using an RNN to derive the land cover class of the patch center pixel (Interdonato et al., 2019; Turkoglu et al., 2021; Masolele et al., 2021; Thorp and Drajat, 2021; Wang et al., 2022). Fully attention based networks (also known as Transformer networks), that perform better than RNN (Vaswani et al., 2017), are starting to be used with CNN and image patch time-series to classify land cover (Liu et al., 2022a; Yang et al., 2022). However, with all patch-based approaches, small and spatially fragmented land cover areas may be blurred because their pixels may be confused with neighboring pixels that have different land cover that occupy the majority of the patch, and certain land cover class boundaries may be overly generalized (Kussul et al., 2017; Stoian et al., 2019; Derksen et al., 2019; Zhang et al., 2020).

Rather than use patches, single pixel time series based land cover classification approaches have been developed by application of a one dimensional (1D) CNN to single pixel time series defined by a $t \times s$ array, where t is the number of observations in the time series and s is the number of spectral bands. Single pixel time series 1D CNN land cover classification has been demonstrated recently over relatively small geographic areas. For example, Pelletier et al. (2019) used a 5-layer 1D CNN to classify Formosat-2 image time series acquired on 46 dates for a 24 × 24 km area in France. Zhong et al. (2019) used a 5-layer 1D CNN to classify Landsat 7 and 8 image time series acquired on 37 dates in 2014 for a county in California. Wang et al. (2020) used a 5-layer 1D CNN to classify Sentinel-2 and Sentinel-1 image time series acquired in years 2017 to 2019 for two states in India. Rußwurm and Körner (2020) used a 5-layer 1D CNN to classify Sentinel-2 image time series for three 100 × 100 km areas in Germany. Debella-Gilo and Gjertsen (2021) used a 5-layer 1D CNN to classify Sentinel-2 image time series acquired across Norway that were temporally composited into 27 7-day, 14 14-day, 9 21-day and 7 28-day intervals. Lobert et al. (2021) used a 4-layer 1D CNN to classify Landsat-8, Sentinel-2, and Sentinel-1 time series at three study sites in Germany; the Sentinel-2 and Landsat-8 data were composited into 44 6-day intervals so that they were consistent with the Sentinel-1 revisit interval. Zhao et al. (2021) used a 5-layer 1D CNN to classify Sentinel-2 image time series acquired on 37 dates across 4 Sentinel-2 110 × 110 km tiles in Hebei, China. Lange et al. (2022) used a 4-layer 1D CNN to classify Sentinel-2 time series for grassland areas in Germany.

The 1D CNN single pixel time series land cover classification approach explicitly supports the extraction of features that capture spectral differences among land cover classes over time, for example, associated with phenological variations in vegetation greenness and growth stage, and does not have the small object misclassification and land cover boundary generalization drawbacks that can occur with the patch-based CNN approach. However, for reliable application, it requires that there are no missing observations in the single pixel time series. Missing satellite observations, often with irregular observation temporal cadence and large temporal gaps, are common in medium resolution time series provided by Landsat or Sentinel-2 (Egorov et al., 2019; Li and Roy, 2017; Yan and Roy, 2020) and by high temporal and spatial resolution commercial satellite data (Roy et al., 2021). One solution, adopted by single pixel 1D CNN land cover classification researchers, is to fill missing and cloud obscured observations by interpolation between consecutive valid observations (Zhong et al., 2019; Pelletier et al., 2019; Wang et al., 2020; Rußwurm and Körner, 2020; Debella-Gilo and Gjertsen, 2021; Lobert et al., 2021; Lange et al., 2022). Interpolation is unreliable however when gaps occur in periods of rapid surface change, for example, in the growing season, or when the gaps have long duration, or when there are residual clouds, shadows, and poorly atmospherically corrected observations that are not reliably flagged and so are used incorrectly in the interpolation (Yan and Roy, 2020).

In this study, we present a large area, conterminous United States (CONUS) assessment of the single pixel time series 1D CNN approach applied to temporal metric percentiles. Temporal metric percentiles decompose irregular distributed time series into a reduced fixed number of features that can be conveniently used for classification purposes. There is a long heritage in their use for large area land cover classification from a variety of sensors including AVHRR (DeFries et al., 1995; De Fries et al., 1998), Landsat (Hansen et al., 2014; Zhang and Roy, 2017) and Sentinel-2 (Schug et al., 2020; Grabska et al., 2020) and, for example, to generate systematically the global NASA MODIS land cover product (Sulla-Menashe et al., 2019). Temporal metric percentiles are extracted at gridded pixel locations by ranking spectral reflectance and/or spectral band ratio values over the image time series, and then selecting percentiles, for example, the 25th, 50th, and 75th percentiles of a spectral band ratio over the time series. The number of percentiles used is discussed in Section 2 but in principle a larger number of percentiles will better capture seasonal surface variations but require a larger number of observations and there must be at least as many observations as there are percentiles. Importantly, temporal metric percentiles are insensitive to phenological differences, as the metrics do not capture the timing but rather the amplitude of the reflectance variation, and are generally insensitive to missing observations in the time series. Despite these advantages, temporal metric percentiles have not been demonstrated for large area CNN land cover classification.

The objectives of this study were to (i) demonstrate, for the first time, large area single pixel time series 1D CNN land cover classification of temporal metric percentiles extracted from 30 m Landsat data, (ii) investigate the sensitivity of the classification results to using different numbers of temporal metric percentiles, 1D CNN architecture complexity, and training data amount, (iii) evaluate the single pixel time series 1D CNN land cover classification accuracy results and compare with the classification results derived using a conventional Random Forest classifier as a benchmark. The land cover classifications were developed for the CONUS at 30 m resolution using 2011 growing season of Landsat-5 Thematic Mapper (TM) and Landsat-7 Enhanced Thematic Mapper Plus (ETM+) surface reflectance provided by the United States Geological Survey (USGS) Landsat analysis ready data (ARD) (Dwyer et al., 2018). Training data land cover class labels were defined using the 16 classes 30 m CONUS 2011 USGS National Land Cover Database

(NLCD) product (Homer et al., 2015). More than 3.31 million 30 m pixel locations sampled across the CONUS, subject to stringent Landsat ARD quality filtering, were used to define a NLCD land cover class labelled pool that was split into 30 m training and evaluation proportions. Two CNN structures designed with different complexity, i.e., a 5-layer CNN with 0.2 million learnable coefficients, and an 8-layer CNN with 2.1 million learnable coefficients, were used to examine the classification accuracy sensitivity to CNN structure complexity. The classification results were quality assessed by visual comparison with the NLCD land cover product at CONUS scale, and also in detail over 150 × 150 km ARD tiles. The classification accuracy was quantified by per-pixel comparison with the evaluation data sample class values and summarized by overall and class-specific producer's and user's land cover classification accuracies. To further demonstrate the approach, the classification accuracies found by application of the trained model to a different year (2006) of growing season Landsat-5 TM and Landsat-7 ETM+ surface reflectance ARD were quantified. The paper concludes with a discussion of the potential of other deep learning (non-CNN structures) for large area and single pixel land cover classification and the training data and codes in this study are publicly available to facilitate future comparison studies.

2. Data and processing undertaken to derive Landsat temporal metrics and land cover training and evaluation data

2.1. Landsat data

Two years of CONUS growing season 30 m Landsat-5 TM and Landsat-7 ETM+ analysis ready data (ARD) for 2011 and 2006, respectively were used. The CONUS growing season was defined as seven months from April 1st to October 31st 2011 following Hansen et al. (2014) and Zhang and Roy (2017). This seven month definition also mitigates the impacts of winter snow and unreliable Landsat discrimination of cloud and snow (Skakun et al., 2022). The Landsat data for 2011 and 2006 were used because of the availability of the 30 m National Land Cover Database (NLCD) CONUS land cover product for these two years that were used as a source of training and evaluation data (Section 2.3). All the Landsat-5 TM and Landsat-7 ETM+ 30 m bands, except the blue band (0.45–0.52 μm) that is highly sensitive to atmospheric scattering and is less reliably atmospherically corrected (Ju et al., 2012; Roy et al., 2014), were used. Specifically, the green 0.52–0.60 μm, red 0.63–0.69 μm, near infrared (NIR) 0.76–0.90 μm, first shortwave infrared (SWIR1) 1.55–1.75 μm, and second shortwave infrared (SWIR2) 2.09–2.35 μm bands were used.

The Landsat ARD are generated by the USGS in fixed non-overlapping 5000 × 5000 30 m pixel (150 × 150 km) tiles in the Albers Equal Area Conic projection (Dwyer et al., 2018). Each individual orbit of Landsat data overlapping an ARD tile is stored independently. There are 512 CONUS ARD land tiles referenced by horizontal (h) and vertical (v) tile coordinates varying from 0 to 32 and from 0 to 21, respectively, and the coordinated are referred to later in this paper (e.g., tile h07v13). The Landsat ARD are available as both top of atmosphere (TOA) and atmospherically corrected (i.e., surface) reflectance and in this study the surface reflectance were used with the associated per-pixel saturation status, cloud, cloud shadow, cirrus cloud, and snow masks. The Landsat ARD have consistent geometric accuracy and are processed to ensure image-to-image tolerances of ≤12 m radial root mean square error (Dwyer et al., 2018). The Landsat ARD are not corrected for bidirectional reflectance distribution function (BRDF) effects, although these effects are non-negligible in Landsat data (Gao et al., 2014; Roy et al., 2016). Therefore, in this study the Landsat ARD surface reflectance for each sensor band and each ARD pixel location was adjusted to a nadir view to provide Landsat nadir BRDF-adjusted reflectance (NBAR). The published *c*-factor approach was used to generate NBAR and is based on multiplying the Landsat surface reflectance with the ratio of the reflectances modeled using fixed global average MODIS BRDF spectral model parameters for the observed Landsat and for a nadir view and a specified solar zenith (Roy et al., 2016). The solar zenith was defined with a model developed for this purpose that provides modelled solar zenith angles close to (<0.5° difference) the observed Landsat solar zenith at the time of overpass and that varies smoothly over the year and latitudinally (Zhang et al., 2016).

A total of 14,785 Landsat-5 and 14,680 Landsat-7 ARD tile granules covering the CONUS over the seven month growing season from April 1st to October 31st 2011 were used. In addition, a total of 14,531 Landsat-5 and 14,351 Landsat-7 ARD tile granules in 2006 were used. The ARD data were quality filtered to remove all ARD pixel observations flagged as cloud, cirrus cloud, or as saturated. Fig. 1 illustrates the irregular spatial and temporal availability of Landsat observations, for a 512 × 512 30 m pixel subset of an ARD tile in 2011. Figs. 1(a)-(c) show example daily ARD tile surface reflectance NBAR for different dates and sensors selected in the seven month growing season, illustrating (a) Landsat-5 imagery containing clouds, (b) Landsat-7 imagery with missing stripes caused by the ETM+ scan line corrector (SLC) anomaly that caused a 22% pixel loss (Maxwell et al., 2007), and (c) the edge of a Landsat-5 image that is jagged because of the staggered spectral band readout at the swath edge, with a region of no data where Landsat-5 did not overpass that day. Notably, the temporal cadence of the Landsat observations is irregular, i.e., not only are there spatial gaps, but the cloud-free observations are not available on a regular temporal basis and this is well documented (Brooks et al., 2012; Egorov et al., 2019; Roy and Yan, 2020). Thus, the total number of quality filtered Landsat-5 and Landsat-7 observations over the seven month growing season shown in Fig. 1(d) varies geographically, and, for example, 14.5% of the illustrated ARD pixel locations had ≤5 quality filtered observations (red) and only 23.5% had ≥11 quality filtered observations (purple).

The number of quality filtered Landsat-5 TM and Landsat-7 ETM+ observations at each CONUS ARD pixel location over the seven month growing season varied from a minimum of zero to a maximum of 54. Fig. 2 shows the percentage of the CONUS ARD 30 m land pixel locations that had at least n_{valid} (1, 2, …, 54) quality filtered observations over the 2011 growing season, with 99.96%, 99.90%, 99.81%, 99.43%, 97.88%, and 93.56% of locations having at least 1, 3, 5, 7, 9 and 11 observations, respectively. This information is used to justify the selection of temporal metric percentiles for the classification experiments described below. The year 2006 growing season data had similar distributions (not illustrated) with 99.99%, 99.98%, 99.87%, 99.27%, 97.32%, and 93.03% of locations having at least 1, 3, 5, 7, 9 and 11 observations, respectively.

2.2. Landsat temporal metric percentile generation

Temporal metric percentiles were derived at each CONUS ARD 30 m pixel location from the quality filtered ARD surface NBAR time series acquired over the seven month growing season. They were derived for the five Landsat bands (green, red, NIR, SWIR1 and SWIR2), and for the eight possible two band normalized ratios of these bands, i.e., (NIR-red)/(NIR + red), (SWIR1-green)/(SWIR1 + green), (SWIR1-red)/(SWIR1 + red), (SWIR1-NIR)/(SWIR1 + NIR), (SWIR2-green)/(SWIR2 + green), (SWIR2-red)/(SWIR2 + red), (SWIR2-NIR)/(SWIR2 + NIR), and (SWIR2-SWIR1)/(SWIR2 + SWIR1). The temporal metric percentiles were extracted at each ARD gridded pixel location by ranking these values over the growing season and then selecting percentile values in the conventional way. For example, given n growing season quality filtered observations at an ARD pixel location, the k^{th} NIR band percentile is the NIR growing season value with (k/100 × n) observations that have smaller or equal NIR value. Following convention, an odd number of percentiles selected with a symmetrical distribution around the 50th percentile (i.e., the median) were used. Typically, Landsat classification studies have been undertaken using five or more percentiles. For example, five percentiles were derived over the growing season from one year of Landsat-7 data (Yan and Roy, 2015), from three

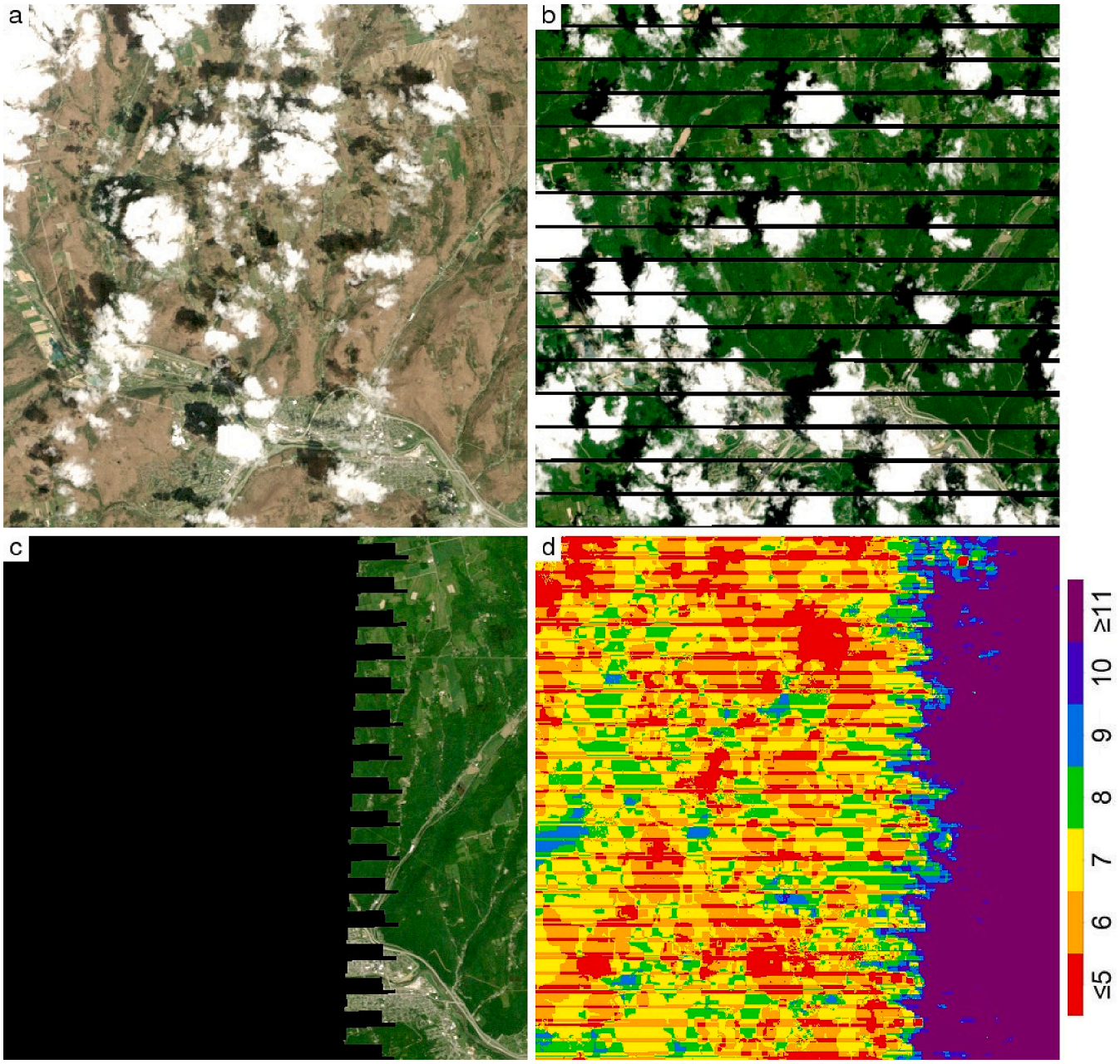

Fig. 1. Illustration of the irregular spatial and temporal availability of Landsat observations for a 512 × 512 30 m pixel subset of ARD tile h27v06 over the West Hill State Forest in New York state (close to Corning, NY). (a) Landsat-5 TM acquired May 9th 2011, (b) Landsat-7 ETM+ acquired June 2nd, 2011, (c) Landsat-5 TM acquired June 3rd 2011, and (d) the number of quality filtered Landsat-5 TM and Landsat-7 ETM+ observations over the April 1st to October 31st 2011 growing season. (a)-(c) show Landsat true color (red, green and blue) surface NBAR. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

years of Landsat-8 data (Pflugmacher et al., 2019), and from four years of Landsat-7 and -8 data (Azzari and Lobell, 2017). Landsat studies to map surface change typically use a greater number of percentiles, for example, seven percentiles were extracted from six years of Landsat-7 data (Potapov et al., 2012) and from six years of Landsat-5 and Landsat-7 data (Margono et al., 2012). In principle using a larger number of percentiles will better capture seasonal surface variations and as a minimum there must be at least as many Landsat observations as there are percentiles. Therefore, in this study, CONUS land cover classification experiments using $n_p = 5$, 7 and 9 percentiles were considered. They were defined specifically by the 10th, 25th, 50th, 75th, and 90th percentiles ($n_p = 5$), 10th, 20th, 35th, 50th, 65th, 80th, and 90th percentiles ($n_p = 7$), and 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentiles ($n_p = 9$) of the quality filtered NBAR for the five Landsat bands and for the eight normalized NBAR band ratios. In this way a total of $n_p \times 13$ temporal metrics were derived for each set of $n_p = 5$, 7 or 9 percentiles.

Land cover classification experiments were not undertaken with $n_p = 11$ percentiles because only 93.56% of the CONUS ARD 30 m pixel locations had ≥11 quality filtered growing season observations (Fig. 2). Therefore, 6.44% of the CONUS ARD pixels would be unclassified with $n_p = 11$ due to the requirement that there must be at least as many Landsat observations as there are percentiles. This is also evident in Fig. 3 that shows the proportion of CONUS ARD pixels in 3 × 3 km grid cells with ≥5, ≥7, ≥9, and ≥ 11 Landsat-5 and Landsat-7 quality filtered growing season observations in 2011. Fewer quality filtered growing

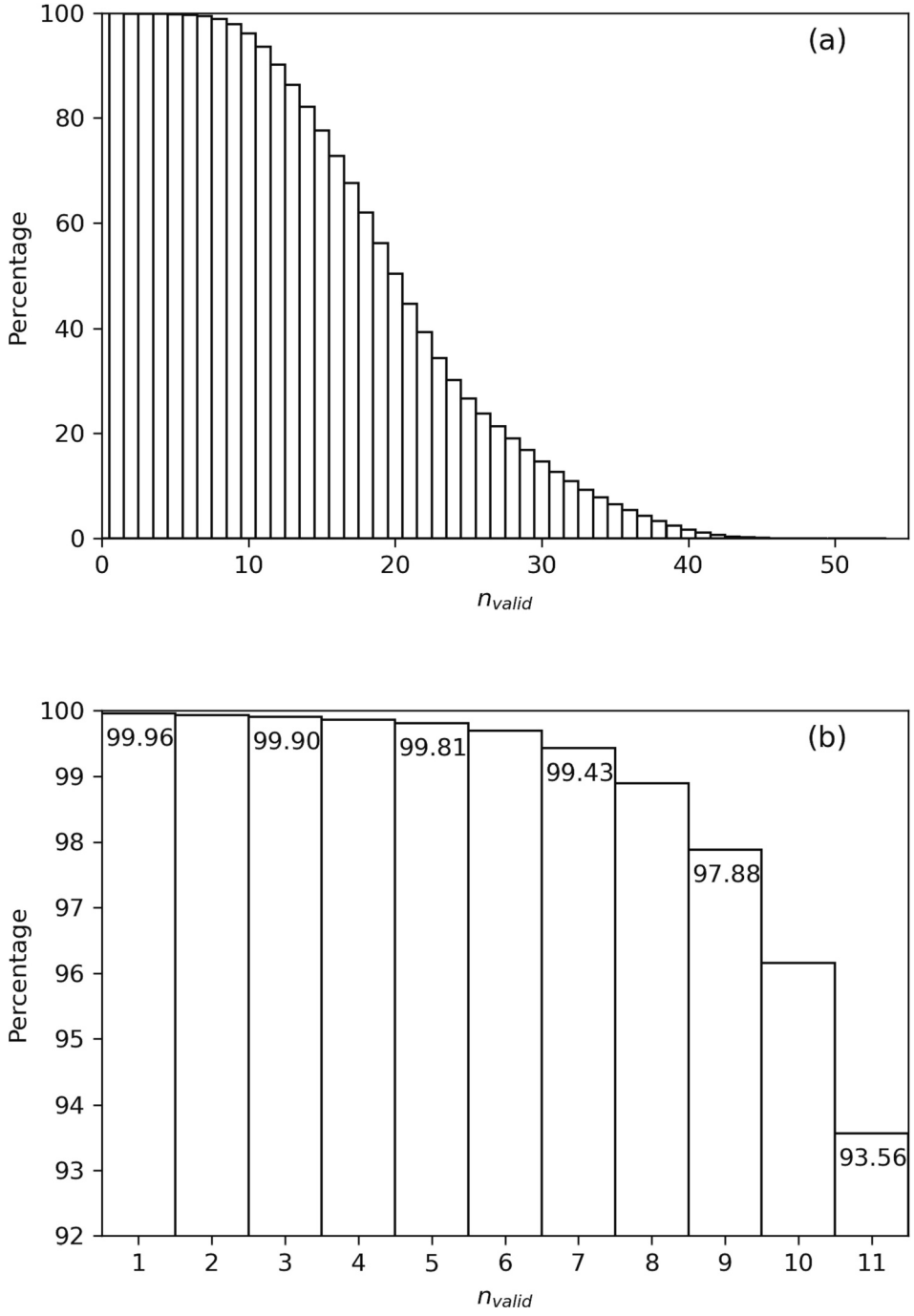

Fig. 2. Percentage of the CONUS ARD 30 m pixel locations that had at least n_{valid} quality filtered Landsat-5 and Landsat-7 observations over the April 1st to October 31st 2011 growing season, (a) shows percentages for $1 \leq n_{valid} \leq 54$, (b) shows percentages for $1 \leq n_{valid} \leq 11$. There were $>$ 8977 million 30 m CONUS ARD pixel locations with $\geq$1 quality filtered growing season observations.

season observations occur in localized regions, predominantly where the CONUS is cloudy at the time of Landsat overpass, i.e., in the north-east and north-west (Kovalskyy and Roy, 2015). In addition, there are fewer observations in the approximately north-south strips where adjacent Landsat orbit swaths do not overlap - the number of observations in the swath overlapping regions is generally twice that in the non-overlapping swath regions (Egorov et al., 2019). Land cover classification experiments were not undertaken using $n_p = 3$ percentiles because so few percentiles are not expected to capture seasonal surface variations and previous Landsat studies have used $n_p \geq 5$ percentiles.

2.3. National Land Cover Database (NLCD) and land cover training and evaluation data generation

An existing 30 m CONUS land cover map was used as a source of land cover class labels needed to train the classifiers and to evaluate the results, which is a common approach (Zhang and Roy, 2017; Johnson and Mueller, 2021; Zhai et al., 2022). The USGS 2011 National Land Cover Database (NLCD), reprocessed in 2014, was used and has 16 land cover classes (Homer et al., 2015) and a reported 86.8% overall land cover classification accuracy (Wickham et al., 2021). The NLCD is stored in a single CONUS image file in the Albers Equal Area Conic projection and was clipped spatially into the 5000 × 5000 30 m pixel ARD tile grid.

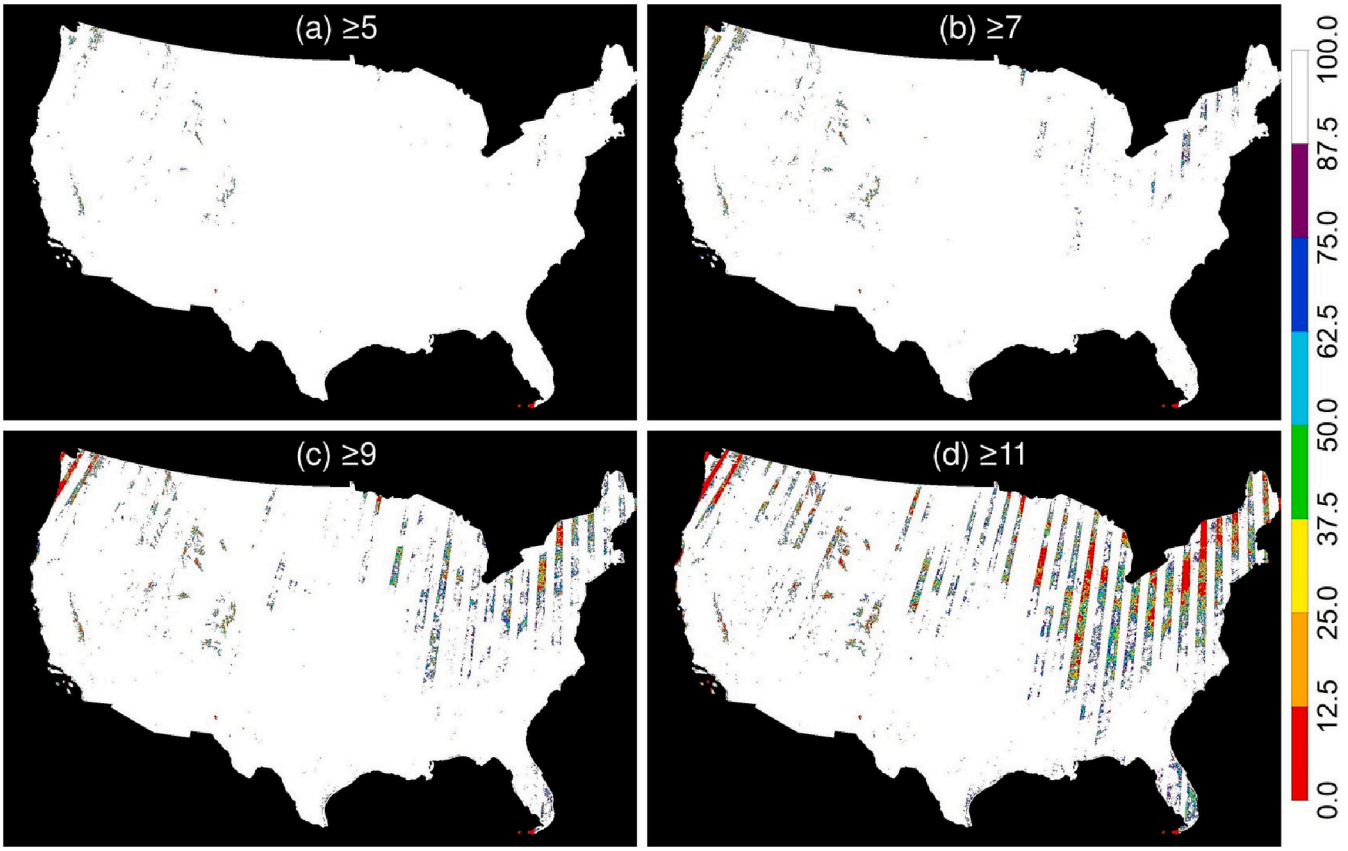

Fig. 3. Proportion of 3 × 3 km (100 × 100 30 m pixel) CONUS grid cells that have (a) ≥5, (b) ≥7, (c) ≥ 9, and (d) ≥11 quality filtered Landsat-5 and Landsat-7 seven month April 1st to October 31st 2011 growing season observations.

Each tile was systematically sampled every 40 pixels (i.e., every 1.2 km) in the column and row directions to reduce land cover spatial autocorrelation effects (Yang et al., 2003; Zhang and Roy, 2017) and then the NLCD land cover class label and quality filtered Landsat 5 and Landsat 7 observations at that pixel location over the 2011 growing season were extracted. The following filtering criteria were applied in the extraction process. Only (i) ARD pixel locations with ≥5 quality filtered year 2011 growing season observations were considered (as land cover classifications based on temporal metrics with $n_p = 5$ and also 7 and 9 percentiles were generated), (ii) ARD pixel locations with the same NLCD land cover class in the surrounding eight 30 m pixels were considered to reduce the impact of Landsat sub-pixel misregistration errors and isolated single pixel NLCD misclassification errors (Colditz et al., 2012; Zhang and Roy, 2017), (iii) NLCD land cover classes with ≥1000 samples were considered to ensure sufficiently representative class samples were used; consequently, the NLCD perennial ice/snow class was not included because of its geographic rarity across the CONUS. A total of 3,314,439 CONUS 30 m pixel locations each with a NLCD land cover class label response variable (Table 1) and predictor variables defined by $n_p \times 13$ temporal metrics derived from the quality filtered Landsat growing season observations at the pixel location, were extracted.

The 3.3 million CONUS samples (Table 1) were divided into proportions used to train the classifiers and to evaluate the accuracy of the classification results, respectively. CONUS land cover classifications were generated independently (Section 3) using 10%, 50%, and 90% of the 3.3 million CONUS samples as training and using predictor variables defined by temporal metrics derived with $n_p = 5$, 7 and 9 percentiles. The different classifications were evaluated with a 10% evaluation proportion so that they could be compared meaningfully. Each set of training and evaluation data was selected randomly from the 3.3 million CONUS samples (Table 1) but ensuring that they were selected at ARD 30 m pixel locations where there were at least as many growing Landsat season observations as the number of percentiles (n_p) used to generate the classification. The numbers of samples for the different proportions are summarized in Table 2. The random selection was undertaken starting for $n_p = 5$, selecting only samples at CONUS ARD pixel locations

Table 1
Number of CONUS ARD 30 m pixel sample locations used to generate the land cover training and evaluation data (Table 2). The numbers in each of the 15 NLCD land cover classes are summarized, with the NLCD legend class ID (Homer et al., 2004) for reference. A total of 3,314,439 locations were used, corresponding to about 0.04% of the number of 30 m CONUS ARD land pixels.

ID	Land cover class	Number of CONUS sample locations	ID	Land cover class	Number of CONUS sample locations
11	Open water	255,725	43	Mixed forest	18,391
21	Developed open-space	11,288	52	Shrub/scrub	823,707
22	Developed low-intensity	3427	71	Grassland/ herbaceous	493,638
23	Developed medium-intensity	2003	81	Pasture/hay	187,039
24	Developed high-intensity	2135	82	Cultivated crops	644,661
31	Barren land	36,562	90	Woody wetlands	103,608
41	Deciduous forest	321,740	95	Emergent herbaceous wetlands	30,714
42	Evergreen forest	379,801			

Table 2
The number of training and evaluation samples used in the different classification experiments. Each sample is composed of the 30 m NLCD land cover class label (i.e., the response variable), and $n_p \times 13$ temporal metrics derived from the quality filtered Landsat-5 and Landsat-7 observations quality filtered Landsat ARD (i.e., the predictor variables).

Proportion (%) selected from Table 1	Number of 30 m CONUS training samples		
	≥5 quality filtered observations, used to train classifiers using $n_p = 5$ percentiles	≥7 quality filtered observations, used to train classifiers using $n_p = 5$ and 7 percentiles	≥9 quality filtered observations, used to train classifiers using $n_p = 5$, 7 and 9 percentiles
~90%	2,982,996	2,975,184	2,935,602
~50%	1,657,220	1,652,880	1,630,890
~10%	331,444	330,576	326,178
Proportion (%) selected from Table 1	Number of 30 m CONUS evaluation samples		
	≥5 quality filtered observations, used to evaluate the above classifications	≥7 quality filtered observations, used to evaluate the above 2 classifications	≥9 quality filtered observations, used to evaluate the above 3 classifications
~10%	331,443	330,571	326,170

with ≥5 quality filtered growing season observations. The $n_p = 7$ samples were selected from the $n_p = 5$ samples, removing samples where there were fewer than 7 quality filtered growing season observations. Similarly, the $n_p = 9$ samples were selected from the $n_p = 7$ samples, removing samples where there were fewer than 9 quality filtered growing season observations. This meant that there were a marginally smaller number of $n_p = 9$ samples than $n_p = 7$ samples (up to 1.33% less for the three proportions), and a marginally smaller number of $n_p = 7$ samples than $n_p = 5$ samples (up to 0.26% less) (Table 2). These random selections were constrained so that the CONUS NLCD land cover class proportions evident in Table 1 were approximately maintained in the resulting training and evaluation samples. This follows conventional non-parametric classification approaches (Weiss and Provost, 2003; Colditz, 2015; Zhang and Roy, 2017) as the naturally occurring class distribution, i.e., a proportional distribution among land cover classes related to the proportion that they occur in reality, will provide training and evaluation data that are representative of CONUS conditions.

In addition, to further demonstrate the applicability of the approach, the NLCD 2006 product was used to derive samples for evaluation of the year 2011 trained classification model applied to year 2006 Landsat ARD (Section 3.4). The NLCD 2006 has the same land cover class and definitions as the NLCD 2011 and was generated with a reported 83.6% overall classification accuracy (Wickham et al., 2021). The NLCD 2006 and corresponding year 2006 seven month growing season Landsat 5 TM and Landsat 7 ETM+ ARD were processed as described above. Only the quality filtered 2006 CONUS pixel samples that had the *same* locations as the NLCD 2011 evaluation samples were used for the experiment. There were slightly fewer year 2006 evaluation samples than in 2011 primarily because of the constraint that the 2006 NLCD land cover class in the surrounding eight 30 m pixels be the same and because of land cover change between 2006 and 2011. For example, there were 325,736 evaluation samples extracted with ≥7 year 2006 growing season quality filtered observations.

3. Deep convolutional neural network (CNN) classification

3.1. Overview of conventional CNN single image patch based land cover classification

To provide context for the single pixel time series CNN land cover classification approach we first overview the conventional patch-based CNN approach that is applied to image patches composed of $n \times n$ pixels and one to several image bands spatially subset from a single image (Huang et al., 2018; Srivastava et al., 2019; Tong et al., 2020; Belenguer-Plomer et al., 2021; Mäyrä et al., 2021; Lu et al., 2022). The CNN structure consists of many sequential convolutional layers followed by several fully connected layers (Fig. 4 black boxes). Each convolutional layer consists of a set of kernels, also known as convolution filters, that are defined by $i \times j \times r$ matrices storing values, termed kernel weights, where i and j are the kernel spatial dimensions and usually the i and j dimensions are the same. Each kernel has an associated single value, termed the bias. The fully connected layers consist of a set of one-dimensional $1 \times l$ vectors storing values, termed vector weights, and each vector has an associated bias term. The number of kernels and vectors in each layer must be pre-defined, although the last fully connected layer must have as many vectors as there are land cover classes. The kernels are three dimensional and r is set as the number of spectral bands in the input image patch for the first convolutional layer and for subsequent layers is equal to the number of kernels in the previous layer. The CNN structure, i.e., the number of layers, the number of kernels per convolutional layer, the kernel spatial dimensions in each convolutional layer, and the number of vectors in the fully connected layers, must be pre-defined before training the CNN. The CNN structure largely dictates the complexity and performance of the classification. Generally, more layers capture different aspects of the training data and aid among-class discrimination but require more training data (Shin et al., 2016; Zhang et al., 2021). The optimal CNN structure is data specific and is hard to derive, and sensitivity analysis approaches, for example, changing the number and spatial dimensions of the kernels in predefined intervals with cross-validation until the best classification accuracy is obtained (Yang et al., 2017; Tan and Le, 2019), are computationally expensive.

The CNN is trained using a large number of patch samples. The training defines the kernel weights and the bias and vector values that collectively are termed the network coefficients. The network coefficients are first initialized randomly. A gradient descent method is used to iteratively update the network coefficients by minimizing a loss function. With each iteration the network coefficients are updated by adding the coefficient gradient values of a loss function perturbed by a small amount (also known as learning rate). The loss function is defined by examination of the difference between the training data land cover labels and the CNN predicted class labels. The gradient values can become explosively large or vanishingly small because of the complexity and inter-dependence of the CNN layers, leading to over-fitting issues (Ruder, 2016) or precluding reliable network coefficient estimation (Glorot and Bengio, 2010). Approaches have been developed to avoid extreme gradient values. He et al. (2015) proposed a network coefficients initialization method by setting the layer kernel and vector weights with random values drawn from a normal distribution with standard deviations related to the number weights in each kernel or the number of values in each vector. Ioffe and Szegedy (2015) proposed a batch normalization process to normalize the feature map and feature vectors using the mean and standard deviation of the feature map and feature vectors derived from the mini-batch of training samples (2015). Regularization of the convolutional layer kernel weights and fully connected layer vector values is required to prevent overfitting (Nowlan and Hinton, 1992). The L2 regularization approach (Neumaier, 1998) is often used where the square root of the sum of all the layer kernel and vector weights is minimized and a single scalar parameter is set to define the relative contribution of L2 regularization minimization with respect to the loss function minimization.

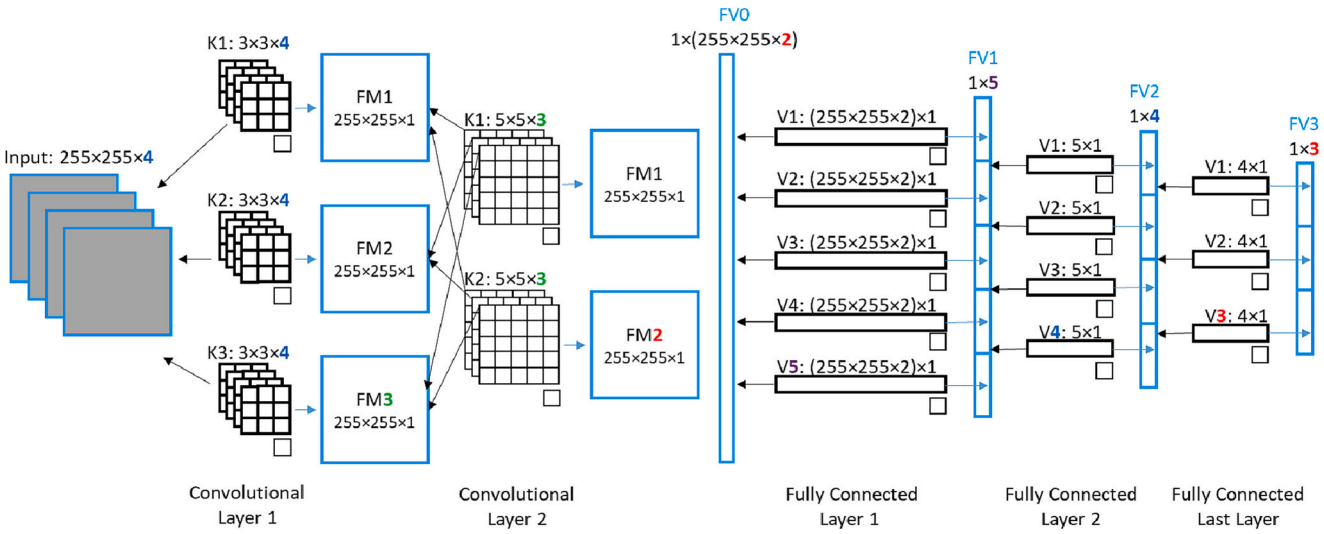

Fig. 4. Example simple CNN structure defined by two convolutional layers followed by three fully connected layers (black boxes) and the generated feature maps (FM) and feature vectors (FV) (blue boxes) used to classify an image patch composed of 255 × 255 pixels and 4 spectral bands into one of three land cover classes. The black and blue arrows show the inputs (i.e., previous layer FM or FV) and outputs (i.e., current layer FM or FV), respectively, of each convolution kernel or vector. The bold and colored numbers indicate the number of feature maps for the convolutional layer and the length of the feature vector for the fully connected layer. Note the structure is simplified for illustrative purposes only, typically the number of convolutional kernels and layers are greater, and in this study far more than four input bands were used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Land cover classification of an individual patch is undertaken as follows. First, each kernel in the first convolutional layer is convolved with the patch $n \times n$ pixel values to generate a two-dimensional feature map. Specifically, for each kernel, a two dimensional spatial convolution is undertaken, whereby the $i \times j \times r = k$ kernel matrix weights are convolved with patch spectral band k, and then the convolution results for each of the r bands are summed together, and the kernel bias term is added. If there are c_1 kernels and associated bias terms in the first convolutional layer then c_1 feature maps are generated. Next, each kernel in the second convolutional layer is convolved with all the c_1 feature maps to generate a two-dimensional $n \times n$ feature map. For each kernel, the $i \times j \times r = c_1$ kernel matrix weights are convolved with the first layer feature map c_1, and then the convolution results for each of the r feature maps are summed together, and the kernel bias term is added to generate a feature map in the second layer. If there are c_2 kernels in the second convolutional layer then c_2 feature maps are generated. This process is cascaded through all the convolution layers. Before the fully connected layers are used, a feature vector with length fv_0 equal to the product of the number and the two side dimensions of the feature maps generated by the last convolution layer is generated. This feature vector is defined by flattening (sometimes termed vectorizing) the two-dimensional feature maps generated by the last convolution layer and concatenating the results together into a single vector. The first fully connected layer is composed of a set of fc_1 vectors that each are length fv_0 (sometime termed the fully connected layer weight matrix (fc_1) × (fv_0)) and have an associated bias term (the fully connected layer bias vector (fc_1) × 1). The first fully connected layer is used to generate a feature vector, length fc_1, (the vector elements are often called neurons) and each value in the future vector is derived as the dot product of the flattened feature vector values (length fv_0) and one of the fc_1 vectors (length fv_0) with the bias terms added. In CNN this is usually implemented using matrix calculation to derive all the neuron values in parallel. The second fully connected layer is composed of fc_2 vectors that each have length fc_1 and have an associated bias term and is used to generate the second layer feature vector with length fc_2. This process is cascaded through all the fully connected layers. The last fully connected layer is composed of as many vectors, with associated bias terms, as there are land cover classes, and the last derived feature vector has a length equal to the number of classes. Each feature map and vector value is nonlinearly transformed before fed into next layer and the rectified linear unit (ReLU) function is a commonly used transform by setting negative values to zero and leaving positive values unchanged (Glorot et al., 2011). The classified land cover class is derived by application of a different nonlinear function, i.e., the normalized exponential function, termed softmax function, to the last feature vector. The land cover class is assigned to the patch center pixel. Consequently, for satellite image CNN classification, the input patch usually has odd side dimensions, e.g., 255 × 255 as in Fig. 4. In order to classify an image that has spatial dimensions greater than a patch, the above process is undertaken sliding the patch spatially one pixel at a time across the image and independently classifying each patch.

3.2. *The single pixel time series based CNN land cover classification methodology*

Rather than classify patches extracted from single images, as described above, single pixel temporal metrics are classified. Specifically, at each CONUS 30 m ARD pixel location, a 1D CNN is applied to an $s \times n_p$ array where $s = 13$ is composed of the five Landsat surface NBAR bands and the eight NBAR ratios (Section 2.2), and n_p (=5, 7 or 9) is the number of percentiles of each of the five NBAR and eight NBAR ratio values. Fig. 5 shows a hypothetical example CNN structure (black boxes) and the feature maps and vectors (blue boxes) for $n_p = 7$. For illustrative purposes Fig. 5 shows only 2 convolutional layers and 3 fully connected layers. The input $s = 13 \times n_p = 7$ predictor array is read by the first convolution layer composed of m_1 3 × 7 kernels and m_1 biases to generate m_1 feature maps. The 3 × 7 kernels are applied to the 13 × 7 array of predictor variables and not to a spatial patch. In this way a single kernel can capture information from the 7 percentile variables derived over the growing season. The m_1 feature maps generated by the first convolutional layer are read by the second convolution layer composed of m_2 3 × m_1 kernels and m_2 biases to generate m_2 feature maps. This process is cascaded through all the convolution layers. The last convolutional layer feature maps are flattened and concatenated into a single feature vector with length fv_0. The dot product of the length fv_0 vector and each of the fc_1 vectors (length fv_0) in the first fully connected layer are summed with each of the fc_1 biases in the first fully connected layer to derive the first fully connected layer feature vector

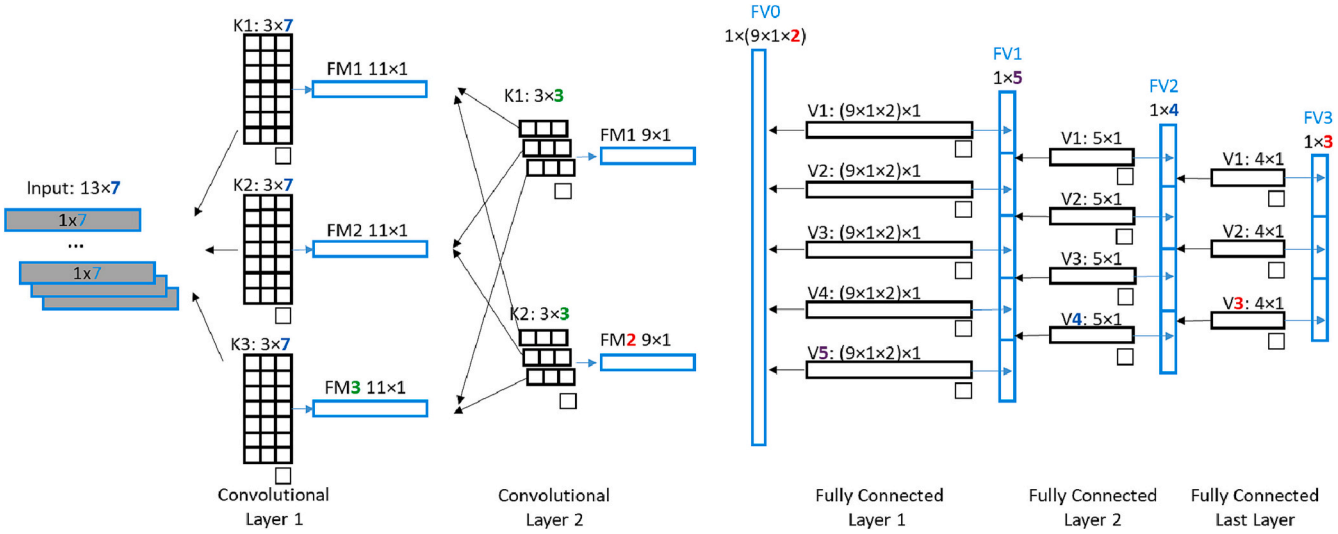

Fig. 5. Example simple 1D CNN structure defined by two convolutional layers followed by three fully connected layers (black boxes) and the generated feature maps (FM) and feature vectors (FV) (blue boxes) used to classify a pixel time series composed of 13 × 7 predictor array into one of three land cover classes. Note that in this example, zero padding is not used and the first layer feature map dimensions (11) are smaller than the input array (13). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with length fc_1. This process is cascaded through all the fully connected layers. As for conventional CNN the last fully connected layer is composed of as many vectors, with associated bias terms, as there are land cover classes. Although a single kernel cannot capture the relationships between the 13 different NBAR and NBAR ratio values, these relationships can be extracted by the kernels in the other convolutional layers and by the first fully connected layer. For example, in Fig. 5 it is evident that the first layer of 3 × 7 kernels can only capture relationships among 3 local neighbor values of all the 13 different NBAR and NBAR ratio predictors. However, the second layer of 3 × 5 kernels can capture relationships among 5 local neighbor values.

The CNN structure dictates the complexity and performance of the classification; however, definition of the optimal structure is data specific and hard to optimize. In this study, two different structures were examined with 5 and 8 layers, and approximately 0.2 and 2.1 million learnable network coefficients, respectively. The 5-layer model is based on the one reported by Pelletier et al. (2019), with 3 convolutional layers each composed of 64 kernels, and 2 fully connected layers composed of 256 vectors and 15 vectors (i.e., the number of NLCD classes). The 8-layer model is a more complex variant, with 6 convolutional layers composed of 64, 128, 256, 256, 512, and 512 kernels, and 2 fully connected layers composed of 1024 vectors and 15 vectors. For both models the kernels with 3 spatial dimensions were used. A total of 6 convolutional layers is used because it is the maximum meaningful number given 13 input predictors (13 different NBAR and NBAR ratio values). Note that the feature map length is reduced by 2 after a three dimensional kernel convolution (e.g., 13 is reduced to 11 after the first convolution in Fig. 5). The feature map length will become 1 after cascading through 6 convolutional layers and so no further convolutions are warranted. For deeper CNN, it is conventional to increase the number of the convolutional kernels for deeper layers (Simonyan and Zisserman, 2014; Tan and Le, 2019).

CNN deep learning has several implementation parameterizations. It is beyond the scope of this paper to describe them in detail but notable ones are overviewed here. The feature maps generated by each kernel convolution with patch-based CNN have undefined values along their boundaries (the undefined boundary width is half the kernel spatial dimension) that can be filled with zero values so that the feature maps have the same spatial dimensions as the input patch, this is usually termed "zero padding" (Sideris and Li, 1993). Zero padding of the $s \times n_p$ arrays was not used as we found no classification improvement. Another strategy is max pooling (Boureau et al., 2010). In this approach the side dimensions of the feature maps are reduced by a factor of $p = 2$ with a filter that calculates the maximum value of each $p \times p$ adjacent region of the feature map to create a downsampled (pooled) feature map. The pooled feature maps are then read by the next convolution layer. Max pooling was not used however because the low dimensionality of the input $s \times n_p$ array imposes limitations on the reduction of the feature map dimensions across the network layers. Skip connection is a strategy developed by He et al. (2016) and implemented by feeding the output of one convolutional layer not only to the following convolution layer (Figs. 4 and 5) but also to deeper convolutional layers. The strategy can make deep CNN easier to train (Radosavovic et al., 2020; Liu et al., 2022b). We found, however, no classification improvement when using skip connections, presumably because the 8- and 5-layer 1D CNNs have sufficient complexity to capture land cover class differences in the $s \times n_p$ arrays.

CNN training requires considerable processing and memory requirements and several strategies have been developed to reduce computational constraints. Notably, the mini-batch gradient descent method (Bottou, 2010) divides the training data randomly without replacement into smaller subsets (each is termed a mini-batch). The CNN is trained with each mini-batch sequentially, i.e., the samples in a single mini-batch are used for each iteration of the network coefficient updates. In this study, the mini-batch gradient descent method was used with 256 training samples per batch; we found that using fewer (128) or more (512) training samples per batch provided negligible classification differences. The stochastic gradient descent (SGD) optimizer, parameterized with a commonly used 0.9 momentum value, was used to reduce gradient oscillations among successive mini-batches due to training sample differences among mini-batches (Qian, 1999). The learning rate, i.e., a scalar value from 0 to 1 to perturb the coefficient gradient values calculated for each mini-batch is an important parameter in gradient descent training (Ruder, 2016). In general, a large learning rate will provide fast convergence of the loss function value to the global minimum. However, this can make the loss function value oscillate, or diverge, rather than converge, with subsequent iterations. In contrast, a small learning rate could make the training process unnecessarily time-consuming. Consequently, in this study, a dynamic learning rate was used, whereby initial iterations used a relatively large learning rate that was reduced, e.g., by a factor of ten, whenever the loss function value stopped converging. The dynamic learning rate was used with initial

learning rate 0.01. Other initial learning rates (0.1, 0.001 and 0.0001) and alternative learning rate adjustment methods, such as the commonly used Adam optimizer (Kingma and Ba, 2014), provided negligible classification differences. An epoch of iterations is completed when all the training samples are used. Typically, tens of epochs are needed, i.e., each sample is used to update the network coefficients tens of times. The epoch number was set to 70 in this study similar to other studies (Simonyan and Zisserman, 2014; Martins et al., 2022) and because we observed negligible accuracy improvement using more epochs.

In order to reduce imprecise storage of very large or very small numerical values in the gradient descent training process the training data predictor variables are conventionally normalized so that each variable has zero mean and one standard deviation after normalization (LeCun et al., 2012). This was undertaken for each predictor of the $s \times n_p$ arrays by subtracting the training sample mean from the predictor values and then divided by the training sample standard deviation. Similarly, batch normalization (see Section 3.1) to normalize the feature map/vector values was used as it can mitigate the vanishing gradient issue (Ioffe and Szegedy, 2015). The L2 regularization (see Section 3.1) was used to prevent overfitting. The relative contribution of the L2 regularization minimization with respect to the loss function minimization, defined by the single scalar parameter (λ), can also be very influential on classification accuracy (Bilgic et al., 2014). Therefore, the λ parameter value was set to 0.001 after examining four different values, specifically 0 (i.e., no L2 regularization), 0.01, 0.001, and 0.0001. The optimal training and structure parameters and the iterations to change the learning rate were determined based on randomly selecting 4% of the training samples (sometimes termed validation samples) following the procedure described in He et al. (2016).

3.3. Land cover classification experiments

Different CONUS land cover 30 m classifications were undertaken for 2011 using the 10%, 50% and 90% training proportions, the temporal metrics derived with $n_p = 5$, 7 and 9 percentiles, and with the 5-layer and 8-layer 1D CNN architectures. Given that there should be as least as many quality filtered observations as n_p, a classification with $n_p = 9$ was generated and evaluated only at CONUS ARD pixel locations with $\geq$9 quality filtered growing season observations. Two classifications with $n_p = 7$ were generated and evaluated i.e., considering CONUS ARD pixel locations with $\geq$7 and, then considering locations with $\geq$9, quality filtered growing season observations. Similarly, three classifications with $n_p = 5$ were generated and evaluated, i.e., considering only CONUS ARD pixel locations with $\geq$5, $\geq$7 and $\geq$ 9 quality filtered growing season observations. Our expectation was that the 90% training proportion and 8-layer 1D CNN architecture would provide the greatest overall classification accuracy. However, although using a greater number of percentiles (i.e., $n_p = 9$) should better capture seasonal reflectance variations and provide higher classification accuracy, this may not be the case for certain land covers. For example, the "cultivated crop" and "developed" classes have complex temporal signatures and significant within class variation across the CONUS (Zhang and Roy, 2017; Roy and Yan, 2020; Sun et al., 2021; Zhou et al., 2020) and so may be better classified using fewer percentiles. To help examine this, classifications were also undertaken with temporal metrics derived with just $n_p = 1$ percentiles defined by the 50th percentile. The 50th percentile is the median value over the growing season and so does not capture temporal surface variations. Despite this, median value composites, derived by taking the median of the growing season reflectance band and band ratios values at each pixel location, have been used for land cover classification (Maxwell and Sylvester, 2012; Hermosilla et al., 2018). We note that researchers have also used monthly composites for land cover classification (Griffiths et al., 2019; Tran et al., 2022) although gaps due to unobserved locations and cloud obscuration become more frequent over smaller time periods (Lindquist et al., 2008; Roy et al., 2010). Therefore, 1D CNN classifications were also undertaken using seven monthly growing season composites. Each composite was derived by selecting the median of the quality filtered observations acquired over the month for each of the five Landsat bands and for each of the eight normalized ratios, and linearly interpolating between months to fill any gaps.

To provide a benchmark, the 1D CNN land cover classifications were compared with CONUS classifications derived by random forest classification of the same predictor variables and using the same training and evaluation data. The random forest classifier is a non-parametric ensemble form of decision tree classifier with each tree grown using a random subset of training data and randomly selected predictor variables to avoid over-fitting (Breiman, 2001). Random forest has been extensively used for large area land cover classification (Zhang and Roy, 2017; Hermosilla et al., 2022; Liu et al., 2021). In this study the random forest was run with default parameter settings (Liaw and Wiener, 2002), specifically, with a total of 500 trees and with each tree considering 63.2% of the training data selected randomly with replacement and with 9 predictor variables selected randomly from the $13 \times n_p$ predictor variables.

3.4. Land cover classification accuracy and quality assessment

The CONUS 2011 land cover classification accuracies were quantified by comparison with the 2011 evaluation data. Recall that the evaluation data were selected considering 10% of the 3.3 million CONUS samples (Table 1) not used in the training and were selected randomly from those ARD 30 m pixel locations where there were at least as many growing Landsat season observations as the number of percentiles (n_p) used to generate the classification. This provided an evaluation data set defined at >325,000 CONUS 30 m pixel locations (Table 2). The CONUS land cover classification accuracies were quantified in the conventional manner by counting the correspondence of the NLCD class values at the evaluation 30 m pixel locations with the classified class values to populate a two-dimensional confusion matrix composed of 15 land cover classes. Land cover class specific producer's and user's accuracies, sometimes referred to as the precision and recall, respectively, and the F1-score that is the harmonic mean of the user's and producer's accuracies were then derived from the confusion matrix (Congalton and Green, 2019). The overall accuracy was also extracted from the confusion matrix and was used as a diagnostic accuracy metric to compare the different classification results. The overall accuracy was derived 7 times for each classification to check for the undue influence of randomness in the training steps. Notably, the CNN has randomness in the network coefficient initialization and mini-batch gradient descent training (Scardapane and Wang, 2017) and the random forest has randomness in the tree sample selection and predictor selection for each tree branch (Breiman, 2001). To examine this, the training was undertaken using different pseudo-random number generation initializations and the mean and standard deviation of the 7 overall classification accuracies were examined. As the standard deviations of the 7 overall classifications were found to be very small ($\leq$0.03%) this 7-fold experiment was not undertaken for the class specific accuracy analysis.

It is well established that large area land cover classifications may contain quality issues, such as stripes at input image boundaries or anomalous spatial patterns, that may not be revealed by the accuracy assessment results that necessarily rely on a limited sample of evaluation data (Boschetti et al., 2019). Therefore, the 30 m CONUS land cover classifications were quality assessed by visual comparison with the NLCD 2011 CONUS land cover map. In addition, the percentage of CONUS ARD land pixels that were classified as each class by the NLCD and the 1D CNN were compared to provide a synoptic assessment of their consistency. Detailed visual comparisons were undertaken at three Landsat ARD 5000 × 5000 30 m tiles that were examined to ascertain if the land cover of spatially fragmented and isolated pixels were preserved, which, as mentioned earlier, can be an issue for conventional

patch-based CNN classifiers and should not be an issue for single pixel time series based 1D CNN classification. The three ARD tiles were selected at location that we have examined in other papers (Yan and Roy, 2020; Zhai et al., 2022) and that encompass a mix of land cover types, predominantly, agriculture in South Dakota, shrubland and urban in Arizona, and wetland, water and urban in Florida.

3.5. Signature extension demonstration - application of the 2011 trained model to generate CONUS year 2006 land cover classification results

To further demonstrate the approach, the 1D CNN trained with the 2011 Landsat ARD was applied to the seven month growing season of year 2006 Landsat 5 TM and Landsat 7 ETM+ ARD. In the remote sensing literature, the application of a trained model to classify different geographic regions or time periods not used for the training is known as signature extension (Woodcock et al., 2001; Gray and Song, 2013). In general, signature extension becomes less appropriate with greater space and/or time separation (Zhang and Roy, 2017). However, the relatively short period between 2006 and 2011, the use of the same Landsat sensors for these two years, and the same CONUS geographic area, means that this is a meaningful demonstration of the 1D CNN capability, although satellite data acquisition and surface differences between the two years will reduce the 2006 classification accuracy. The percentage of CONUS ARD land pixels that were classified as each class by the 2006 classification and by NLCD 2006 were compared to provide a synoptic assessment of their consistency. The 2006 land cover classification accuracy was quantified by comparison with the 2006 evaluation samples (described at the end of Section 2.2).

4. Results

4.1. CONUS land cover classification model inter-comparison experiment results

Fig. 6 shows the CONUS overall land cover classification accuracies derived with $n_p = 9$ percentiles (i.e., 13 × 9 predictor variables) for the 8-layer 1D CNN (open dots), 5-layer 1D CNN (open diamonds) and random forest (stars), trained with the 10%, 50% and 90% training proportions (about 0.3, 1.6 and 2.9 million CONUS samples, Table 2). As there should be as least as many quality filtered observations as n_p, the classifications were generated and evaluated considering only CONUS ARD pixel locations with ≥9 quality filtered growing season observations. As expected, the overall classification accuracies increased with the training sample size with the highest accuracies for the 90% training proportion. The random forest provided consistently the lowest overall classification accuracy, and, for example, with the 90% training proportion provided an 83.6% accuracy. The two 1D CNN structures trained with the 90% proportion provided 86.1% and 86.4% accuracies for the 5-layer and 8-layer structures, respectively. Notably, the 1D CNN improvement over random forest was larger than the accuracy differences between the two CNN structures. The 8-layer CNN had 0.7% and 0.6% higher overall classification accuracy than the 5-layer CNN for the 90% and 50% training proportions, respectively, whereas the 5-layer CNN had 0.1% higher accuracy for the 10% training proportion.

Fig. 7 shows the CONUS overall land cover classification accuracies derived as Fig. 6 but using one growing season median value composite defined by the 50th percentile (median value) (black) and using seven monthly median value composites derived over the growing season (magenta). Similar results as Fig. 6 are apparent, i.e., the overall classification accuracies increase with the training sample size, the highest accuracies are for the 90% training proportion, and the random forest provides systematically lower accuracies than the 1D CNN. Notably, however, the $n_p = 1$ (Fig. 7, black) classification accuracies are

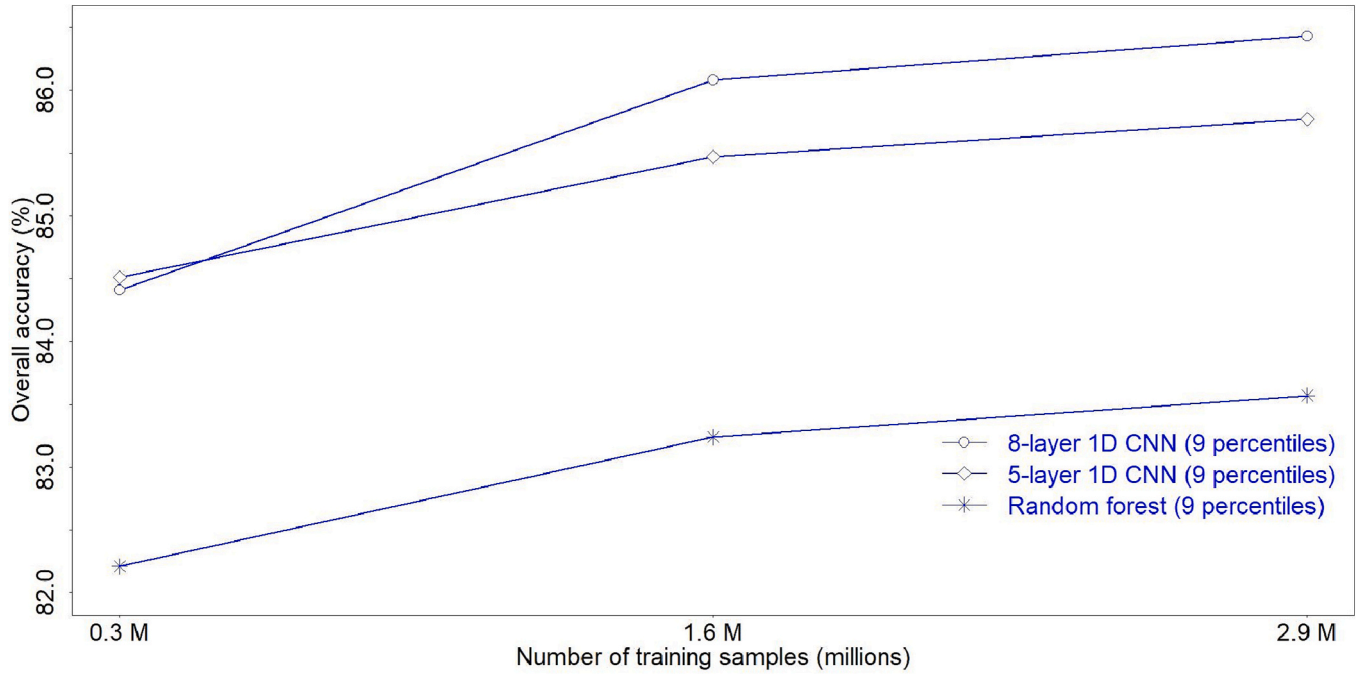

Fig. 6. Overall CONUS land cover classification accuracies derived with $n_p = 9$ percentiles (i.e., 13 × 9 predictor variables) for the 8-layer and 5-layer 1D CNN structures and random forest, each trained with the same 10%, 50% and 90% training proportions (about 0.3, 1.6 and 2.9 million CONUS samples), respectively, and evaluated using an independent 10% evaluation proportion (about 0.3 million CONUS samples) (Table 2). The classifications were derived and evaluated considering only CONUS ARD pixel locations with ≥9 quality filtered observations. The overall accuracy was derived 7 times for each classification to check for the undue influence of randomness in the training, and the symbols show the mean overall accuracy values (the standard deviation of each set of 7 results was negligible and less than the plotted symbol dimensions).

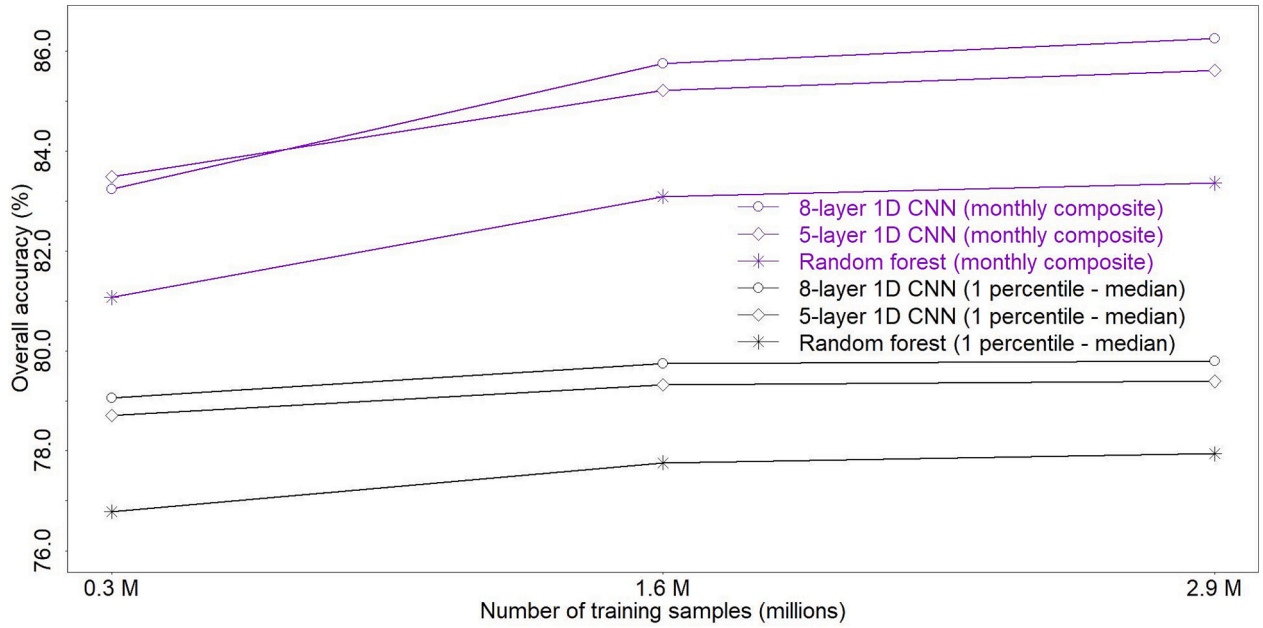

Fig. 7. As Fig. 6 but generated using one growing season median value composite defined by the 50th percentile ($n_p = 1$) (black) and using seven monthly median value composites, i.e., April, May …, October (magenta). The classifications were derived and evaluated considering only CONUS ARD pixel locations with ≥9 quality filtered observations. The overall accuracy was derived 7 times for each classification to check for the undue influence of randomness in the training, and the symbols show the mean overall accuracy values (the standard deviation of each set of 7 results was negligible and less than the plotted symbol dimensions). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

significantly and systematically lower than the $n_p = 9$ (Fig. 6) accuracies by at least 5.35% (8-layer CNN using 10% training samples) and up to 6.53% (8-layer CNN using 90% training samples). Similarly, the monthly composite (Fig. 7, magenta) classification accuracies are systematically lower than the $n_p = 9$ (Fig. 6) classification accuracies by at least 0.18% (8-layer CNN using 90% training samples) and up to 1.17% (8-layer CNN using 10% training samples). These differences illustrate the utility of using many percentiles to better capture seasonal reflectance variations.

In both Figs. 6 and 7 the overall accuracies were derived 7 times for each classification to check for the undue influence of randomness in the training, and the plotted symbols show the mean overall accuracy values

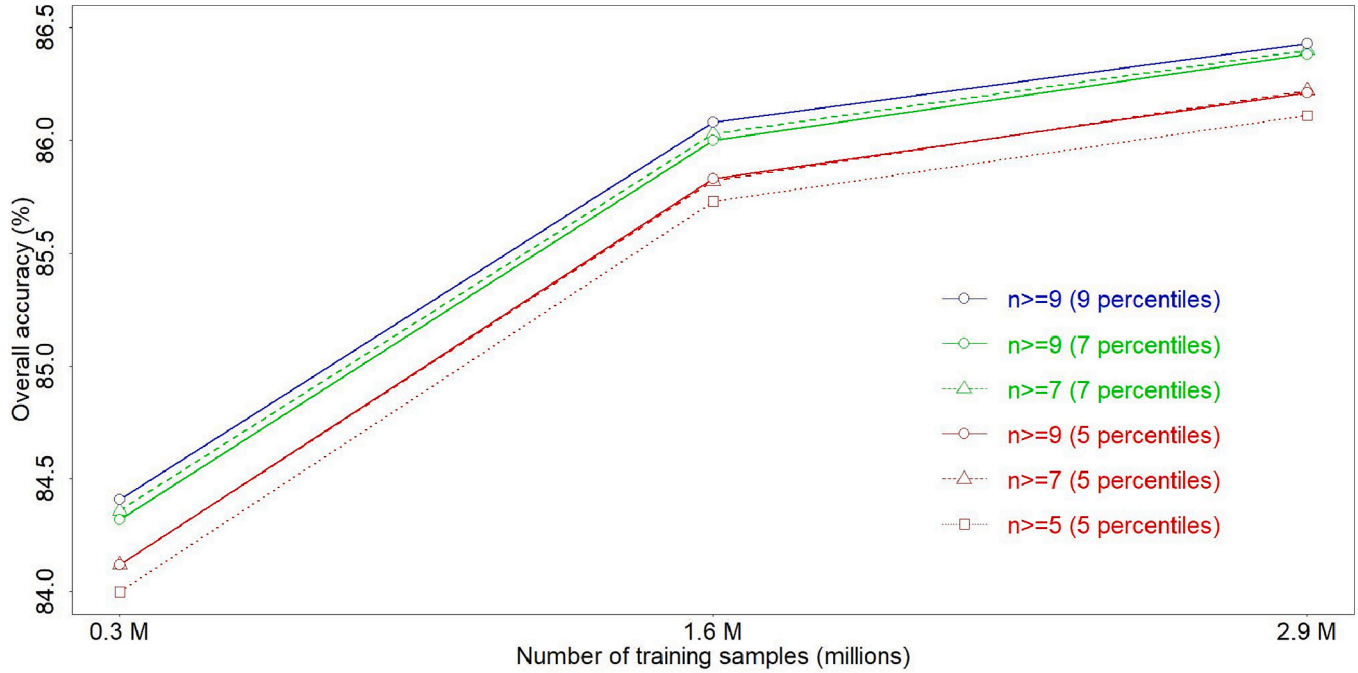

Fig. 8. Overall CONUS land cover classification accuracies derived with the 8-layer 1D CNN structure and generated using temporal metrics derived with $n_p = 5$, 7, or 9 percentiles and considering CONUS ARD pixel locations with ≥n quality filtered growing season observations. The classifications were trained using 10%, 50% and 90% training proportions (about 0.3, 1.6 and 2.9 million CONUS samples), and evaluated using an independent 10% evaluation proportion (about 0.3 million CONUS samples) (Table 2). The overall accuracy was derived 7 times for each classification to check for the undue influence of randomness in the training, and the symbols show the mean overall accuracy values.

derived from the 7 sets of overall accuracy results. The standard deviations of the 7 overall accuracies are not shown as their values are too small to be meaningfully plotted. The 50% and 90% training samples experiments had standard deviations ≤0.02%. The 10% training sample experiments had slightly greater standard deviations (≤0.03%) and this is likely due to the smaller number of training samples used (Table 2). Given these findings, the randomness standard derivation results are not discussed in the following figures.

On the basis of the Fig. 6 and 7 results, the 8-layer 1D CNN with $n_p >$ 1 was used for the remainder of this study. Fig. 8 shows the overall CONUS land cover classification accuracies provided by the 8-layer 1D CNN trained using 10%, 50% and 90% training proportions (about 0.3, 1.6 and 2.9 million CONUS samples, Table 2) and derived using temporal metrics with $n_p = 5$, 7 and 9 percentiles (shown by red, green, blue colors). Given that there should be as least as many quality filtered observations as n_p, the classification with $n_p = 9$ was undertaken and evaluated at CONUS ARD pixel locations with $n \geq 9$ quality filtered growing season observations. The classification with $n_p = 7$ were undertaken and evaluated twice – first at CONUS ARD pixel locations with $n \geq 7$ and, then again at locations with $n \geq 9$, quality filtered growing season observations. Similarly, the classifications with $n_p = 5$ were generated and evaluated three times i.e., considering only CONUS ARD pixel locations with $n \geq 5$, ≥7, and ≥ 9 quality filtered growing season observations. In all cases, the overall classification accuracies increased with the training sample size, and the 9 and 5 percentile classifications had the highest and lowest overall accuracies, respectively. Considering the 90% training proportion results, the highest accuracies were obtained for $n_p = 9$ (86.43%, $n \geq 9$), with marginally lower accuracies for $n_p = 7$ (86.38% for $n \geq 9$, and 86.40%, for $n \geq 7$) and the $n_p = 5$ results always had the lowest accuracies and were lower accuracy than $n_p = 9$ by 0.17% to 0.29%. The greatest accuracy differences among the classifications occurred for the 10% training proportion results, likely reflecting that too few training data were used. On the basis of these results, the 8-layer 1D CNN with $n_p = 7$ and $n_p = 9$ were used for the remainder of this study.

4.2. Mapped CONUS land cover classification results

Fig. 9 shows the CONUS 8-layer 1D CNN land cover classification derived using the 90% training proportion and $n_p = 7$ percentiles (i.e., 13 × 7 predictor variables). The CONUS ARD encompasses more than 150 thousand and 88 thousand 30 m pixels east-west and north-south respectively. Therefore, to visualize the results, Fig. 9 shows the majority land cover class in adjacent non-overlapping 50 × 50 30 m pixel (i.e., 1.5 × 1.5 km) regions. Consequently, the small minority of CONUS pixels with ≤7 growing season quality filtered observations (0.57%, Fig. 2) are not apparent. At this synoptic scale the 8-layer 1D CNN land cover classification derived with $n_p = 9$ percentiles appears the same, and so is not shown. The classification shows no stripes or anomalous spatial patterns and appears quite plausible. The spatial variation in land cover across the CONUS exhibits generally expected geographic differences with, for example, shrub/scrub in the dry south-west; cultivated

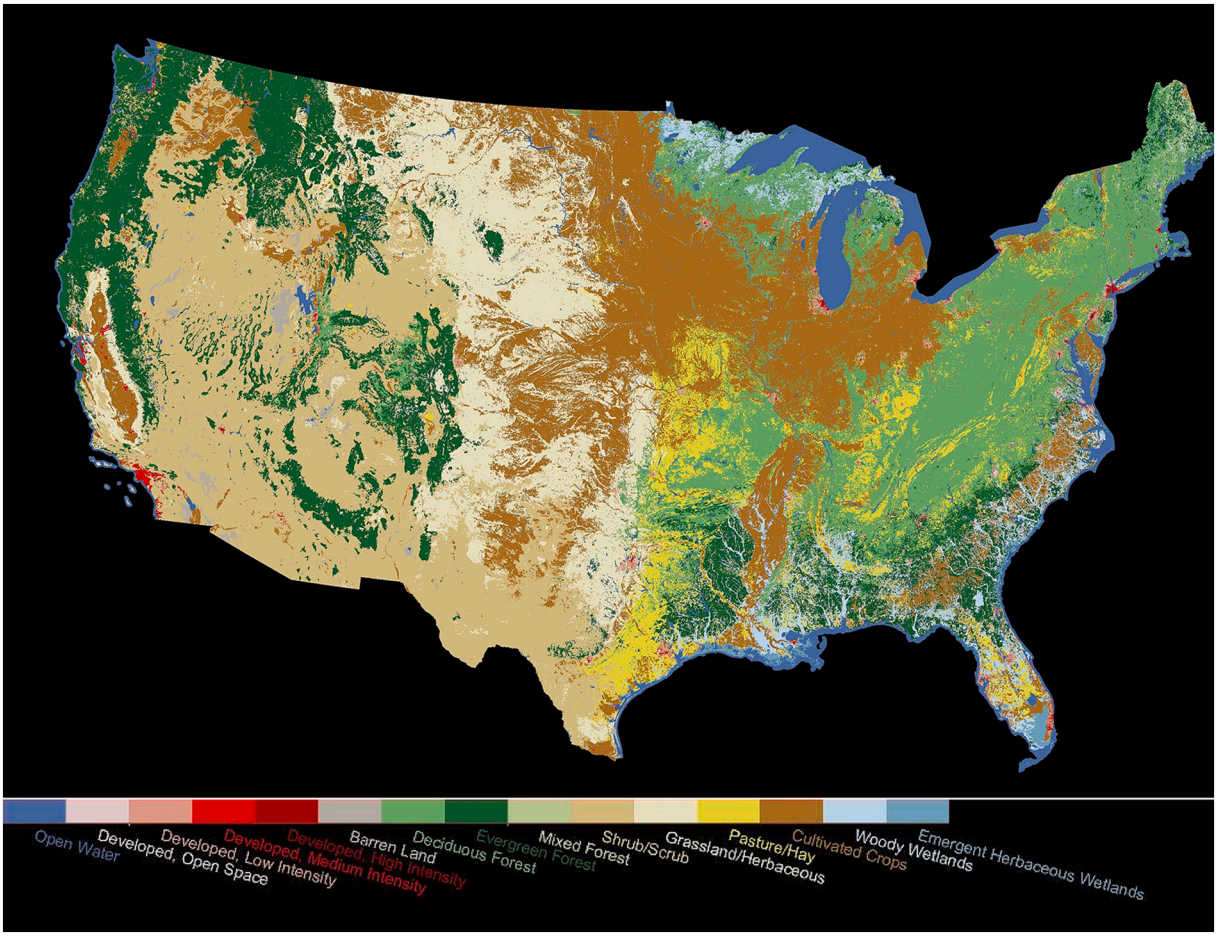

Fig. 9. The 15-class CONUS 2011 land cover map derived with the 8-layer 1D CNN structure and $n_p = 7$ percentiles (i.e., 13 × 7 predictor variables) trained with the 90% training proportion (about 2.9 million CONUS samples, Table 2). The classification was derived at CONUS 30 m pixel locations with ≥7 growing season quality filtered observations (i.e., covering 99.43% of the CONUS, Fig. 2). For visualization purposes, the majority land cover class in adjacent non-overlapping 50 × 50 30 m pixel regions is shown.

crops in the Great Plains and the Mississippi watershed and California interior; open water over the Great Lakes near the Canadian border; evergreen forest dominating in the north-west and deciduous forest in the eastern states; and developed medium-intensity and developed high-intensity classes evident in major urban areas such a Los Angeles on the south-west coast and New York on the north-east coast.

Fig. 10 shows for comparison the CONUS NLCD 2011 illustrated at the same scale as Fig. 9. The spatial variation in the land cover classes is similar between the two land cover classifications. However, two apparent differences are evident: (i) more cultivated crops but less pasture/hay in the agricultural heartland between the Missouri and Mississippi rivers is apparent in the 8-layer 1D CNN classification (Fig. 9) compared to the NLCD (Fig. 10); and (ii) less developed open-space around cities including Houston, Phoenix, Chicago, and Detroit, is apparent in the 8-layer 1D CNN classification compared to the NLCD. Note that only 0.0178% of the CONUS land pixels were classified as "perennial ice/snow" in the 2011 NLCD and this class, although present, is not apparent in Fig. 10.

Fig. 11 shows a scatterplot comparing the CONUS land cover class percentages extracted from the 7 percentile 8-layer 1D CNN classification (y-axis) and from the NLCD (x-axis). The 0.0178% of CONUS 30 m pixel locations that were classified as "perennial ice/snow" in the 2011 NLCD classification were discarded from this comparison, thus Fig. 11 shows results for 15 of the 16 NLCD classes (Table 1). The class percentages between the two classifications are similar and have a 0.99 correlation. The scatterplot comparing the CONUS land cover class percentages for the 9 percentile 8-layer 1D CNN land cover classification and the NLCD is not shown but was similar with a 0.98 correlation. The greatest relative class percentage difference was for the developed open space class (3.3% in the NLCD and 0.4% in the 1D CNN) and is quite evident, as noted above, when comparing Figs. 9 and 10 (light red colors) around cities, such as Houston, Phoenix, Chicago, and Detroit. The mixed forest class percentages were also quite different (2.0% in the NLCD and 0.6% in the 1D CNN) but this is hard to visually assess due to the spatial arrangement of the mixed forest classified pixels and the Figs. 9 and 10 synoptic image scale. The greatest absolute class percentage was for the cultivated crop class that was 15.5% in the NLCD and 19.2% in the 1D CNN. These differences are particularly apparent in the detailed ARD tile results and in CONUS class specific user's and producer's accuracies that are presented below.

Figs. 12-14 show the NLCD and 8-layer 1D CNN land cover classifications for the three 5000 × 5000 30 m ARD tiles that are located over an agricultural area in central South Dakota (Fig. 12), a shrub, forest and urban mixed area around Phoenix, Arizona (Fig. 13), and a wetland and urban mixed area around Miami, Florida (Fig. 14). True color images are also shown to provide geographic context. Both the 7-percentile and 9-percentile 1D CNN classification results are illustrated with little difference between them except that the 9-percentile classifications have fewer classified pixels. This is particularly apparent for the Florida tile (Fig. 14) with fewer available growing season observations that is also clearly apparent in Fig. 3 relative to the rest of the CONUS. Notably, the 1D CNN classification land cover class boundaries are not smoothed and the boundary detail is preserved for features with small axis dimensions greater than a 30 m pixel, including roads and buildings (Figs. 13 and 14), and lakes and rivers (Figs. 12 and 13). This is because, as discussed earlier, the 1D CNN is applied to single ARD pixel temporal metrics

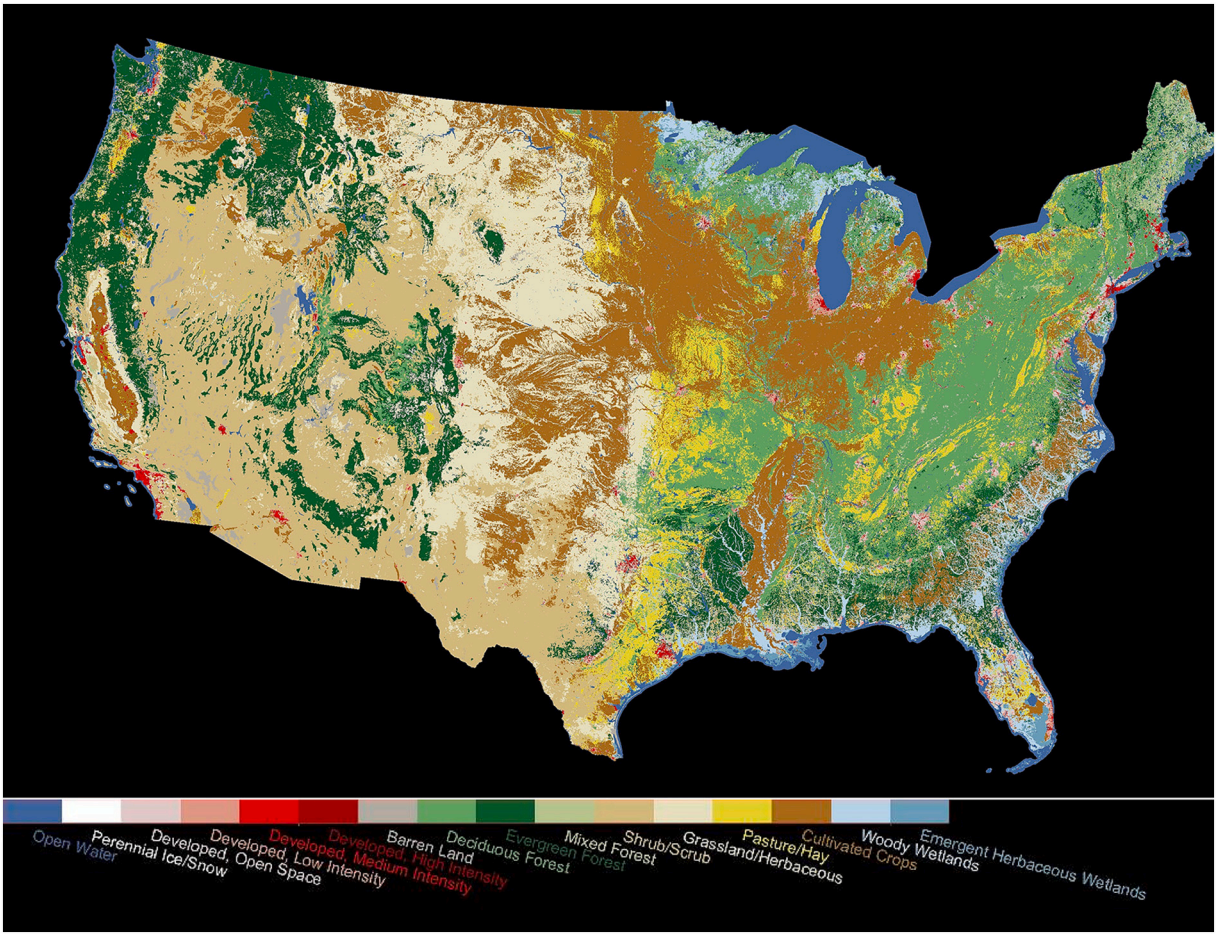

Fig. 10. The NLCD 16-class 2011 land cover map (Homer et al., 2015). For visualization purposes, the majority land cover class in adjacent non-overlapping 50 × 50 30 m pixel regions is shown in the same way as for Fig. 9.

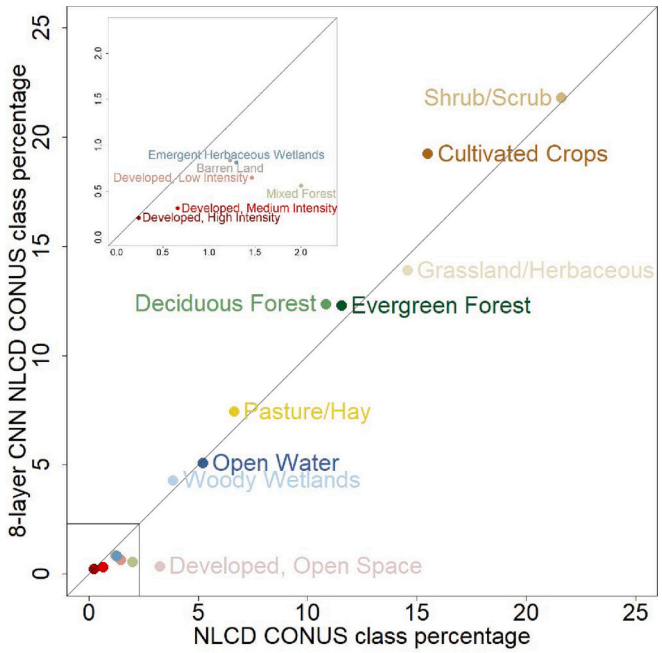

Fig. 11. Scatterplot comparing the percentage of CONUS year 2011 30 m ARD land pixels classified into each of 15 classes (Table 1) by the 8-layer 1D CNN (Fig. 9) and by the 2011 NLCD (Fig. 10). The percentages were derived considering only the CONUS 30 m pixel locations that were not classified as "perennial ice/snow" in the 2011 NLCD classification, as this class was not classified in the 8-layer 1D CNN model. The 1:1 line is shown for reference.

rather than to spatial patches. There is a high level of agreement between the NLCD and 1D CNN land cover classifications apparent in Figs. 12-14 except over certain urban/suburban and agricultural areas. The discrepancies between the NLCD and 1D CNN land cover classifications for the developed open-space class evident at CONUS scale (Figs. 9-11) are particularly apparent and about half of the developed open-space pixels in the NLCD are classified by the 1D CNN as shrub/scrub around Phoenix (Fig. 13) or as cultivated crops to the south and north of Miami (Fig. 14). There are also pronounced NLCD and 1D CNN classification differences between the distribution of pasture/hay and cultivated crops in the South Dakota tile (Fig. 12) that was also evident as noted above at CONUS scale.

4.3. Quantification of class-specific CONUS land cover classification accuracy

Fig. 15 shows the producer's (red shades) and user's (blue shades) accuracies of the 8-layer 1D CNN that was trained using the 90% training proportion. The accuracies are shown for the CONUS classifications derived using $n_p = 7$ percentiles at locations with $n \geq 7$ and $n \geq 9$ quality filtered growing season observations and for the classification derived using $n_p = 9$ percentiles at locations with $n \geq 9$ quality filtered growing season observations. The results for these three classifications are shown as they had the highest overall classification accuracies (Fig. 8). For all three classifications, the open water (11), barren land (31), deciduous forest (41), evergreen forest (42), shrub/scrub (52), grassland/herbaceous (71) and cultivated crops (82) had high (>80%) user's and producer's accuracies. This is likely because these classes have distinct growing season reflectance variations. For example, the open water class (11) had the highest producer's and user's accuracies (>98%) likely due to the characteristically low and relatively temporally unchanging reflectance of open water relative to land surfaces (Pahlevan et al., 2019; Zhai et al., 2022). The developed open-space (21) class had the lowest accuracy with 14.18% - 15.55% producer's and 46.27% - 47.89% user's accuracies, indicating that this class occurred more rarely in the CONUS 1D CNN land cover classifications than it should. This is consistent with the findings commented on above concerning Figs. 9, 13, and 14. The three other developed land classes (22–24) had intermediate producer's and user's accuracies that increase with development intensity in the range 51.24% - 58.99% (developed low-intensity, class 22), to 61.50–66.49% (developed medium-intensity, class 23), to 61.92% - 71.05% (developed high-intensity, class 24). This is likely because impermeable surfaces (concrete, asphalt etc.) have temporally more consistent reflectance than vegetated surfaces (Small, 2002; Schug et al., 2020) and a spatially greater proportion of impermeable surfaces is present with increasing development intensity as defined by the NLCD classification scheme (Yang et al., 2003). Notably, the mixed forest class (43) producer's accuracies were about half the user's accuracies, indicating that this class occurred more rarely in the CONUS 1D CNN land cover classifications than it should.

There was no systematic pattern in the class specific accuracy results with respect to the number (n) of quality filtered growing season observations for the two $n_p = 7$ classifications (Fig. 15). For the 11 non-developed classes, the difference in the producer's accuracies comparing $n \geq 7$ and $n \geq 9$ was <2.5% and the difference in the user's accuracies was <2.0%. For the four developed classes the difference in the producer's accuracies was <2.5% and the difference in the user's accuracies was <3.0%. The only large difference between the $n \geq 7$ and $n \geq 9$ results was for the developed high-intensity class (class 24) that had a 6.3% higher user's accuracy for $n \geq 9$ than $n \geq 7$. This indicates that fewer of the other classes are misclassified as developed high-intensity class for $n \geq 9$ than for $n \geq 7$; this is discussed in the next paragraph.

There was also no systematic pattern in the class specific accuracy results with respect to the number of percentiles ($n_p = 7$ or $n_p = 9$) that were classified (Fig. 15). The only class specific accuracy differences between the $n_p = 7$ and $n_p = 9$ classifications that were not small were for the developed classes. The developed open-space (21) and developed low-intensity (22) classes had $n_p = 9$ producer's and user's accuracies that were up to 3.6% lower than the $n_p = 7$ classifications. This is likely because these classes have a multitude of land cover and land uses, with significant within class variation across the CONUS, and so geographically variable and complex temporal signatures that are better generalized in the 1D CNN using fewer percentiles. Conversely, the developed high-intensity class (24) had 0.0% - 0.6% and 2.9% - 9.2% higher producer's and user's accuracies, respectively, than the two $n_p = 7$ classifications. This is likely because the developed high-density class had less within class spatial variation across the CONUS and more stable temporal reflectance due to the predominance of impermeable surfaces as noted earlier. Nominally we expected greater class specific accuracies with $n_p = 9$ reflecting the greater overall classification accuracy obtained with $n_p = 9$ than $n_p = 7$ (Fig. 8). However, the $n_p = 9$ producer's accuracies were marginally smaller (within 1.5%) than the $n_p = 7$ class producer's accuracies for the woody wetlands (90) and emergent herbaceous wetlands (95) classes, and the $n_p = 9$ user's accuracies were marginally smaller than the $n_p = 7$ user's accuracy for the barren land class (31). This is likely for the reasons described above with respect to the developed open-space and developed low-intensity classes, as wetlands and barren land can exhibit considerable spatial and seasonal variability due to, for example, precipitation and soil moisture changes.

Fig. 16 shows the class specific F1-scores derived as the harmonic mean of the user's and producer's accuracies shown in Fig. 15. The F1-score is low (close to 0) if either the user's or producer's accuracy is low and so the Fig. 16 has a similar class specific accuracies as Fig. 15. For all classifications, the open water (11), deciduous forest (41), evergreen forest (42), shrub/scrub (52), grassland/herbaceous (71) and cultivated crop (82) classes had high (>0.8) F1-scores, and the developed open-space and the mixed forest class had the lowest F1-scores (<0.5). There was no systemic pattern in the F1-scores with respect to the number of quality filtered growing season observations ($n \geq 7$ or $n \geq 9$) or to the number of percentiles ($n_p = 7$ or $n_p = 9$) used.

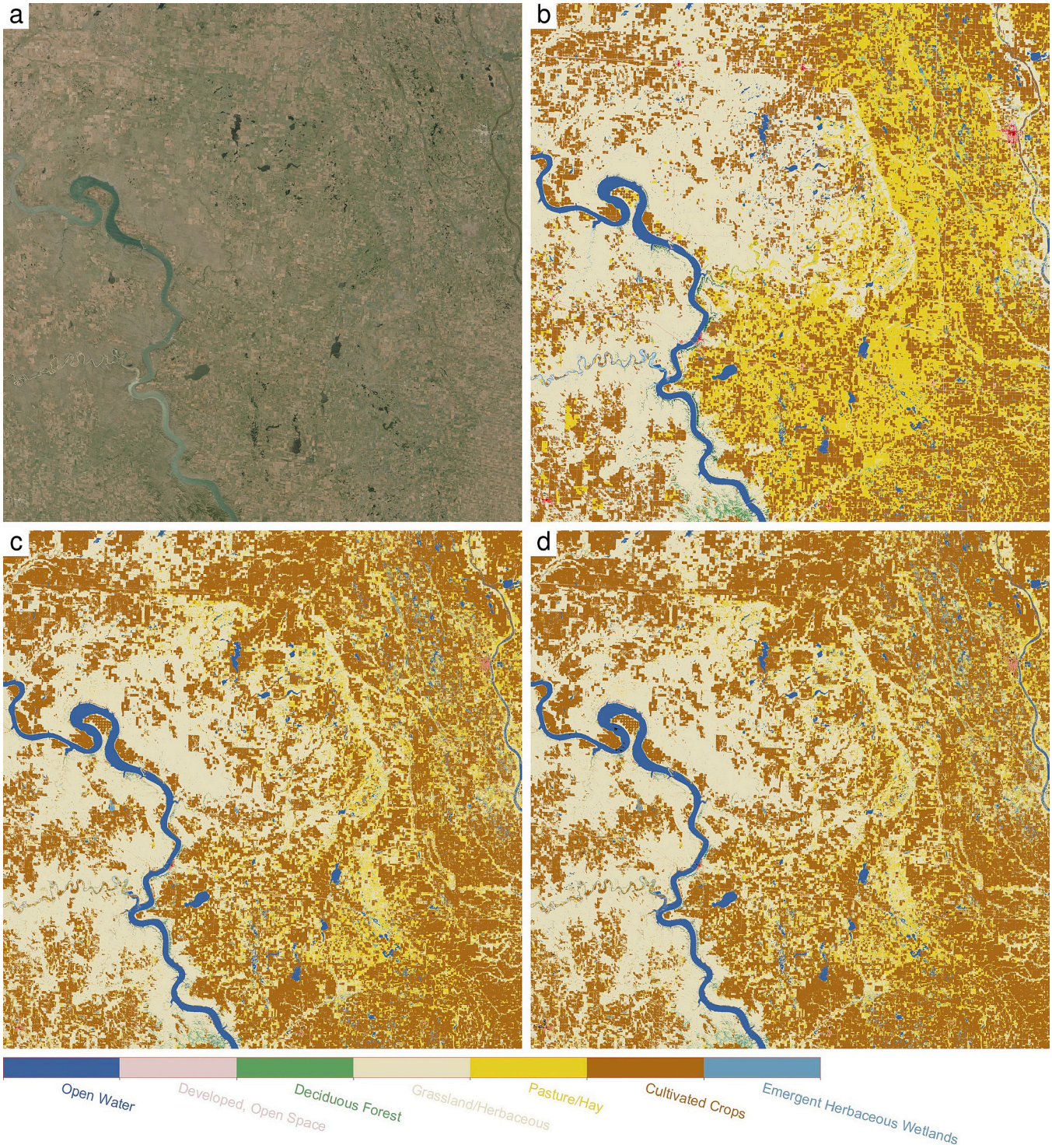

Fig. 12. South Dakota 5000 × 5000 30 m ARD tile results: (a) true color reflectance derived from the median surface NBAR of the 7 months of quality filtered growing season Landsat 5 and 7 observations in the red, green and blue bands, (b) 2011 NLCD classification (spatially subset from Fig. 10), (c) 8-layer 1D CNN classification derived with $n_p = 7$ percentiles (spatially subset from Fig. 9), (d) 8-layer 1D CNN classification derived with $n_p = 9$ percentiles. Only the land cover classes present in (b-d) are included in the figure legend; pixel locations that could not be classified due to insufficient growing season observations are colored black in (c) and (d). Illustrated 150 × 150 km area is Landsat ARD tile h15v06 (centered on 99.0209°W, 44.0124°N). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.4. Signature extension demonstration - application of the 2011 trained 1D CCN model to generate CONUS year 2006 land cover classification results

The 8-layer 1D CNN used to generate the 2011 CONUS land cover classification (shown in Fig. 9, derived with $n_p = 7$ percentiles and a 90% training proportion derived from the NCLD 2011 and 2011 growing season Landsat ARD) was applied to the year 2006 growing season Landsat ARD. Fig. 17 shows a scatterplot comparing the resulting 2006 CONUS land cover class percentages (*y*-axis) with the 2006 NLCD

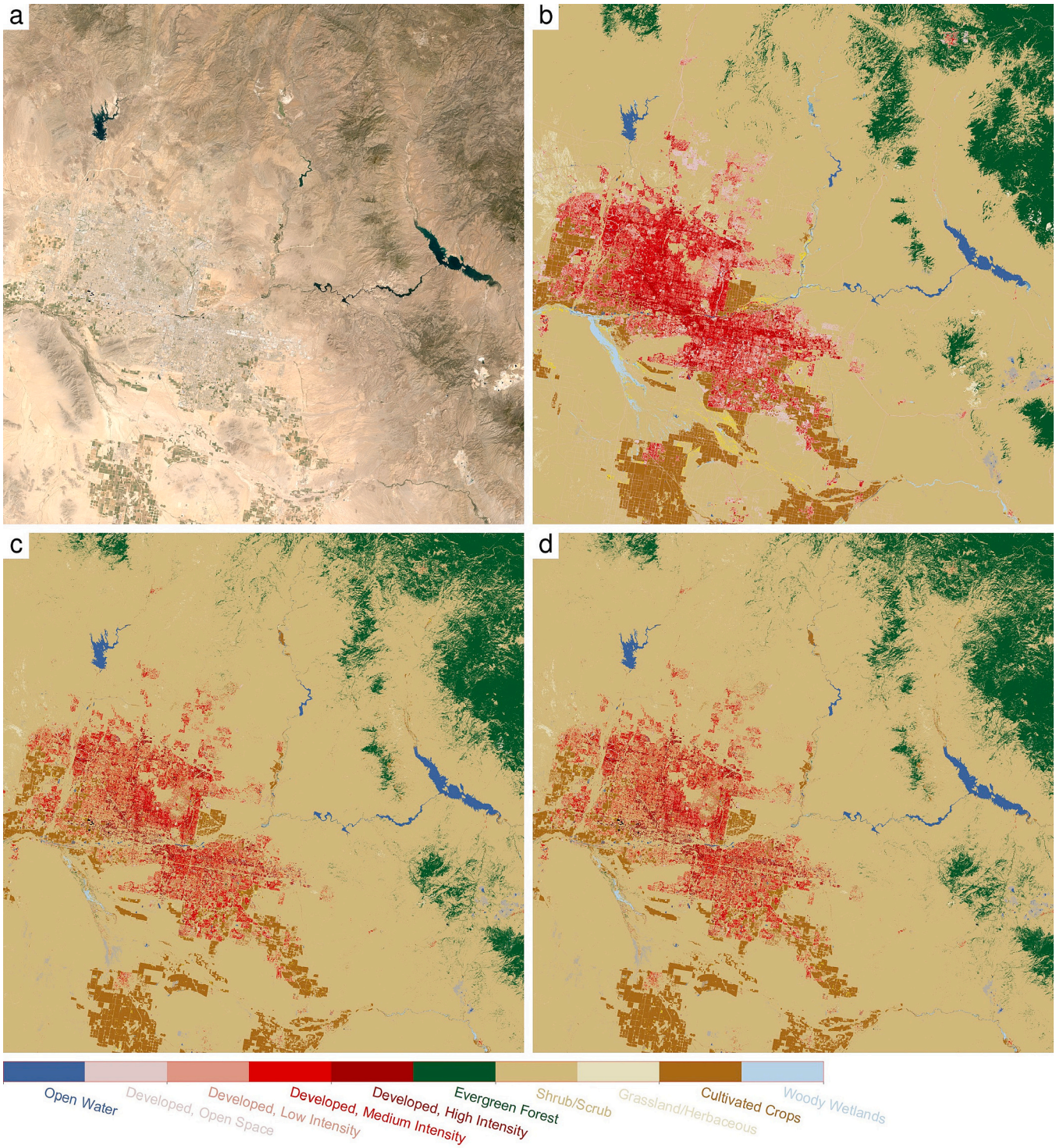

Fig. 13. Arizona 5000 × 5000 30 m ARD tile results: (a) true color reflectance derived from the median surface NBAR of the 7 months of quality filtered growing season Landsat 5 and 7 observations in the red, green and blue bands, (b) 2011 NLCD classification (spatially subset from Fig. 10), (c) 8-layer 1D CNN classification derived with $n_p = 7$ percentiles (spatially subset from Fig. 9), (d) 8-layer 1D CNN classification derived with $n_p = 9$ percentiles. Only the land cover classes present in (b-d) are included in the figure legend; pixel locations that could not be classified due to insufficient growing season observations are colored black in (c) and (d). Illustrated 150 × 150 km area is Landsat ARD tile h07v13 (centered on 111.7090°W, 35.5997°N). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

percentages (*x*-axis). The class percentages between the two classifications have a high 0.99 correlation and similar pattern as the 2011 CONUS scatterplot results illustrated in Fig. 11. However, several classes deviate more from the 1:1 line for the 2006 results, including the mixed forest, emergent herbaceous wetlands, and barren land classes.

The CONUS 2006 overall classification accuracy was quantified by comparison with 3,257,362,006 evaluation samples derived from the NLCD 2006. The CONUS 2006 overall classification accuracy was 80.79% which is less than the 86.40% overall classification accuracy for 2011. This 5.61% accuracy decrease is likely to be influenced not just by

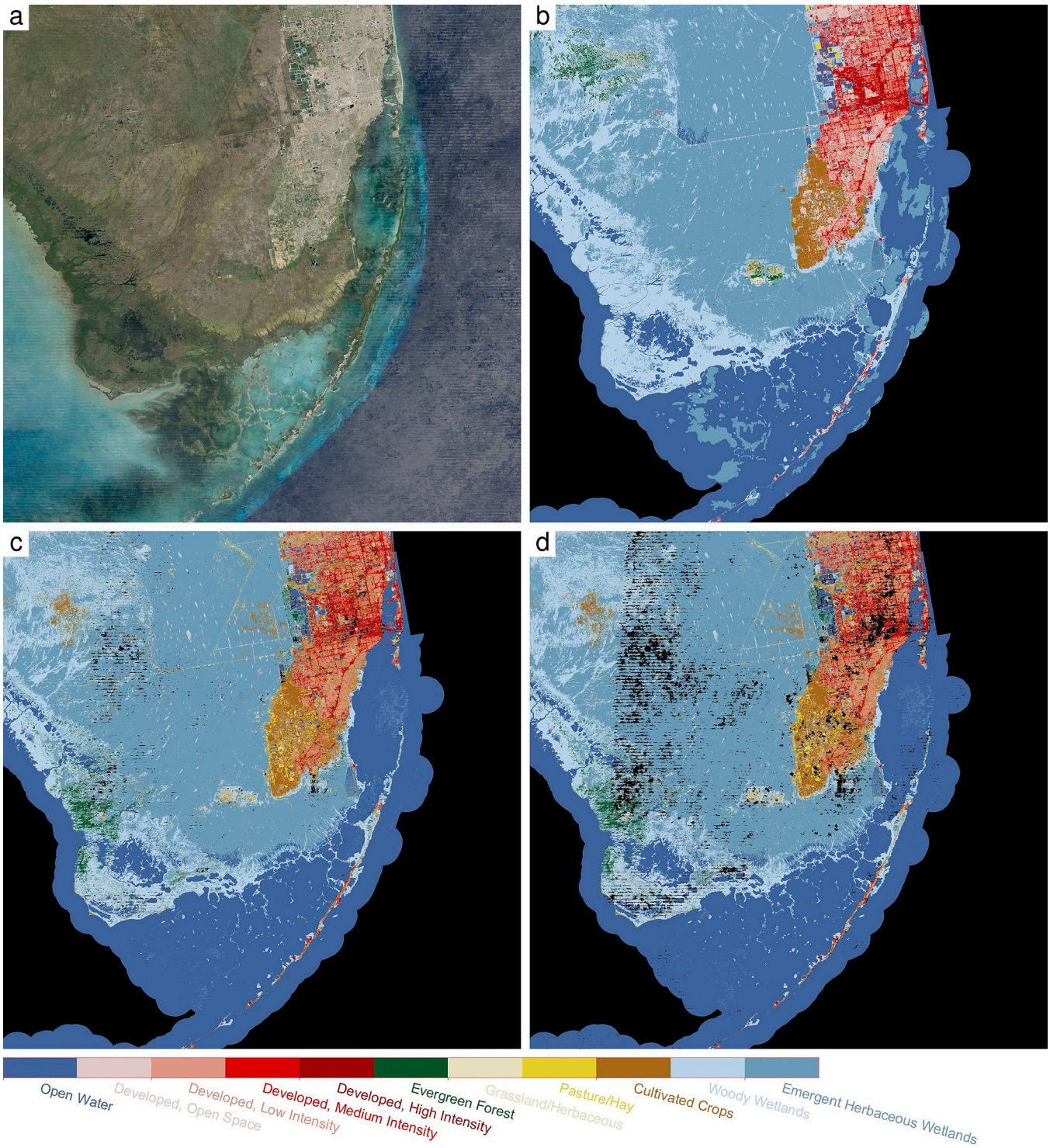

Fig. 14. Florida 5000 × 5000 30 m ARD tile results: (a) true color reflectance derived from the median surface NBAR of the 7 months of quality filtered growing season Landsat 5 and 7 observations in the red, green and blue bands, (b) 2011 NLCD classification (spatially subset from Fig. 10), (c) 8-layer 1D CNN classification derived with $n_p = 7$ percentiles (spatially subset from Fig. 9), (d) 8-layer 1D CNN classification derived with $n_p = 9$ percentiles. Only the land cover classes present in (b-d) are included in the figure legend; pixel locations that could not be classified due to insufficient growing season observations are colored black in (c) and (d). Illustrated 150 × 150 km area is Landsat ARD tile h27v19 (centered on 80.5952°W, 25.4087°N). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the classification methodology but also by satellite data acquisition and surface differences between the two years. These include different intra-annual variations in surface conditions (e.g., vegetation productivity and phenology, soil moisture) in 2006 and 2011, land cover and land use change from 2006 to 2011, changes in agricultural crop types between 2006 and 2011, differences in the cloud cover at the time of overpass between 2006 and 2011, and Landsat orbit drift (and so reflectance differences caused by differences in the solar position at Landsat overpass time) (Roy et al., 2020).

To examine the potential influence of differences in the satellite data

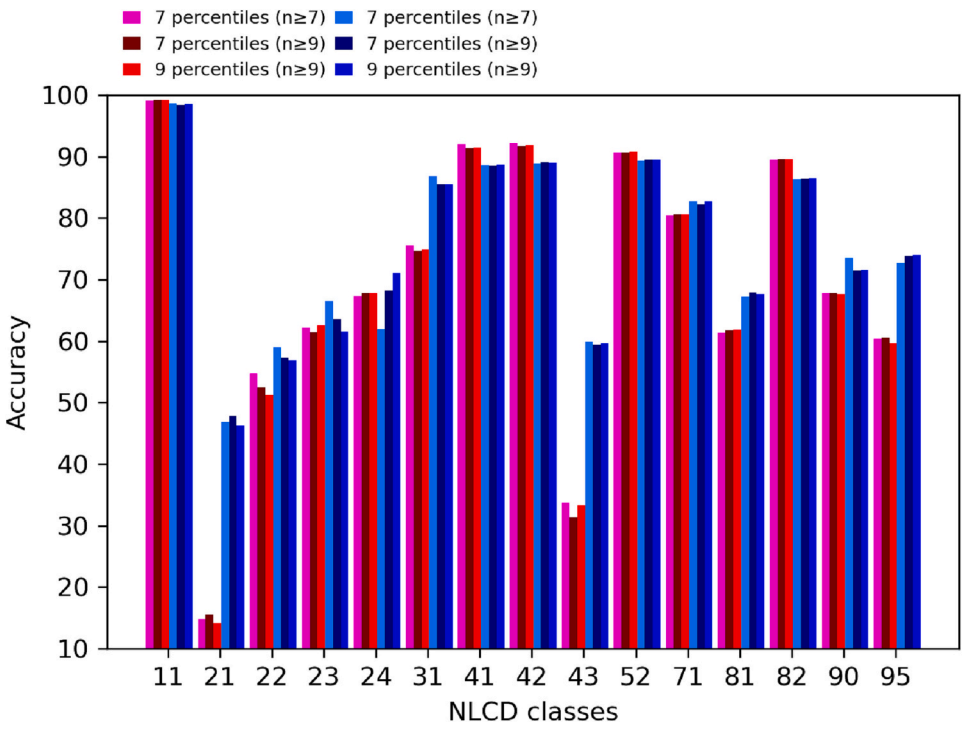

Fig. 15. CONUS class specific producer's accuracies (red shades) and user's accuracies (blue shades) for the 15 NLCD land cover classes (see Table 1). Results for the 8-layer 1D CNN classification derived using $n_p = 7$ percentiles considering CONUS ARD pixel locations with ≥7 and ≥ 9 quality filtered growing season observations, and derived using $n_p = 9$ percentiles considering CONUS ARD pixel locations with ≥9 quality filtered growing season observations. The classifications were trained using the 90% training proportions (about 2.9 million CONUS samples) and evaluated using an independent 10% evaluation proportion (about 0.3 million CONUS samples) (Table 2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

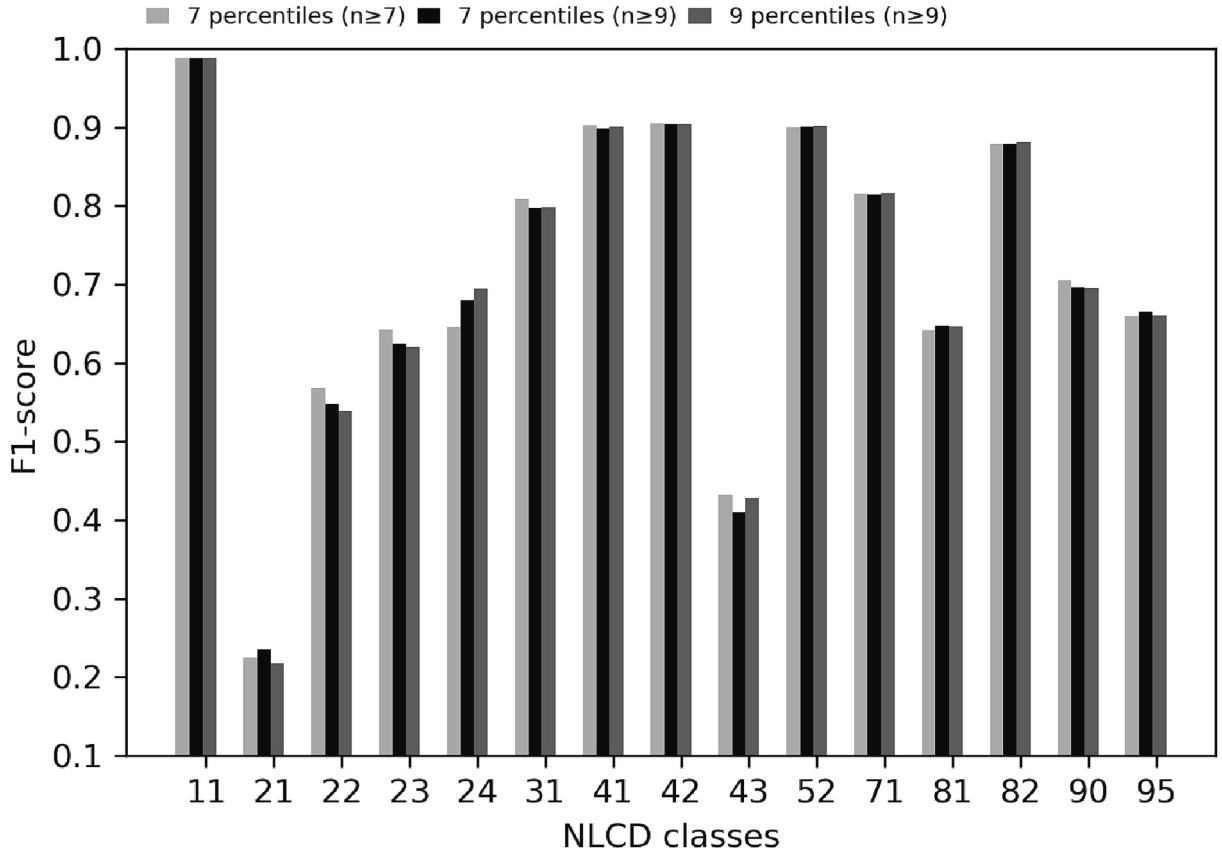

Fig. 16. CONUS class specific F1-score results derived from the Fig. 15 8-layer 1D CNN classification producer's accuracies and user's accuracies.

and the surface between years, the same signature extension experiment was undertaken but using $n_p = 1$ percentile (the 50th percentile) rather than $n_p = 7$ percentiles. Notably, the overall CONUS classification accuracy for $n_p = 1$ was 77.27% for 2006, and 79.75% for 2011. This 2.48% decrease is relatively smaller than the 5.61% decrease reported above for $n_p = 7$ percentiles. Recall that the Figs. 6 and 7 results showed that, regardless of the model used, the CONUS overall classification accuracies with $n_p = 1$ were consistently lower than with $n_p = 9$. This is

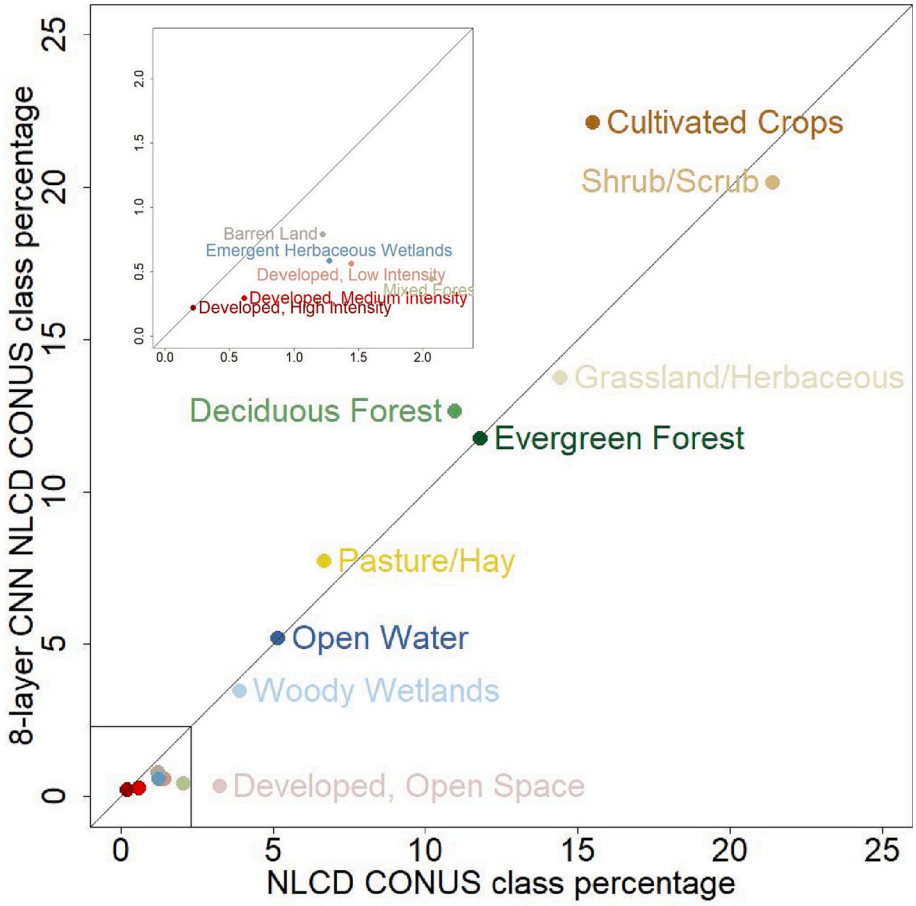

Fig. 17. Scatterplot, as Fig. 11, but comparing the CONUS class percentages defined by the 8-layer 1D CNN trained with 2011 Landsat ARD (with $n_p = 7$ percentiles and the 90% training proportion derived from the NLCD 2011) used to classify the 2006 Landsat ARD (y-axis), with the 2006 NLCD class percentages (x axis).

because using a greater number of percentiles better captured seasonal reflectance variations. This is evident in Fig. 18 that shows class specific F1-scores comparing the 2011 (y-axis) and 2006 (x-axis) results derived for $n_p = 7$ percentiles (the start of each arrow) and $n_p = 1$ (the end of each arrow, i.e., the arrow tips). The majority of the arrows (starts and ends) are located above the 1:1 line indicating that, as expected, the year 2011 classification F1-scores are greater than the 2006 classification F1-scores. Notably, most of the arrows in Fig. 18 point toward the origin (i.e., the slope of the arrow is greater than one) indicating that the F1-score decrease from 2011 to 2006 is greater (signature extension issue is more serious) with $n_p = 7$ than $n_p = 1$. This is likely because using $n_p = 1$ is less sensitive to differences in the satellite data and the surface between years. Notably, the classes with arrows further from the 1:1 line are mixed forest, emergent herbaceous wetlands, woody wetlands, grassland/herbaceous, and shrub/scrub, that all typically exhibit significant seasonal reflectance variation. Conversely, the open water class F1-scores are located almost on the 1:1 line indicating that signature extension is not an issue for this class which is likely because of the characteristically low and relatively temporally unchanging reflectance of open water relative to land surfaces.

5. Discussion and conclusion

Medium spatial resolution multi-spectral satellite data acquired by sensors, such as Landsat for more than 50 years (Wulder et al., 2022), and in the last decade by Sentinel-2 (Drusch et al., 2012), have been used to derive national to global scale land cover maps, predominantly using supervised decision tree classifiers. Over large areas, land cover classification has conventionally been undertaken using satellite time series, typically using temporal metric percentiles derived from annual growing season time series (Wulder et al., 2018). Deep convolutional neural networks (CNNs) were first demonstrated for application to single date high spatial resolution images (Zhang et al., 2018; Kellenberger et al., 2018; Mahdianpari et al., 2018; Srivastava et al., 2019; Tong et al., 2020; Yuan et al., 2020) and their considerable potential makes them attractive for large area land cover classification. For example, recently, the Environmental Systems Research Institute (ESRI) (Karra et al., 2021) and the Google Dynamic World (Brown et al., 2022) initiatives have derived global coverage land cover maps by patch-based CNN classification of single date predominantly cloud-free and non-hazy Sentinel-2 10 m images using fully convolution networks that normally classify the entire patch rather than the center patch pixel (Ronneberger et al., 2015). The application of CNNs to satellite image time series, to take advantage of spectral differences among land cover classes over time, is complicated because of missing observations due to clouds and irregular surface observation temporal cadence. These issues commonly occur in medium spatial resolution multi-spectral satellite data (Egorov et al., 2019). For example, the percentage of the CONUS ARD 30 m pixel locations that had at least n good quality cloud-free Landsat 5 TM and Landsat 7 ETM+ observations over the seven month growing season of 2011 examined in this study declined rapidly with n from 99.90% for $n = 3$, 97.88% for $n = 9$, 93.56% for $n = 11$, to 50.29% for $n = 20$ (Fig. 2). It is well established that spatial and temporal differences in land cover class spectral signatures increase with geographic coverage and so the incidence of missing observations can be particularly problematic when a single classification model is applied. Recent studies have applied 1D CNN architectures to classify single pixel reflectance time series and have interpolated missing or cloud-flagged observations using preceding

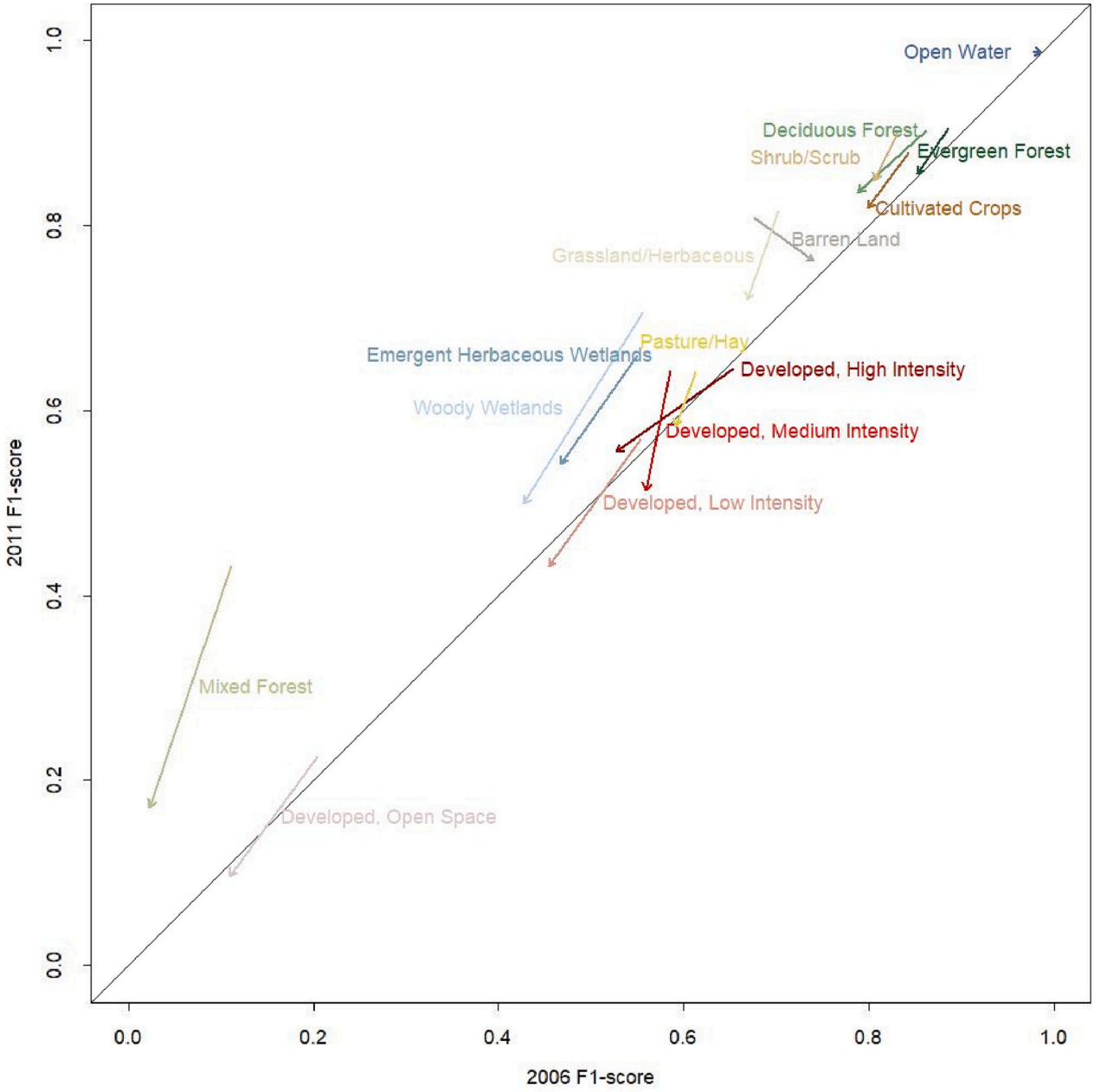

Fig. 18. F1-scores for each land cover class for the year 2011 (y-axis) and 2006 (x-axis) classifications using $n_p = 7$ (arrow start) and $n_p = 1$ (arrow end) percentiles. The 1:1 line is shown for reference.

and subsequent cloud-free observation values. Interpolation is unreliable, however, when gaps occur in periods of rapid surface change, when the gaps have long duration, and when there are undetected clouds or shadows that are used incorrectly in the interpolation (Yan and Roy, 2020).

This study demonstrates, for the first time, 1D CNN single pixel time series land classification derived using temporal percentile metrics and demonstrates this at scale for all the CONUS. Temporal metric percentiles decompose irregular distributed satellite time series into a reduced fixed number of features that can be conveniently used for classification purposes and were developed for national to global scale land cover classification as they are robust to missing data and reduce the impact of spatial and temporal differences in land cover spectral signatures (DeFries et al., 1995; Friedl et al., 2010; Hansen et al., 2014; Zhang and Roy, 2017). The temporal metric percentiles used in this study were derived at each CONUS ARD 30 m pixel location from quality filtered Landsat 5 TM and Landsat 7 ETM+ ARD surface NBAR time series acquired over the seven month growing season of 2011. Percentiles of the five Landsat bands and of the eight possible two band normalized NBAR ratios were derived to provide $n_p \times 13$ temporal metrics where n_p is the number of percentiles and in this study was set as 5, 7 or 9. The temporal metric percentiles were derived at each CONUS ARD pixel and used as classification predictor variables. A pool of >3.3 million CONUS 30 m pixels was used to derive independent training and evaluation data. The pool was derived by systematic sampling the year 2011 (reprocessed in 2014) National Land Cover Database (NLCD) 16 classes land cover product. The NLCD is a widely applied CONUS land cover product that is generated on a systematic basis by the USGS and has been robustly validated with a reported 86.8% overall land cover classification accuracy (Wickham et al., 2021). The pool was derived carefully to minimize NLCD classification errors by selecting only 30 m pixel locations with the same NLCD land cover class in the surrounding eight 30 m pixels (to

reduce the impact of Landsat misregistration and isolated single pixel NLCD misclassification errors). Further, only locations with $\geq n_p$ quality filtered year 2011 growing season observations needed to undertake the classification were selected.

The sensitivity of the CONUS 1D CNN classification results to using different numbers of temporal metric percentiles (n_p), CNN architecture complexity, and training data amount was investigated. A 5-layer 1D CNN based on the structure used by Pelletier et al. (2019) and an 8-layer 1D CNN model that extended the 5-layer model to include 3 more convolutional layers, with 0.2 and 2.1 million learnable coefficients, respectively, were assessed. The influence of using different training sample sizes (10%, 50% and 90% of the 3.3 million sample pool) was also examined. As expected, the overall classification accuracies increased with the training sample size for the two 1D CNN architectures, and also for the random forest classifier that was included as a benchmark. Notably, both 1D CNN architectures provided higher CONUS overall classification accuracies than random forest by 1.9% - 2.8% which, given the high overall CONUS classification accuracies (>83%) is a meaningful increase. The 8-layer 1D CNN provided the highest overall classification accuracies for the 50% and 90% training proportion experiments and the overall classification accuracies between the two CNN models differed by <0.73%. Thus, on the basis of these results, the 8-layer 1D CNN is recommended rather than the 5-layer 1D CNN model, and either 1D CNN model is recommended over random forest.

In principle, using a larger number of temporal percentiles (n_p) should better capture seasonal surface variations and so provide higher CONUS classification accuracy. This was illustrated by an experiment using only one ($n_p = 1$) percentile (defined by the median growing season value for each of the five Landsat bands and of the eight possible two band normalized NBAR ratios) that provided consistently lower classification accuracies than found using $n_p = 9$ percentiles regardless of the model (random forest or 1D CNN). Similarly, using seven monthly median composites, provided systematically lower accuracies than the $n_p = 9$ classification accuracies. Previously, researchers have used Landsat temporal metrics defined by five or seven percentiles for large area land cover classification using random forest and decision tree classifiers (Potapov et al., 2012; Margono et al., 2012; Yan and Roy, 2015; Azzari and Lobell, 2017; Pflugmacher et al., 2019). In this study nine percentiles were considered, even though 2.12% of the CONUS pixels had <9 quality filtered growing season observations and so could not be classified with $n_p = 9$. The CONUS overall classification accuracy with $n_p = 9$ was 86.43% and was marginally (to the second decimal place) greater than the $n_p = 7$ overall accuracy (86.40%), greater than the $n_p = 5$ overall accuracy (86.21%) and significantly higher than the $n_p = 1$ overall accuracy (79.90%). Class specific producer and user accuracies were also quantified with respect to the number of temporal percentiles ($n_p = 7$ and 9). There was no systematic pattern in the class specific accuracy results with respect to n_p. A total of 99.43% and 97.88% of CONUS ARD pixels had sufficient quality filtered time series observations to support $n_p = 7$ or $n_p = 9$ land clover classification, respectively. On the basis of these results, the 8-layer 1D CNN with $n_p = 7$ or $n_p = 9$ is recommended. We note that the temporal observation coverage of the CONUS Landsat 5 TM and Landsat 7 ETM+ data may be different in other regions of the world, such as the tropics that are typically cloudy at the time of Landsat overpass, and also that two Landsat sensor coverage is not always available even over the CONUS (Kovalskyy and Roy, 2013; Wulder et al., 2016). Consequently, using a 1D CNN with five percentiles may be more appropriate in these instances to increase the number of pixel locations that can be classified. Conversely, land cover classifications undertaken with data sensed by both Sentinel-2 sensors or in combination with Landsat may have higher accuracy with $n_p = 9$ due to the greater temporal observation coverage provided by these sensors (Li and Roy, 2017).

The 30 m CONUS and detailed 30 m ARD tile mapped classification results presented in this study demonstrate that the 1D CNN single pixel temporal metric land classification approach is effective at scale and locally. The CONUS 8-layer 1D CNN and NLCD classification maps illustrated at 1.5 × 1.5 km resolution were qualitatively similar (Figs. 9 and 10). However, the developed open-space class was less apparent around certain cities in the 8-layer 1D CNN map relative to the NLCD, and more cultivated crops but less pasture/hay in the U.S. agricultural heartland was apparent in the 8-layer 1D CNN classification than the NLCD. These differences were quantified in a scatterplot comparing the percentage of CONUS 30 m ARD land pixels classified into each of 15 classes by the 8-layer 1D CNN and NLCD classifications (Fig. 11). The developed open-space class covered 3.3% of the CONUS 30 m NLCD pixels and only 0.4% in the 1D CNN classification, and the cultivated crop class covered of 15.5% of NLCD and 19.2% of the 1D CNN CONUS 30 m pixels. These discrepancies can be explained. It is well established that urban/suburban areas can be composed of different land cover types and land uses (e.g., impervious surfaces, grass lawns, swimming pools, bare soil) that can be mixed spatially within 30 m pixels and so are easily confused with other land cover classes (Small, 2005; Griffiths et al., 2010; Zhang and Roy, 2017). In addition, the four developed land classes had the smallest number of training samples across all the classes (Table 1) and so may be less accurately classified due to class imbalance issues (Chawla, 2003; Mellor et al., 2015). Indeed, due to this complexity, the four NLCD developed classes were derived using additional road vector, night-time light, and digital elevation data (Yang et al., 2003; Homer et al., 2015), that were not used in the 1D CNN classifications, and the developed open-space class had the lowest reported classification accuracy of the four NLCD developed classes (Wickham et al., 2021). It is also well established that cultivated crops and pasture/hay can be hard to differentiate reliably in satellite imagery (Hill et al., 1999; Kuchler et al., 2020), and the pasture/hay class also had low reported NLCD producer's and user's accuracy (Wickham et al., 2021). Despite these differences, the scatterplot correlations comparing the NLCD CONUS class percentages with the 8-layer 1D CNN CONUS class percentages for $n_p = 7$ and $n_p = 9$, were 0.99 and 0.98, respectively (Fig. 11). These correlations indicate good overall correspondence, particularly as the 1D CNN training data corresponded to 0.037% of the >8.5 billion CONUS land 30 m pixels that were classified. Detail results for three 5000 × 5000 30 m pixel ARD tiles were also presented and demonstrated that at native resolution the 1D CNN classification land cover class boundaries were preserved for features with small axis dimensions, including roads, buildings, lakes and rivers. In addition, the 1D CNN tile results had no stripes or anomalous spatial patterns that can sometimes be observed in large area single image patch-based CNN land cover classifications (Karra et al., 2021; Brown et al., 2022).

To further demonstrate the 1D CNN approach, the 8-layer 1D CNN $n_p = 7$ model used to generate the 2011 CONUS land cover classification was applied to the seven month growing season of year 2006 Landsat ARD. The resulting 2006 land cover classification had similar CONUS class percentages documented for 2011. However, an accuracy assessment undertaken by comparison of the 2006 classification with NLCD 2006 evaluation samples revealed a 5.61% lower overall classification accuracy than the 86.40% accuracy found for 2011. This is likely due to a number of factors including satellite data acquisition and surface differences between 2006 and 2011. This was indicated by examination of the class F1-scores for 2011 and 2006 results derived using the 8-layer 1D CNN $n_p = 7$ model and the 8-layer 1D CNN $n_p = 1$ model. For most classes the F1-score decreased from 2011 to 2006 and the decrease was greater with $n_p = 7$ than $n_p = 1$. Notably, the mixed forest, emergent herbaceous wetlands, woody wetlands, grassland/herbaceous, and shrub/scrub classes, that all typically exhibit significant seasonal reflectance variation, had greater differences. Conversely, the open water class F1-scores were similar for 2011 and 2007 and using $n_p = 7$ or $n_p = 9$, which is likely because of the characteristically low and relatively temporally unchanging reflectance of open water compared to the other land cover classes.

Temporal metric percentiles can be used for large area land cover

classification with other deep learning models, such as RNN and fully attention based networks (as discussed in the introduction with respect to patch-based implementations). Notably, these models have been applied to classify single pixel time series by using interpolated values for cloud contaminated or missing observations, e.g., with RNN (Ienco et al., 2017; Campos-Taberner et al., 2020) and with fully attention based networks (Rußwurm and Körner, 2020; Yuan and Lin, 2021). We note that studies comparing these different models over small areas have reported mixed results. For example, Xu et al. (2020) found the RNN provided higher accuracy than the fully attention based network for classification of Landsat reflectance time series over six US 51× 51 km sites into three classes (corn, soybean, other). Rußwurm and Körner (2020) found negligible difference among 1D CNN, RNN and fully attention based models for classification of Sentinel-2 reflectance time series into 23 land cover classes for three <50 × 50 km regions in Bavaria, Germany. Yuan and Lin (2021) reported that the fully attention based network performed better than 1D CNN and RNN for classification of Sentinel-2 reflectance time series over 110 × 110 km areas into agricultural and non-agricultural classes in California and Missouri, and into 5 land cover classes in Beijing, China. In principle the RNN and fully attention models can handle variable length time series data (Cho et al., 2014; Devlin et al., 2018) without the need to interpolate missing data. Indeed, Rußwurm et al. (2023) recently demonstrated this for in-season mapping of crop types using a year of Sentinel-2 images over a 27,200 km^2 area in France and a 1400 km^2 area in Germany. For each study area they trained and applied a single RNN model with reasonable classification accuracies given the number of classes (75% overall accuracy for 14 crop types in the French site and 86% overlap accuracy for 7 crop types in Germany).

Finally, we note that the 1D CNN structure used in this study was implemented because it is straightforward to apply to single pixel time series. However, a 2D CNN structure, usually applied to image patches, could be applied to the $n_p \times 13$ temporal metrics by treating them as an image patch composed of ($n_p \times 13$) pixels and 1 image band. We undertook experiments to check this and found only slightly poorer classification accuracy. For example, using an 8-layer 2D CNN with the same number of layers and kernels as the 1D implementation provided only slightly smaller (<0.5%) overall classification accuracies than the 1D CNN using the 90% training proportion. Notably, the 1D CNN algorithm is computationally quite efficient. It was implemented in Python on a Linux server with 40 Intel Xeon CPU cores, 768 GB RAM and two NVIDIA Tesla P100 PCIe 16 GB GPUs. Using this architecture, the 8-layer 1D CNN took 8.85 h to train 2.9 million CONUS training samples and 2.5 s to classify 0.3 million CONUS evaluation samples. The random forest was implemented on the same server but without using the GPUs and took 6.59 h and 113.2 s to train and classify the same data, respectively.

In summary, the 1D CNN single pixel temporal metric land classification approach presented in this paper has several potential advantages over conventional patch-based CNN land cover classification that can have the following issues. First, patch-based CNN may blur small and spatially fragmented surface features within the image, and methods to reduce this issue, for example, by training multiple CNN models with different patch sizes, although useful, do not provide a universal solution (Martins et al., 2020; Zhang et al., 2020). Second, discontinuities along image boundaries can occur in conventional patch-based CNN classifications, particularly when adjacent images are acquired on different calendar dates and so have different vegetation cover, condition, and soil moisture. Third, CNN patch-based classification is less reliable if one or more of the patch pixels are missing or contaminated by cloud or shadow. These issues do not occur with the developed approach that uses single pixel temporal metrics to take advantage of spectral differences among land cover classes over time and to remove the need to interpolate missing observations. As noted above, other deep learning architectures can also be used. To facilitate future comparison studies, the training and evaluation samples and manipulation code developed in this study are publicly available. In addition, the code could be adapted for application to other multi-temporal satellite data sets and/or land cover training data.

Code and training and evaluation data availability

The training and evaluation samples used in this study are available at https://zenodo.org/record/7106054 and python manipulation codes are available at https://github.com/hankui/cnn_Landsat_time_series_classification_v2-Python. The CONUS 30 m land cover product derived using the 8-layer CNN and 7 percentile model (Fig. 9) is available at https://zenodo.org/record/77405#.ZCb5YXaZNaQ.

CRediT authorship contribution statement

Hankui K. Zhang: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **David P. Roy:** Conceptualization, Writing – original draft, Writing – review & editing, Funding acquisition. **Dong Luo:** Software, Formal analysis.

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.

Data availability

We share the code and product - as stated at end of paper.

Acknowledgments

This research was funded by the NASA Advancing Collaborative Connections for Earth System Science (ACCESS) Program (grant 80NSSC21M0023) and by the U.S. Geological Survey Landsat science team (grant G12PC00069). The USGS Landsat program management and staff are thanked for the free provision of the Landsat ARD used in this study. Dr. Vitor Martins is thanked for CNN discussions in the early stages of this research.

References

Azzari, G., Lobell, D.B., 2017. Landsat-based classification in the cloud: an opportunity for a paradigm shift in land cover monitoring. Remote Sens. Environ. 202, 64–74.
Belenguer-Plomer, M.A., Tanase, M.A., Chuvieco, E., Bovolo, F., 2021. CNN-based burned area mapping using radar and optical data. Remote Sens. Environ. 260, 112468.
Bilgic, B., Chatnuntawech, I., Fan, A.P., Adalsteinsson, E., 2014. Fast image reconstruction with L2-regularization. J. Magn. Reson. Imaging 40 (1), 181–191.
Boschetti, L., Roy, D.P., Giglio, L., Humber, M.L., 2019. Global validation of the collection 6 MODIS burned area product. Remote Sens. Environ. 235, 111490.
Bottou, L., 2010. Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT'2010, pp. 177–186. Physica-Verlag HD.
Boureau, Y.L., Ponce, J., LeCun, Y., 2010. A theoretical analysis of feature pooling in visual recognition. In: Proceedings of the 27th International Conference on Machine Learning, pp. 111–118.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Brooks, E.B., Thomas, V.A., Wynne, R.H., Coulston, J.W., 2012. Fitting the multitemporal curve: a Fourier series approach to the missing data problem in remote sensing analysis. IEEE Trans. Geosci. Remote Sens. 50 (9), 3340–3353.
Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Tait, A.M., 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. Scientific Data 9 (1), 1–17.
Campos-Taberner, M., García-Haro, F.J., Martínez, B., Izquierdo-Verdiguier, E., Atzberger, C., Camps-Valls, G., Gilabert, M.A., 2020. Understanding deep learning in land use classification based on Sentinel-2 time series. Sci. Rep. 10 (1), 1–12.
Chawla, N.V., 2003, August. C4. 5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure. In: Proceedings of the ICML, Vol. 3. CIBC, Toronto, ON, Canada, p. 66.
Chen, T.H.K., Qiu, C., Schmitt, M., Prishchepov, A.V., 2020. Mapping horizontal and vertical urban densification in Denmark with Landsat time-series from 1985 to 2018: a semantic segmentation solution. Remote Sens. Environ. 251, 112096.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
Colditz, R.R., 2015. An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms. Remote Sens. 7 (8), 9655–9681.
Colditz, R.R., Saldaña, G.L., Maeda, P., Ressl, R., 2012. Generation and analysis of the 2005 land cover map for Mexico using 250m MODIS data. Remote Sens. Environ. 123, 541–552.
Congalton, R.G., Green, K., 2019. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices (Third edition), 346 pages. CRC Press, Boca Raton, FL, USA.
De Fries, R.S., Hansen, M., Townshend, J.R.G., Sohlberg, R., 1998. Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. Int. J. Remote Sens. 19 (16), 3141–3168.
Debella-Gilo, M., Gjertsen, A.K., 2021. Mapping seasonal agricultural land use types using deep learning on Sentinel-2 image time series. Remote Sens. 13 (2), 289.
DeFries, R., Hansen, M., Townshend, J., 1995. Global discrimination of land cover types from metrics derived from AVHRR pathfinder data. Remote Sens. Environ. 54, 209–222.
Derksen, D., Inglada, J., Michel, J., 2019. A metric for evaluating the geometric quality of land cover maps generated with contextual features from high-dimensional satellite image time series without dense reference data. Remote Sens. 11 (16), 1929.
Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv: 1810.04805*.
Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens. Environ. 120, 25–36.
Dwyer, J.L., Roy, D.P., Sauer, B., Lymburner, L., 2018. Analysis Ready Data: enabling analysis of the Landsat archive. Remote Sens. 10 (9), 1363.
Egorov, A.V., Roy, D.P., Zhang, H.K., 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 Sens. 11 (4), 447.
Fazzini, P., De Felice Proia, G., Adamo, M., Blonda, P., Petracchini, F., Forte, L., Tarantino, C., 2021. Sentinel-2 remote sensed image classification with patchwise trained ConvNets for grassland habitat discrimination. Remote Sens. 13 (12), 2276.
Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114 (1), 168–182.
Gao, F., He, T., Masek, J.G., Shuai, Y., Wang, Z., 2014. Angular effects and correction for medium resolution sensors to support crop monitoring. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 7 (11), 4480–4489.
Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256.
Glorot, X., Bordes, A., Bengio, Y., 2011, June. Deep sparse rectifier neural networks. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pp. 315–323.
Grabska, E., Frantz, D., Ostapowicz, K., 2020. Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. Remote Sens. Environ. 251, 112103.
Gray, J., Song, C., 2013. Consistent classification of image time series with automatic adaptive signature generalization. Remote Sens. Environ. 134, 333–341.
Griffiths, P., Hostert, P., Gruebner, O., van der Linden, S., 2010. Mapping megacity growth with multi-sensor data. Remote Sens. Environ. 114 (2), 426–439.
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, 135–151.
Hansen, M., Egorov, A., Potapov, P., Bentsg, T., 2014. Monitoring conterminous United States (CONUS) land cover change with web-enabled Landsat data (WELD). Remote Sens. Environ. 140, 466–484.
He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
Hermosilla, T., Wulder, M.A., White, J.C., Hobart, G.W., 2018. Disturbance-informed annual land cover classification maps of Canada's forested ecosystems for a 29-year landsat time series. Can. J. Remote. Sens. 44 (1), 67–87.
Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., 2022. Land cover classification in an era of big and open data: optimizing localized implementation and training data selection to improve mapping outcomes. Remote Sens. Environ. 268, 112780.
Hill, M.J., Vickery, P.J., Furnival, E.P., Donald, G.E., 1999. Pasture land cover in eastern Australia from NOAA-AVHRR NDVI and classified Landsat TM. Remote Sens. Environ. 67 (1), 32–50.
Homer, C., Huang, C., Yang, L., Coan, M., 2004. Development of a 2001 national land-cover database for the United States. Photogramm. Eng. Remote. Sens. 70, 829–840.
Homer, C.G., Dewitz, J.A., Yang, L., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogramm. Eng. Remote. Sens. 81, 345–354.
Hosseiny, B., Mahdianpari, M., Brisco, B., Mohammadimanesh, F., Salehi, B., 2021. WetNet: a spatial–temporal ensemble deep learning model for wetland classification using Sentinel-1 and Sentinel-2. IEEE Trans. Geosci. Remote Sens. 60, 1–14.
Huang, B., Zhao, B., Song, Y., 2018. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. Remote Sens. Environ. 214, 73–86.
Ienco, D., Gaetano, R., Dupaquier, C., Maurel, P., 2017. Land cover classification via multitemporal spatial data by deep recurrent neural networks. IEEE Geosci. Remote Sens. Lett. 14 (10), 1685–1689.
Interdonato, R., Ienco, D., Gaetano, R., Ose, K., 2019. DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn. ISPRS J. Photogramm. Remote Sens. 149, 91–104.
Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:*1502.03167.
Johnson, D.M., Mueller, R., 2021. Pre-and within-season crop type classification trained with archival land cover information. Remote Sens. Environ. 264, 112576.
Ju, J., Roy, D.P., Vermote, E., Kovalskyy, V., 2012. Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods. Remote Sens. Environ. 122, 175–184.
Karakizi, C., Karantzalos, K., Vakalopoulou, M., Antoniou, G., 2018. Detailed land cover mapping from multitemporal Landsat-8 data of different cloud cover. Remote Sens. 10 (8), 1214.
Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M., Brumby, S.P., 2021, July. Global land use/land cover with Sentinel 2 and deep learning. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. IEEE, pp. 4704–4707.
Kellenberger, B., Marcos, D., Tuia, D., 2018. Detecting mammals in UAV images: best practices to address a substantially imbalanced dataset with deep learning. Remote Sens. Environ. 216, 139–153.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
Kovalskyy, V., Roy, D.P., 2013. The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30 m Landsat data product generation. Remote Sens. Environ. 130, 280–293.
Kovalskyy, V., Roy, D.P., 2015. A one year Landsat 8 conterminous United States study of cirrus and non-cirrus clouds. Remote Sens. 7 (1), 564–578.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105.
Kuchler, P.C., Begue, A., Simoes, M., Gaetano, R., Arvor, D., Ferraz, R.P., 2020. Assessing the optimal preprocessing steps of MODIS time series to map cropping systems in Mato Grosso, Brazil. Int. J. Appl. Earth Obs. Geoinf. 92, 102150.
Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14 (5), 778–782.
Kwak, G.H., Park, C.W., Lee, K.D., Na, S.I., Ahn, H.Y., Park, N.W., 2021. Potential of hybrid CNN-RF model for early crop mapping with limited input data. Remote Sens. 13 (9), 1629.
Lange, M., Feilhauer, H., Kühn, I., Doktor, D., 2022. Mapping land-use intensity of grasslands in Germany with machine learning and Sentinel-2 time series. Remote Sens. Environ. 277, 112888.
LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L. D., 1989. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1 (4), 541–551.
LeCun, Y.A., Bottou, L., Orr, G.B., Müller, K.R., 2012. Efficient backprop. In: Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, pp. 9–48.
Li, J., Roy, D.P., 2017. A global analysis of Sentinel-2A, Sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring. Remote Sens. 9 (9), 902.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2 (3), 18–22.
Lindquist, E.J., Hansen, M.C., Roy, D.P., Justice, C.O., 2008. The suitability of decadal image data sets for mapping tropical forest cover change in the Democratic Republic of Congo: implications for the global land survey. Int. J. Remote Sens. 29 (24), 7269–7275.
Liu, H., Gong, P., Wang, J., Xu, B., 2021. Production of global daily seamless data cubes and quantification of global land cover change from 1985 to 2020-iMap World 1.0. Remote Sens. Environ. 258, 112364.
Liu, M., Chai, Z., Deng, H., Liu, R., 2022a. A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 15, 4297–4306.
Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022b. A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11976–11986.
Lobert, F., Holtgrave, A.K., Schwieder, M., Pause, M., Vogt, J., Gocht, A., Erasmi, S., 2021. Mowing event detection in permanent grasslands: systematic evaluation of input features from Sentinel-1, Sentinel-2, and Landsat 8 time series. Remote Sens. Environ. 267, 112751.
Lu, W., Tao, C., Li, H., Qi, J., Li, Y., 2022. A unified deep learning framework for urban functional zone extraction based on multi-source heterogeneous data. Remote Sens. Environ. 112830.
Mahdianpari, M., Salehi, B., Rezaee, M., Zhang, Y., 2018. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. Remote Sens. 10 (7), 1119.
Margono, B.A., Turubanova, S., Zhuravleva, I., Potapov, P., Tyukavina, A., Baccini, A., Hansen, M.C., 2012. Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010. Environ. Res. Lett. 7 (3), 034010.
Martins, V.S., Kaleita, A.L., Gelder, B.K., Abe, C.A., 2020. Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image

classification at high spatial resolution. ISPRS J. Photogramm. Remote Sens. 168, 56–73.
Martins, V.S., Roy, D.P., Huang, H., Boschetti, L., Zhang, H.K., Yan, L., 2022. Deep learning high resolution burned area mapping by transfer learning from Landsat-8 to PlanetScope. Remote Sens. Environ. 280, 113203.
Masolele, R.N., De Sy, V., Herold, M., Marcos, D., Verbesselt, J., Gieseke, F., Martius, C., 2021. Spatial and temporal deep learning methods for deriving land-use following deforestation: a pan-tropical case study using Landsat time series. Remote Sens. Environ. 264, 112600.
Maxwell, S.K., Sylvester, K.M., 2012. Identification of "ever-cropped" land (1984–2010) using Landsat annual maximum NDVI image composites: Southwestern Kansas case study. Remote Sens. Environ. 121, 186–195.
Maxwell, S.K., Schmidt, G.L., Storey, J.C., 2007. A multi-scale segmentation approach to filling gaps in Landsat ETM+ SLC-off images. Int. J. Remote Sens. 28 (23), 5339–5356.
Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., Vihervaara, P., 2021. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. Remote Sens. Environ. 256, 112322.
Mellor, A., Boukir, S., Haywood, A., Jones, S., 2015. Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. ISPRS J. Photogramm. Remote Sens. 105, 155–168.
Neumaier, A., 1998. Solving ill-conditioned and singular linear systems: a tutorial on regularization. SIAM Rev. 40 (3), 636–666.
Nowlan, S.J., Hinton, G.E., 1992. Simplifying neural networks by soft weight-sharing. Neural Comput. 4 (4), 473–493.
Pahlevan, N., Chittimalli, S.K., Balasubramanian, S.V., Vellucci, V., 2019. Sentinel-2/Landsat-8 product consistency and implications for monitoring aquatic systems. Remote Sens. Environ. 220, 19–29.
Pascanu, R., Gulcehre, C., Cho, K., & Bengio, Y. (2013). How to construct deep recurrent neural networks. *arXiv preprint arXiv:1312.6026*.
Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. Remote Sens. 11 (5), 523.
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, 583–595.
Potapov, P.V., Turubanova, S.A., Hansen, M.C., Adusei, B., Broich, M., Altstatt, A., Justice, C.O., 2012. Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM+ data. Remote Sens. Environ. 122, 106–116.
Qian, N., 1999. On the momentum term in gradient descent learning algorithms. Neural Netw. 12 (1), 145–151.
Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., Dollár, P., 2020. Designing network design spaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10428–10436.
Ronneberger, O., Fischer, P., Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, pp. 234–241.
Rosentreter, J., Hagensieker, R., Waske, B., 2020. Towards large-scale mapping of local climate zones using multitemporal Sentinel 2 data and convolutional neural networks. Remote Sens. Environ. 237, 111472.
Roy, D.P., Yan, L., 2020. Robust Landsat-based crop time series modelling. Remote Sens. Environ. 238, 110810.
Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M., Zhang, C., 2010. Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. Remote Sens. Environ. 114 (1), 35–49.
Roy, D.P., Qin, Y., Kovalskyy, V., Yan, L., 2014. Conterminous United States demonstration and characterization of MODIS-based Landsat ETM+ atmospheric correction. Remote Sens. Environ. 140, 433–449.
Roy, D.P., Zhang, H.K., Ju, J., Kovalskyy, V., 2016. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. Remote Sens. Environ. 176, 255–271.
Roy, D.P., Li, Z., Zhang, H.K., Huang, H., 2020. A conterminous United States analysis of the impact of Landsat 5 orbit drift on the temporal consistency of Landsat 5 Thematic Mapper data. Remote Sens. Environ. 240, 111701.
Roy, D.P., Huang, H., Houborg, R., Martins, V.S., 2021. A global analysis of the temporal availability of PlanetScope high spatial resolution multi-spectral imagery. Remote Sens. Environ. 264, 112586.
Ruder, S., 2016. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv*:1609.04747.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.
Rußwurm, M., Körner, M., 2020. Self-attention for raw optical satellite time series classification. ISPRS J. Photogramm. Remote Sens. 169, 421–435.
Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D., Tavenard, R., 2023. End-to-end learned early classification of time series for in-season crop type mapping. ISPRS J. Photogramm. Remote Sens. 196, 445–456.
Sahiner, B., Chan, H.P., Petrick, N., Wei, D., Helvie, M.A., Adler, D.D., Goodsitt, M.M., 1996. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. IEEE Trans. Med. Imaging 15 (5), 598–610.
Scardapane, S., Wang, D., 2017. Randomness in neural networks: an overview. Wiley Interdisc. Rev. Data Min. Knowl. Discov. 7 (2), e1200.
Schug, F., Frantz, D., Okujeni, A., van Der Linden, S., Hostert, P., 2020. Mapping urban-rural gradients of settlements and vegetation at national scale using Sentinel-2 spectral-temporal metrics and regression-based unmixing with synthetic training data. Remote Sens. Environ. 246, 111810.
Shin, H.C., Roth, H.R., Gao, M., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging 35 (5), 1285–1298.
Sideris, M.G., Li, Y.C., 1993. Gravity field convolutions without windowing and edge effects. Bull. Geodesique 67 (2), 107–118.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*:1409.1556.
Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Žust, L., 2022. Cloud Mask Intercomparison eXercise (CMIX): an evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. Remote Sens. Environ. 274, 112990.
Small, C., 2002. Multitemporal analysis of urban reflectance. Remote Sens. Environ. 81 (2–3), 427–442.
Small, C., 2005. A global analysis of urban reflectance. Int. J. Remote Sens. 26 (4), 661–681.
Srivastava, S., Vargas-Muñoz, J.E., Tuia, D., 2019. Understanding urban landuse from the above and ground perspectives: a deep learning, multimodal solution. Remote Sens. Environ. 228, 129–143.
Stoian, A., Poulain, V., Inglada, J., Poughon, V., Derksen, D., 2019. Land cover maps production with high resolution satellite image time series and convolutional neural networks: adaptations and limits for operational systems. Remote Sens. 11 (17), 1986.
Sulla-Menashe, D., Gray, J.M., Abercrombie, S.P., Friedl, M.A., 2019. Hierarchical mapping of annual global land cover 2001 to present: the MODIS Collection 6 Land Cover product. Remote Sens. Environ. 222, 183–194.
Sun, L., Gao, F., Xie, D., Anderson, M., Chen, R., Yang, Y., Chen, Z., 2021. Reconstructing daily 30 m NDVI over complex agricultural landscapes using a crop reference curve approach. Remote Sens. Environ. 253, 112156.
Tan, M., Le, Q., 2019. Efficientnet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. PMLR, pp. 6105–6114.
Thorp, K.R., Drajat, D.E.N.A., 2021. Deep machine learning with Sentinel satellite data to map paddy rice production stages across West Java, Indonesia. Remote Sens. Environ. 265, 112679.
Tong, X.Y., Xia, G.S., Lu, Q., Zhang, L., 2020. Land-cover classification with high-resolution remote sensing images using transferable deep models. Remote Sens. Environ. 237, 111322.
Townshend, J.G., 1992. Land cover. Int. J. Remote Sens. 13 (6–7), 1319–1328.
Tran, K.H., Zhang, H.K., McMaine, J.T., Zhang, X., Luo, D., 2022. 10 m crop type mapping using Sentinel-2 reflectance and 30 m cropland data layer product. Int. J. Appl. Earth Obs. Geoinf. 107, 102692.
Turkoglu, M.O., D'Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K., Wegner, J. D., 2021. Crop mapping from image time series: deep learning with multi-scale label hierarchies. Remote Sens. Environ. 264, 112603.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Proces. Syst. 30.
Wang, S., Di Tommaso, S., Faulkner, J., Friedel, T., Kennepohl, A., Strey, R., Lobell, D.B., 2020. Mapping crop types in southeast India with smartphone crowdsourcing and deep learning. Remote Sens. 12 (18), 2957.
Wang, H., Zhang, X., Du, S., Bai, L., Liu, B., 2022. Mapping Annual Urban Evolution Process (2001–2018) at 250 m: a normalized multi-objective deep learning regression. Remote Sens. Environ. 278, 113088.
Weiss, G.M., Provost, F., 2003. Learning when training data are costly: the effect of class distribution on tree induction. J. Artif. Intell. Res. 19, 315–354.
Wickham, J., Stehman, S.V., Sorenson, D.G., Gass, L., Dewitz, J.A., 2021. Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States. Remote Sens. Environ. 257, 112357.
Woodcock, C.E., Macomber, S.A., Pax-Lenney, M., Cohen, W.B., 2001. Monitoring large areas for forest change using Landsat: generalization across space, time and Landsat sensors. Remote Sens. Environ. 78 (1–2), 194–203.
Wulder, M.A., White, J.C., Loveland, T.R., Woodcock, C.E., Belward, A.S., Cohen, W.B., Roy, D.P., 2016. The global Landsat archive: status, consolidation, and direction. Remote Sens. Environ. 185, 271–283.
Wulder, M.A., Coops, N.C., Roy, D.P., Hermosilla, T., 2018. Land cover 2.0. Int. J. Remote Sens. 39 (12), 4254–4284.
Wulder, M.A., Roy, D.P., Radeloff, V.C., Loveland, T.R., Anderson, M.C., Johnson, D.M., Cook, B.D., 2022. Fifty years of Landsat science and impacts. Remote Sens. Environ. 280, 113195.
Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., Lin, T., 2020. DeepCropMapping: a multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. Remote Sens. Environ. 247, 111946.
Yan, L., Roy, D.P., 2015. Improved time series land cover classification by missing-observation-adaptive nonlinear dimensionality reduction. Remote Sens. Environ. 158, 478–491.
Yan, L., Roy, D.P., 2020. Spatially and temporally complete Landsat reflectance time series modelling: the fill-and-fit approach. Remote Sens. Environ. 241, 111718.
Yang, L., Huang, C., Homer, C.G., Coan, M.J., 2003. An approach for mapping large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. Can. J. Remote. Sens. 29, 230–240.
Yang, J., Zhao, Y.Q., Chan, J.C.W., 2017. Learning and transferring deep joint spectral–spatial features for hyperspectral classification. IEEE Trans. Geosci. Remote Sens. 55 (8), 4729–4742.
Yang, L., Yang, Y., Yang, J., Zhao, N., Wu, L., Wang, L., Wang, T., 2022. FusionNet: a convolution–transformer fusion network for hyperspectral image classification. Remote Sens. 14 (16), 4066.

Yuan, Y., Lin, L., 2021. Self-supervised pretraining of transformers for satellite image time series classification. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 14, 474–487.
Yuan, Q., Shen, H., Li, T., Zhang, L., 2020. Deep learning in environmental remote sensing: achievements and challenges. Remote Sens. Environ. 241, 111716.
Zhai, Y., Roy, D.P., Martins, V.S., Zhang, H.K., Yan, L., Li, Z., 2022. Conterminous United States Landsat-8 top of atmosphere and surface reflectance tasseled cap transformation coefficients. Remote Sens. Environ. 274, 112992.
Zhang, H.K., Roy, D.P., 2017. Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. Remote Sens. Environ. 197, 15–34.
Zhang, H.K., Roy, D.P., Kovalskyy, V., 2016. Optimal solar geometry definition for global long-term Landsat time-series bidirectional reflectance normalization. IEEE Trans. Geosci. Remote Sens. 54 (3), 1410–1418.
Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., Atkinson, P.M., 2018. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. ISPRS J. Photogramm. Remote Sens. 140, 133–144.
Zhang, C., Harrison, P.A., Pan, X., Atkinson, P.M., 2020. Scale Sequence Joint Deep Learning (SS-JDL) for land use and land cover classification. Remote Sens. Environ. 237, 111593.
Zhang, C., Bengio, S., Hardt, M., Vinyals, O., 2021. Understanding deep learning (still) requires rethinking generalization. Commun. ACM 64 (3), 107–115.
Zhao, H., Duan, S., Liu, J., Sun, L., Reymondin, L., 2021. Evaluation of five deep learning models for crop type mapping using Sentinel-2 time series images with missing information. Remote Sens. 13 (14), 2790.
Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430–443.
Zhou, Q., Tollerud, H., Barber, C., Smith, K., Zelenak, D., 2020. Training data selection for annual land cover classification for the land change monitoring, assessment, and projection (LCMAP) initiative. Remote Sens. 12 (4), 699.