Remote Sensing of Environment 233 (2019) 111411

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Deep learning based winter wheat mapping using statistical data as ground references in Kansas and northern Texas, US

Liheng Zhong[a,*], Lina Hu[b], Hang Zhou[c], Xin Tao[d]

[a] Descartes Labs, Inc., San Francisco, CA 94107, United States
[b] Department of Sociology, Tsinghua University, Beijing 100084, China
[c] Descartes Labs, Inc., Santa Fe, NM 87501, United States
[d] Department of Geography, University at Buffalo, The State University of New York, Buffalo, NY 14261, United States

ARTICLE INFO

Edited by Emilio Chuvieco

Keywords:
Crop classification
Deep learning
Artificial neural network
Convolutional neural network
MODIS
Winter wheat
USDA Quick Stats

ABSTRACT

Winter wheat is a major staple crop and it is critical to monitor winter wheat production using efficient and automated means. This study proposed a novel approach to produce winter wheat maps using statistics as the training targets of supervised classification. Deep neural network architectures were built to link remotely sensed image series to the harvested areas of individual administrative units. After training, the resultant maps were generated using the activations on a middle layer of the deep model. The direct use of statistical data to some extent alleviates the shortage of ground samples in classification tasks and provides an opportunity to utilize a wealth of statistical records to improve land use mapping. The experiments were carried out in Kansas and Northern Texas during 2001–2017. For each study area the goal was to create winter maps that are consistent with USDA county-level statistics of harvested areas. The trained deep models automatically identified the seasonal pattern of winter wheat pixels without using pixel-level reference data. The winter wheat maps were compared with the Cropland Data Layer (CDL) for years when the CDL is available. In Kansas where the winter wheat extent of the CDL has high reported accuracy and agrees well with county statistics, the maps produced from the deep model was evaluated using the CDL as an independent test set. Northern Texas was selected as an example where the winter wheat area of the CDL is very different from official statistics, and the maps by the deep model enabled a map-to-map comparison with the CDL to highlight the areas of discrepancy. Visual representation of the deep model behaviors and recognized patterns show that deep learning is an automated and robust means to handle the variability of winter wheat seasonality without the need of manual feature engineering and intensive ground data collection. Showing the possibility of generating maps solely from regional statistics, the proposed deep learning approach has great potential to fill the historical gaps of conventional sample-based classification and extend applications to areas where only regional statistics are available. The flexible deep network architecture can be fused with various statistical datasets to fully employ existing sources of data and knowledge.

1. Introduction

While remote sensing studies provide efficient means to map crop type and extent, the availability of ground reference data is one of the critical limiting factors on the applicability of crop mapping approaches (Song et al., 2017). Crop classification relies on reference data collected via field work or visual image interpretation to develop classification algorithms and acquire agricultural information. The shortage of high-quality ground samples is a major challenge to regular and rapid monitoring of cropland, as it is time-consuming and expensive to collect reference information to train the classifiers for crop mapping, especially over large areas and long periods (Phalke and Özdoğan, 2018). Although numerous studies have contributed to the field of crop classification and a variety of map products have been published, training samples are rarely distributed publicly. For example, the United States Department of Agriculture (USDA) releases the Cropland Data Layer (CDL) on an annual basis (Boryan et al., 2011), but the farm-level land use in the Farm Service Agency's Common Land Unit dataset can only be accessed internally by the institute.

There are various methods developed to mitigate the dependency on ground samples in crop mapping. The general idea is to apply training statistics to an extended spatial-temporal range without severe loss of

* Corresponding author at: Room 233, 1416 9th Street, Sacramento, CA 95814, United States.
E-mail address: lihengzhong@berkeley.edu (L. Zhong).

https://doi.org/10.1016/j.rse.2019.111411
Received 6 December 2018; Received in revised form 12 August 2019; Accepted 2 September 2019

accuracy, or "signature extension". The first method is to train the classifier with ground samples in one or a few years and then apply the trained classifier to another year, which eliminates the need of repeated ground data collection year by year (Zhong et al., 2014; Zhong et al., 2016b; Massey et al., 2017). The method requires multi-temporal observations at a relatively high frequency to extract images at the same growing stage or derive phenological metrics with cross-year consistency. Similarly, it is possible to extend trained algorithms spatially to regions other than the training area to improve the efficiency of ground sample use. Another method is to apply automated algorithms to all years and study areas (Zhong et al., 2012; Thenkabail and Wu, 2012; Zhong et al., 2016a; Xiong et al., 2017). The automated algorithms are usually developed by intensive exploratory analysis and feature engineering based on rules summarized from ground reference information, and the accuracy depends on expert experience and the availability of local agricultural knowledge. So far, ground samples at point or parcel (object) level are still the main data source for all categories of classification approaches (supervised/unsupervised/rule-based). Although some of the existing studies notably reduce the cost of data collection in crop mapping, the lack of ground references remains a challenge. Moreover, for some applications ground reference collection is infeasible, for example, retrospective mapping for a historical period.

Agricultural statistics collected and distributed by public and private agencies are a rich source of information on crop production with great spatial coverage and temporal continuity. Previous studies created crop maps based on the statistical records at regional or country level (Ramankutty et al., 2008; Monfreda et al., 2008). For major crop production areas, statistical data are increasingly available at the level of fine administrative unit, for example, county-level statistics of various crop types in the US are published by the USDA National Agricultural Statistics Service (NASS, WWW1, n.d.), and municipal-level statistics of Brazilian agriculture are published by Brazilian Institute of Geography and Statistics on an annual basis (WWW2, n.d.). Although the statistical datasets provide information on the general spatial distribution of croplands, agricultural statistics are rarely employed in remote sensing classification to produce pixelwise maps. Most classifiers cannot take statistics directly as input data for training or validation. Even the fine spatial unit of statistics like county or municipal corresponds to a large and varying number of pixels on remotely sensed images, which is subject to the curse of dimensionality and is likely to result in an ill-posed problem. Per-pixel crop maps are useful in various applications such as crop yield forecast (Xin et al., 2013; Sakamoto et al., 2014), water use estimate (Bastiaanssen et al., 1998; Allen et al., 2007; Tang et al., 2009; Fisher et al., 2017), carbon cycle modeling (Irwin and Geoghegan, 2001; Harou et al., 2009), and economic modeling (Baret et al., 2007; Searchinger et al., 2008; Marsden et al., 2013). Existing statistical datasets may contribute a great wealth of information to crop mapping, but it is still challenging to combine statistics with remote sensed observations and transfer the knowledge to the pixel level.

In this study, we utilized a deep learning based approach to jointly utilize agricultural statistics and 250 m pixelwise image data and generate 250 m crop maps using only statistical records as the training set. Deep learning is deemed as a recent breakthrough technology in machine learning including the research field of remote sensing classification (Zhu et al., 2017). Deep neural networks are advantageous for automated extraction and identification of complex patterns in large datasets (Chen et al., 2014a; Li et al., 2016; Wan et al., 2017; Mou and Zhu, 2018; Mou et al., 2018b). Various specialized architectures have been designed in past studies to handle patterns in remotely sensed images, including spatial patterns in high resolution images (Chen et al., 2014b; Penatti et al., 2015; Hu et al., 2015b; Kampffmeyer et al., 2016; Sherrah, 2016; Audebert et al., 2018; Maggiori et al., 2017; Li et al., 2017a; Volpi and Tuia, 2017; Kussul et al., 2017; Marcos et al., 2018; Marmanis et al., 2018), spectral patterns in hyperspectral images (Hu et al., 2015a; Li et al., 2017b; Guidici and Clark, 2017; Lyu et al., 2018), and temporal patterns in multi-temporal image series (Lyu et al., 2016; Rußwurm and Körner, 2017; Mou et al., 2018a; Rußwurm and Körner, 2018). Although not tested so far, deep learning may provide a unique opportunity to utilize agricultural statistics directly as training data. By using a deep architecture with an ordered combination of many specialized layers, deep neural network models can regulate the solution space with customized model constraints, achieve high algorithmic stability and generalization, learn complex patterns from limited data, and to some extent overcome the curse of dimensionality (Zhang et al., 2018).

It is a new classification task to map crop extent using statistics for training. The present study aimed to perform the task with county-level statistics from the USDA. The experiments were conducted in two winter wheat production areas where the CDL is available for pixelwise accuracy assessment. The rest of the paper is organized as follows. Section 2 describes the input data and the method to develop deep learning architectures, optimize the models, and evaluate the classification maps. Section 3 presents the results. Section 4 then interprets the behavior of the optimized model to understand how it works and discusses the feasibility of using statistics as the reference data in classification applications. Section 5 concludes the paper.

2. Method and materials

2.1. Study areas

The study was carried out in two areas, the whole state of Kansas and the northern region of Texas (Fig. 1). Kansas is the largest production state of winter wheat in the U.S. There are 105 counties in Kansas, most of which are intensively cultivated by winter wheat, corn, soybean and other crops. Northern Texas includes 67 counties in four agricultural districts (Northern High Plains, Northern Low Plains, Southern High Plains, and Southern Low Plains) and is the major winter wheat production area of the Texas state. The climate of the study areas is semi-arid in the west and humid subtropical in the east with the majority of precipitation occurring in spring and summer. Annual precipitation of Kansas ranges from ~400 mm to ~1100 mm, and Northern Texas from ~400 mm to ~700 mm. Winter wheat is planted in fall and harvested in early-summer. The start of the growing season is from September to October, and the end is from mid-May to mid-July.

While Kansas and Northern Texas are both important growing areas of winter wheat in the US, they were selected to conduct independent experiments as two extreme cases. The CDL of Kansas has high self-reported accuracies for winter wheat (user's accuracy from 92.3% to 96.8%, and producer's accuracy from 90.0% to 96.3%), and the total winter wheat acreage of the CDL only differs from USDA NASS statistics by 3.6% on average. Crop acreage in the NASS statistics has been widely used in agricultural applications and is considered as a reliable reference. Therefore, the CDL of Kansas can be used as an accurate raster map for pixel-wise accuracy assessment. By contrast, the CDL of Northern Texas does not agree well with the NASS statistics. Although the self-reported accuracies of the CDL suggest moderate error rates (user's accuracy of winter wheat varies from 81.9% to 90.6%, and producer's accuracy from 86.5% to 93.3%), the total winter wheat acreage of the CDL is 155.2% larger than that of the NASS statistics on average. In the development of the CDL and also in the crop mapping efforts using the CDL as training data, the classifier trained with pixel-level samples has difficulty in ensuring consistency with total cropped areas at the county level. We took Northern Texas as an example to generate per-pixel maps that are consistent with the statistical data and made comparison with the CDL.

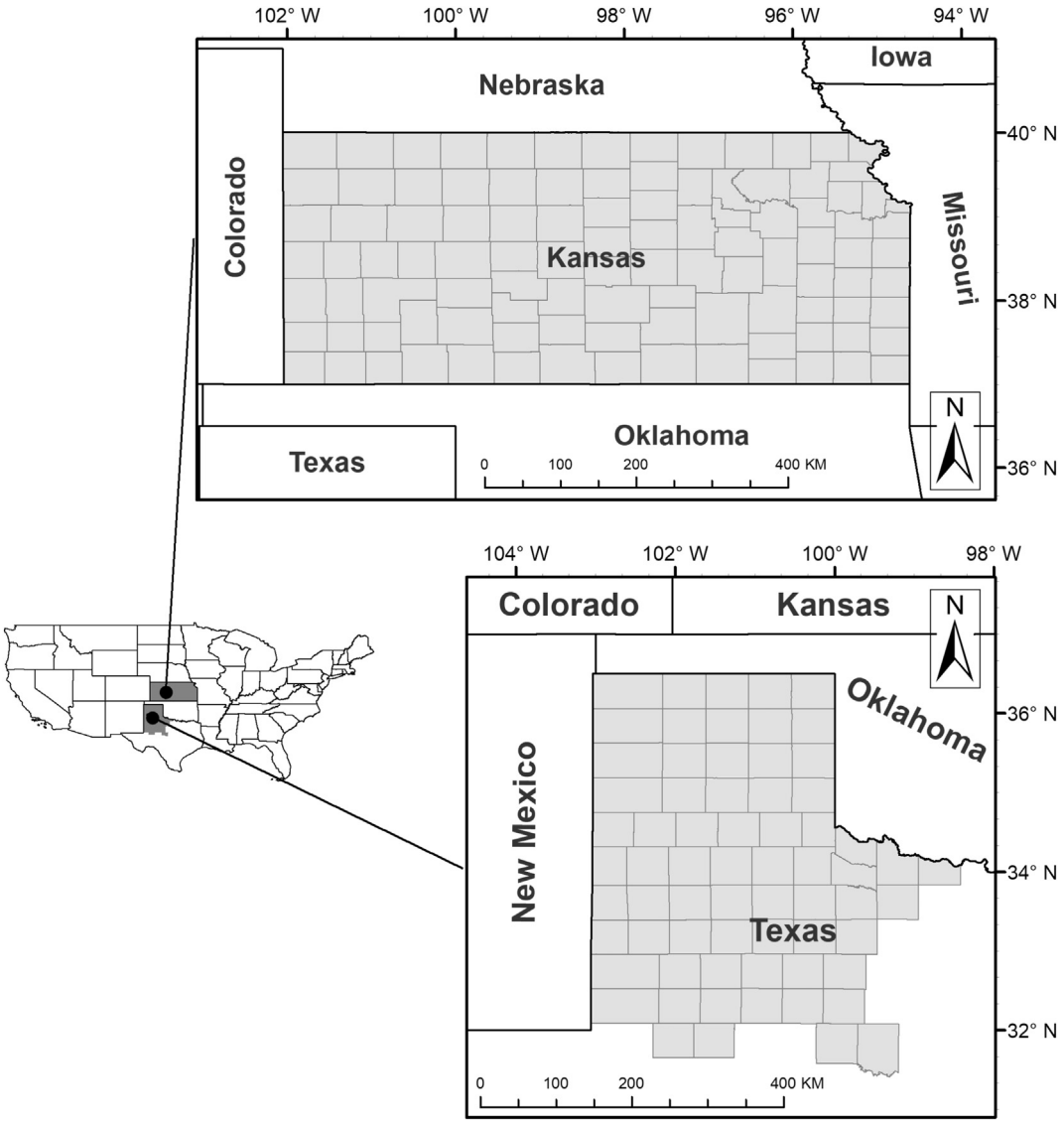

Fig. 1. One study area is the state of Kansas (105 counties), and the other is the 67 counties in four agricultural districts in northern Texas.

2.2. Data

2.2.1. MODIS imagery

The primary input data are MODerate-resolution Imaging Spectroradiometer (MODIS) product MOD13Q1 Collection 6, 16-day Vegetation Index at 250 m resolution. The MOD13Q1 product includes Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index, reflectance, quality and information bands since February 2000. The composite interval is 16 days starting from the first day of each year, and each year has 23 images. The product is in the sinusoidal projection which is equal-area. MODIS observations are subject to sensor degradation issue especially for long-term applications, and the use of Collection 6 data largely eliminates the negative bias (Wang et al., 2012; Lyapustin et al., 2014; Sayer et al., 2015). In this study, we attempted to explore the viability of identifying winter wheat with NDVI. NDVI is one of the most popular vegetation indices to characterize seasonal crop growth and is available from many other sensors. The quality band was employed to exclude cloud-covered observations in the NDVI series. Gaps were filled by per-pixel linear interpolation using the closest high-quality observations before and after the gap date. Gaps of the 16-day composite are short and rare thanks to the daily revisit frequency of the MODIS platform. By visually inspecting the filled NDVI time series and the frequency of high-quality observations in the study areas, we considered the simple gap-filling method as sufficient to depict the seasonality of the NDVI time series (Kandasamy et al., 2013). Snow may occur in the growing season of winter wheat. In our approach the algorithm was trained by input time series to learn to reduce the influence by snow cover, and no specific snow masks were developed.

MOD13Q1 images were clipped using the boundaries of counties in the study areas. For each county in each year, a corresponding data cube with a shape of *width* by *height* by 23 was created. *Width* and *height* are spatial dimensions and 23 is the number of observations in the NDVI time series. In our study areas, the maximum *width* and *height* of all counties are 601 and 296 in Kansas, and 325 and 568 in Northern Texas. To cover the growing season of winter wheat, the beginning of the annual time series was chosen as August 29th (August 28th in leap years), the start of the 16th image in the calendar year before the year of harvest. The earliest growing season that the MODIS data could completely cover is 2000–2001, and MODIS images from August 2000 to August 2017 were acquired and processed. Margin pixels outside of the county boundary were set to value −1.0 throughout the year. The dynamic range of NDVI is appropriate for deep learning models and further scaling is unnecessary.

2.2.2. USDA statistics

USDA NASS continuously publishes statistics of various agricultural commodities acquired through the survey and the census programs. The data item used in this study is harvested acres of winter wheat per county retrieved from the NASS Quick Stats portal (WWW1, n.d.). The census is a complete count of croplands which is taken every five years. For census years the census data were used and for other years the annual survey data were used to generate an annual statistical record. By the time of this study, census data in 2002, 2007, and 2012 are available within the observation period of MOD13Q1. The final study years are from 2001 to 2017, giving 1785 possible data points for Kansas (17 years by 105 counties) and 1139 for Northern Texas (17 years by 67 counties). Counties with relatively small harvested areas may be combined by the NASS in the statistical record and are not useful for county-level analysis. As a result, the county-level statistics are not available for every county in every year, and the actual numbers of county & year combinations are 1641 for Kansas and 956 for Northern Texas. The harvested acres were converted into the areal percentage of winter wheat of each county. Then MOD13Q1 data cubes of each county were associated with the corresponding value of the winter wheat percentage as a pair of independent/dependent variables. For example, the 23 MOD13Q1 images from August 28th, 2000 to August 28th, 2001 clipped to a county's boundary were associated with the winter wheat percentage of the county in year 2001. The reference percentage data along with the associated MOD13Q1 images were split into training (years 2001–2016) and validation (year 2017) sets. Model parameters (weights) were trained by the training set, and the validation set was for network architecture searching and hyper-parameter selection to reduce over-fitting. The network model was trained and tuned on a per-county basis using the training set and the validation set, while the final results of per-pixel classification maps were extracted from a middle layer of the network (explained in Section 2.3) on which evaluation was conducted. Therefore, the test set of this study was not split by year but included per-pixel winter wheat coverage of all years.

2.2.3. USDA Cropland Data Layer

The Cropland Data Layer (CDL) was used as the reference data for pixelwise assessment. The CDL is a raster, georeferenced, crop-specific land-cover map created using satellite imagery and extensive agricultural ground truth by the USDA, NASS on an annual basis. Since 2008, the CDL program has provided cropland area estimates and digital products of spatial distribution at a resolution of 30 m for all states in the continental US. Kansas has two more years of state-level coverage at a resolution of 56 m prior to the full CDL of the continental US, and the record of Kansas CDL starts from 2006. The CDL was acquired in the Albers equal-area conic projection specified by the USDA for contiguous US. Each pixel of the CDL is identified as one of the 131 crop-specific land-cover classes. Beginning in 2008, the CDL also includes per-pixel predicted confidence of the given classification, with 0 being the least confident and 100 the most confident. The confidence value is a percent measure of how well the decision to identify a pixel as a specific category fits the rules in the decision tree classifier that was used to produce the CDL (Boryan et al., 2011).

All land-cover classes of the CDL containing winter wheat were combined and extracted: winter wheat (single crop), double crop winter wheat/soybeans, double crop winter wheat/corn, double crop winter wheat/sorghum, and double crop winter wheat/cotton. All the double crop classes consist of winter wheat and a summer crop. The producer's accuracy of winter wheat in the Kansas CDL ranges between 88% and 96%, and the user's accuracy between 91% and 96% as reported in the metadata of the CDL. Compared to harvested acres of winter wheat in the statistics from survey or census, the winter wheat area estimated by the Kansas CDL is larger by ~6% with inter-annual variance. The accuracy of winter wheat in the CDL of Texas is lower. The self-reported producer's accuracy varies from 87% to 93%, and the user's accuracy from 82% to 91%. In Northern Texas the winter wheat area of the CDL is 155% higher on average (from 59% to 249%) than the survey/census statistics.

The winter wheat pixels of the 30 m CDL were re-projected to the projection of the MOD13Q1 product and aggregated to the 250 m footprint. During aggregation, the binary raster of winter wheat and other land-cover were converted into the winter wheat coverage of each MODIS pixel in percentage. When available, the confidence layer was also employed to extract high-confidence pixels at 250 m resolution by averaging 30 m confidence values of winter wheat within each 250 m footprint (other land-cover types were set to zero, and more details are in Section 2.4). For Kansas the processed CDL is available annually from 2006 to 2017, and for Northern Texas from 2008 to 2017. High-confidence pixels are from 2008 to 2017 for both the study areas. In our experiment the CDL was only used to test mapping results. Information at the pixel level from the CDL was not fed into the classification models during the training process.

2.3. Classification

As one of the most prominent characteristics of crops, the seasonal dynamics of vegetation index are the basis of crop classification studies as well as the driver of temporal representation method development (Zhong et al., 2019). Different crops exhibit different temporal profiles of phenology as manifested in the NDVI (Jakubauskas et al., 2001). Winter wheat has distinct crop calendar from other major crops in the Central US and NDVI values at some point during the growing season have been reported to be separable (Wardlow and Egbert, 2008). The shape of the NDVI temporal trajectories can be used to effectively identify winter wheat (Shao et al., 2010). To quantitatively characterize winter wheat seasonality, methods have been developed to derive phenological metrics as the input to classification algorithms (Howard and Wylie, 2014). A major source of uncertainty in phenology-based classification is the inter-annual and inter-regional variability of crop progress and conditions, which is usually addressed by supervised learning using training statistics from widely-distributed ground samples or existing land use maps across years and areas.

Unlike traditional supervised classification efforts that use pixel-level or object-level ground reference data for training, our mapping study only employed county-level statistics which do not provide detailed spatial information at the resolution of the input imagery. We designed deep learning models with specialized architectures to directly establish a relationship between remotely-sensed images and winter wheat percentage. The model input is an NDVI cube in the shape of *width* by *height* by 23, and the model output is a single value. Because the input dimension is high, it is infeasible to train traditional algorithms with only 1641 (Kansas) or 956 (Northern Texas) samples.

The general architecture of our models consists of convolutional layer(s) and/or other functional layers followed by a global average layer. Convolution can be along the spatial dimensions (width and height) and/or the temporal dimension. For flexibility the implementation of 3-dimensional convolutional layers (Conv3D) was used. When the kernel sizes of both the spatial dimensions are one, Conv3D is equivalent to one-dimensional convolution in the temporal domain. When all the three kernel sizes are larger than one, spatio-temporal convolution is applied. The global average layer is the last layer of the model which reduces the incoming matrix to a scalar by taking the average. The models were designed in a way that the shape of the incoming matrix is *width* by *height*, and values in the matrix are between 0 and 1. By such settings, the incoming matrix to the global average layer can be considered as a land cover map in which the value of each pixel represents the percentage of a certain type. After the model is fully trained, the global average layer can be removed, and the remaining sub-model will output land cover maps at the same resolution as the input MODIS images.

There are many forms of Conv3D layers and other units as the candidate components of deep models. Because of the versatility of the

specialized architectures, there is no standard procedure to search for the optimal combination of hyper-parameters and various types of layers. In this study, the implementation of Conv3D was combined with pooling layers and dropout. A variety of units and techniques were tested but not selected in the final network, for example, batch normalization (Ioffe and Szegedy, 2015) and inception modules (Szegedy et al., 2015). Various kernel sizes (window widths) of Conv3D layers were tested. Kernels for the spatial dimension are either 1-by-1 (temporal convolution only) or 3-by-3. We did not try larger spatial sizes because stacking multiple small-kernel layers is a more computational efficient way to enlarge the receptive field (Simonyan and Zisserman, 2014). Similarly, the kernel size in the temporal domain was fixed to 3, except the last Conv3D layer in the model in which the kernel size is exactly the temporal length of the input so that the output length is reduced to one in the temporal dimension. The last Conv3D layer also utilizes a sigmoid activation function to output values strictly between 0 and 1 as the crop cover percentage, unlike other Conv3D layers with ReLU (Rectified Linear Unit) as the activation function. Multiple Conv3D layers could be stacked to learn temporal/spatial features in a hierarchical manner. The first Conv3D layer has 4, 8, or 16 channels/filters and the channel number may increase when going deeper. Pooling layers were fixed as max-pooling with a window size of 2 along the temporal dimension. The use of max-pooling is to reduce the sensitivity to small shifts in crop phenology. No pooling is performed in the spatial dimensions and the spatial resolution of the input images is kept throughout the model. Dropout is a regularization technique that randomly drops some neurons in a layer during training so that the output of the layer does not rely on only a few neurons (Srivastava et al., 2014). The probability of dropping neurons was set to 20%, 30%, 40%, or 50%.

As a result, there are an extremely large number of potential network architectures and it is impossible to try them all. The selection of hyper-parameters was conducted step by step based on experience and the performance on the validation set. While the training set (years 2001–2016) was used to train network parameters (weights), the validation set (year 2017) was for the selection of network architectures and hyper-parameters. We started with a small model with only one layer of convolution along the temporal dimension, which is equivalent to per-pixel linear regression using the 23 values in the NDVI time series, the simplest learning approach. Then we generated new models by changing one or two hyper-parameters, adding a new layer, re-ordering layers, or replacing a part of the network with a more complex component. Among the new models, models with promising performance on the validation set were used as the seeds to begin a new round of searching. In this way, the tested model grew in size and complexity until classification results did not improve further.

All the three types of ANNs were trained using the Adam optimizer (Kingma and Ba, 2014). Parameters of Adam were fixed as: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1e-07$, and a learning rate decay of 0.001. The initial learning rate was set to 0.001 and the batch size was set to 4. Mean-squared-error was used as the loss function to represent the difference between reference and predicted winter wheat coverage. Model training continued until the loss on the validation set stopped decreasing. Classification models were built and evaluated using the Keras library (WWW3, n.d.) on top of Tensorflow (WWW4, n.d.).

2.4. Evaluation

While the training process is at the county level using county statistics as the input data, model evaluation is pixelwise because the main objective of this study is to map crop coverage at the resolution of the input imagery. We first performed wall-to-wall assessment for each of the study areas using the CDL. The CDL was not used as ideal ground reference for both study areas. For Kansas the CDL is considered as a reliable reference for accuracy assessment because the self-reported accuracy of the Kansas CDL is high and the total winter wheat acreage of the CDL is very close to the NASS statistics. As for Northern Texas, the total winter wheat acreage of the CDL is 155% larger than the statistics on average and the self-reported accuracy of the Texas CDL is not as high as Kansas, and so the assessment using the CDL can be seen as a pixel-wise comparison between the CDL and the winter wheat maps derived only from the NASS statistics. Winter wheat coverage was categorized as “winter wheat” or “other” pixels using the majority rule. Metrics including the overall accuracy, the producer's accuracy, the user's accuracy, and the F1 score were reported. The overall accuracy is inflated as the dataset is very unbalanced with much more non-winter-wheat areas than winter wheat, and it is reported only to provide a general measure of areal assessment. The F1 score is the harmonic mean of the producer's accuracy and the user's accuracy:

$$F1_{wheat} = \frac{1}{\frac{1}{2}\left(\frac{1}{A_{prod}} + \frac{1}{A_{user}}\right)} = \frac{2\,A_{prod} \bullet A_{user}}{A_{prod} + A_{user}}$$

where $F1_{wheat}$ is the F1 score of the winter wheat class, A_{prod} is the producer's accuracy of the class, and A_{user} is the user's accuracy of the class.

As a classification product, the CDL is subject to errors especially for the Northern Texas site where severe overestimation occurs. Furthermore, there is spatial uncertainty when aligning the CDL and the MOD13Q1 footprints to aggregate the CDL pixels into the 250-m resolution. We tested the sensitivity of classification accuracy to footprint alignment by adding a half-pixel shift (~115 m) to the MOD13Q1 footprints and comparing the shifted 250-m CDL to the same CDL aggregated without any shift. In all study years, the overall accuracy ranged from 0.917 to 0.956, and the producer's/user's accuracy from 0.742 to 0.838. The producer's accuracy is equal to the user's accuracy because the comparison was between two identical maps with different spatial aggregation. The results of the sensitivity test suggest that the impact of possible spatial uncertainty on accuracy assessment is not negligible. To reduce the influence of reference data error and spatial uncertainty during aggregation, we extracted very pure and high-quality pixels to conduct small-set assessment by using the confidence (0 to 100) raster associated with CDLs since 2008. The confidence raster of each CDL was also aggregated into the 250-m footprints using the mean value of winter wheat areas. A 250-m pixel is considered as high-confidence winter wheat if pixels in its 3-by-3 neighborhood have winter wheat confidence values larger than 90, and is considered as high-confidence other land covers if the confidence values of its 3-by-3 neighborhood pixels are zero. About 10% of the 250 m winter wheat pixels were treated as high-confidence pixels. The small set of high-confidence pixels separates other factors from inherent model errors and provides more reliable measures on the classification performance of the model. The same set of metrics as the wall-to-wall comparison was reported for the “high-confidence pixel only” evaluation.

3. Results

3.1. Model architecture

We selected two models for each study area according to the best results on the validation set: one model with temporal convolution only (which means the kernel sizes of Conv3D layers in the spatial dimension are always 1-by-1), and the other with convolution in both spatial and temporal dimensions. The former model only utilizes information in the temporal dimension or seasonality in the NDVI time series to identify winter wheat, while the latter considers additional spatial patterns of fields in classification. The difference between the two models was used to evaluate the contribution of spatial information to classification results. The two models are named as “temporal-only” and “spatio-temporal”, respectively. The selected model architectures are identical between the study areas. The temporal-only model has two temporal convolution layers, followed by max-pooling along the temporal

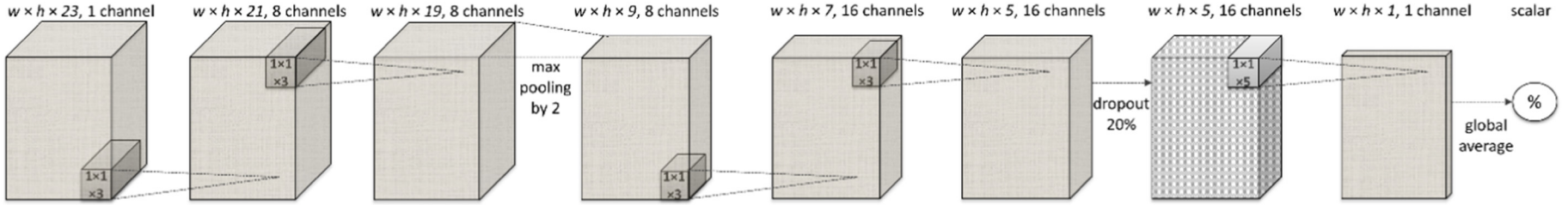

Fig. 2. Architecture of the temporal-only model. Small cubes represent convolutional kernels with kernel sizes labeled inside. Data dimensions are labeled on the top.

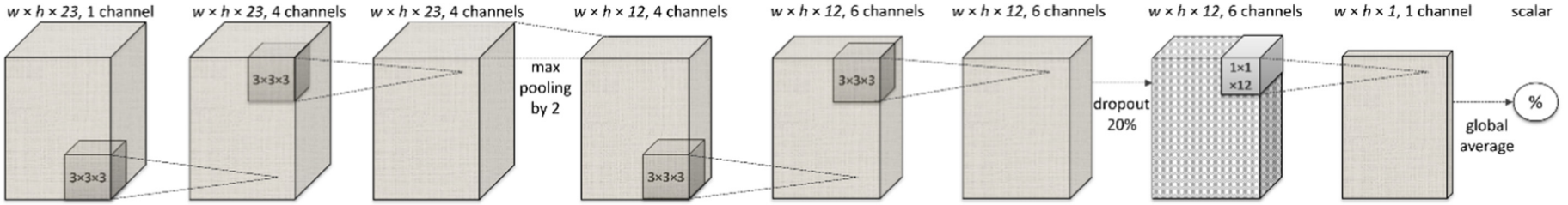

Fig. 3. Architecture of the spatiotemporal model. Small cubes represent convolutional kernels with kernel sizes labeled inside. Data dimensions are labeled on the top.

dimension, another two temporal convolution layers, dropout, a convolution layer that shrinks the temporal length to one, and finally a global average operation (Fig. 2). The architecture is fully-convolutional that reduces the input dimension of *width*-by-*height*-by-*23* to a *width*-by-*height*-by-*1* map and then a scalar regardless of the actual values of *width* and *height*. The activation function of convolutional layers is ReLU, except the last convolutional layer that uses the sigmoid activation function so that values in the *width*-by-*height*-by-*1* map are between 0 and 1.

The spatiotemporal architecture is very similar to the temporal-only model. The types and the order of layers follow the same logic (Fig. 3). The differences of the spatiotemporal model are: i) the spatiotemporal convolutional layers have a kernel size of 3-by-3-by-3, ii) padding with the same values at the edge was applied, so that the input size does not decrease after convolution, and iii) the number of channels in each Conv3D layer is smaller (4 or 6) than the temporal-only model (8 or 16). By using a smaller number of channels, replacing temporal-only convolution with spatiotemporal convolution did not considerably increase the model complexity or the total number of weights. The complexity of the models is suitable for the size of the dataset, and further adding more layers, more channels, or larger kernels worsened the training results.

3.2. Winter wheat maps

A winter wheat map was produced for each combination of years (2001–2017), study areas (Kansas and Northern Texas), and models (temporal-only and spatiotemporal). Maps since year 2006 (Kansas) or 2008 (Northern Texas) were compared with the CDL. Fig. 4 and Fig. 6 show the CDL winter wheat percentage at 250 m and the resultant maps for the whole study areas in year 2016, in which the reported quality of the CDL is relatively high. Fig. 5 and Fig. 7 provide zoom views of three sub-areas in each of the study area. In general, the spatial distributions of winter wheat in the resultant maps by both models are consistent with the CDL. Western and middle Kansas is intensively cultivated by single-cropping winter wheat with patterns following the terrain. In the southeast region of the state, the double-cropping type of winter wheat and summer soybean is common. Both the single- and double-cropping types were well depicted by the resultant maps. In Northern Texas, the major areas of winter wheat were identified by the maps, but the mapped coverage of winter wheat was apparently much lower than that of the CDL. For both the study areas, compared to the winter wheat cover mapped by the temporal-only model, winter wheat pixels by the spatiotemporal model appear to be spatially agglomerated to form smaller but denser clusters. The possible reason for the discrepancy is further discussed in later sections. Winter wheat maps of other years were also visually inspected, and the same patterns and issues were found.

3.3. Pixelwise accuracy assessment

Table 1 and Table 2 present the overall accuracy (OA), the producer's accuracy (PA), the user's accuracy (UA), and the F1 score of winter wheat maps in Kansas and Northern Texas, respectively. These metrics were evaluated using the CDL as the reference data. For Kansas, the winter wheat areas of the CDL are quite reliable as indicated by the high self-reported accuracies and the agreement with statistics from survey/census. In the wall-to-wall assessment (all pixels in the whole area were used for accuracy assessment), the temporal-only model yielded 91.9% OA, 61.4% PA, 82.3% UA, and 0.702 F1 on average, while the results by the spatiotemporal model was poorer with 87.1% OA, 57.1% PA, 58.8% UA, and 0.579 F1. When using only high-confidence pixels in assessment, all accuracies were much higher. The temporal-only model gave 99.6% OA, 93.8% PA, 94.8% UA, and 0.942 F1, and the spatiotemporal model gave 99.3% OA, 92.5% PA, 88.5% UA, and 0.903 F1. As described in Section 2.4, high-confidence pixels were extracted by considering the confidence value provided by the CDL, the purity within each 250-m pixel, and the spatial neighborhood of the pixel. The much better results in the high-confidence-pixel-only assessment suggest that the discrepancy in the wall-to-wall assessment may be caused by the inherent uncertainty of the CDL, the mixed-pixel effect at the 250-m resolution, and the uncertainty in the alignment of the CDL and the MOD13Q1 pixel footprints.

In general, for the study area of Kansas the winter wheat maps by the deep models are comparable to the CDL at 250 m resolution. The classifier of the CDL was trained by extensive ground reference data at 30 m resolution, and the Kansas CDL was reported to be one of the most accurate CDL products. Given that the deep models conducted mapping without using any pixel-level training data at a much coarser 250 m resolution, the agreement between the CDL and the maps by the deep models looks quite promising. The deep learning approach shows a great potential to produce maps for earlier years that are not covered by the CDL (in this study, years from 2001 to 2005 or 2007) and for places where there are no pixel-level ground samples available but only statistical data.

As for the winter wheat maps of Northern Texas, the UA is generally comparable to that of Kansas, but the PA is much lower. In the wall-to-wall assessment, the average PA is 28.8% by the temporal-only model and 34.7% by the spatiotemporal model. Even using high-confidence pixels for assessment, the PA values increased to only 56.7% and 56.2%, respectively. The low PA is consistent with the visual inspection in which apparent omission of winter wheat was found. The main purpose of the study in Northern Texas is to generate winter wheat maps that are consistent with the statistics and compare with the CDL to understand the spatial distribution of the large differences between the NASS statistics and the CDL. The lower accuracies evaluated using the CDL are expected when selecting the study area.

The comparison between the two deep models suggests that in both Kansas and Northern Texas the spatiotemporal model is inferior to the temporal-only model. The use of spatial convolution resulted in particularly lower UA. The spatiotemporal model utilizes textural and contextual information in addition to the temporal patterns used by the temporal-only model, but the additional spatial information contributes negatively to the pixelwise accuracy. In the next section, we will further consider the agreement with statistical data and continue discussing the possible reason.

3.4. Areal comparison

To further analyze the uncertainty in the mapping process, the mapped winter wheat areas were compared with the USDA statistics. The USDA statistics from censuses and surveys are usually recognized as a reliable data source of crop cultivation. Fig. 8 and Fig. 9 present the winter wheat acreage from the USDA statistics, the CDL, the map by the temporal-only model, and the map from the spatiotemporal model for Kansas and Northern Texas, respectively. The winter wheat areas of maps were calculated using the winter wheat percentage of 250 m pixels. For Kansas, the total winter wheat acreage of the CDL agrees well with the statistics. The percentage difference in most years is less than 5.0%, the maximum difference (in 2013) is 10.1%, and the Mean Absolute Percentage Error (MAPE) is 3.6%. The temporal-only model resulted in 13.0% maximum difference (in 2014) and 6.8% MAPE, and the spatiotemporal model resulted in 12.1% maximum difference (in 2014) and 6.4% MAPE. The annual trends of winter wheat acreage from the results of the deep models are consistent with the statistics and the CDL.

In the pixelwise accuracy assessment, the spatiotemporal model resulted in much lower user's accuracy than the temporal-only model (58.8% vs 82.3% in the wall-to-wall assessment, and 88.5% vs 94.8% in the high-confidence-pixel-only assessment). However, the total winter wheat acreage mapped by the spatiotemporal model has the same level of agreement with the statistics as the temporal-only model, if not

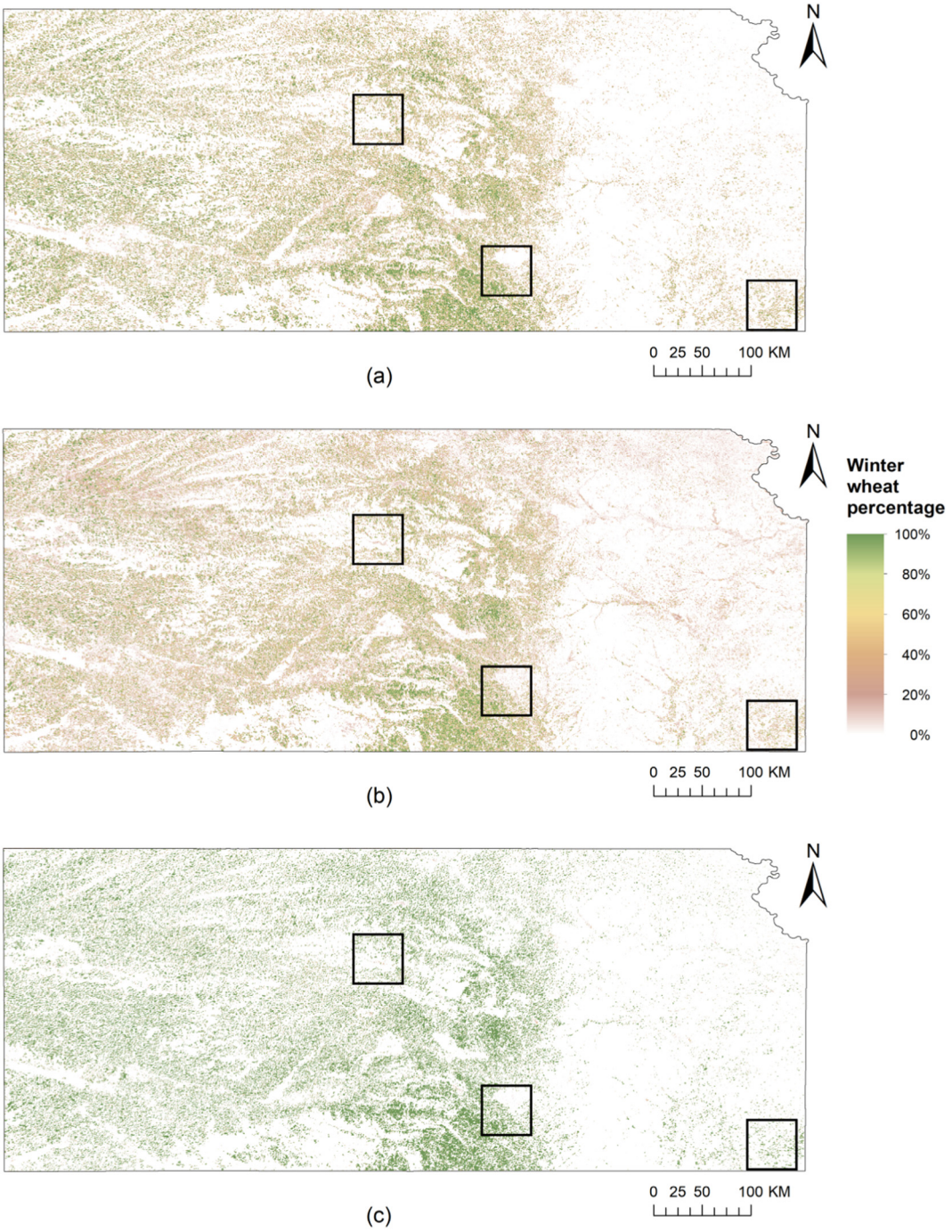

Fig. 4. Winter wheat cover of the CDL aggregated to 250 m (a) and winter wheat fuzzy maps produced by the temporal-only model (b) and the spatiotemporal model (c) for Kansas in year 2016. Squares show the extent of the zoom views in Fig. 5.

better than. When inspecting the map details like Fig. 5, we found that the parcels mapped by the spatiotemporal model were consistent with the spatial pattern of the CDL, but the parcel extent did not align well with the CDL. The mapped winter wheat pixels show an "agglomerated" pattern with small and dense clusters, which is likely to be caused by the use of spatial convolution. The spatiotemporal model includes 4 layers of 3-by-3 spatial convolution, and a pixel in the final map has a receptive field corresponding to a 9-by-9 window in the input images. In this study we aimed to explore the possibility of training classification models without pixel-level reference data. As a result, pixel-level spatial constraints were not implemented in the loss function of the spatiotemporal model, and the information carried by a pixel in the input images might "drift" to any place within the 9-by-9 receptive field rather than the middle point of the receptive field. The spatiotemporal model can utilize spatial information to improve the estimate of overall coverage, but the exact locations of the mapped pixels may be inaccurate. To benefit from the advantage of spatiotemporal convolution, at least a small set of pixel-level reference data is still needed to tune the model to learn the exact locations of the crop pixels and reduce the "pixel drift" effect.

For Northern Texas, the winter wheat acreage of the CDL is known to be much larger than that of the NASS statistics. As the maps produced by the deep models are completely trained with statistical data rather than ground samples, they are able to ensure consistency with

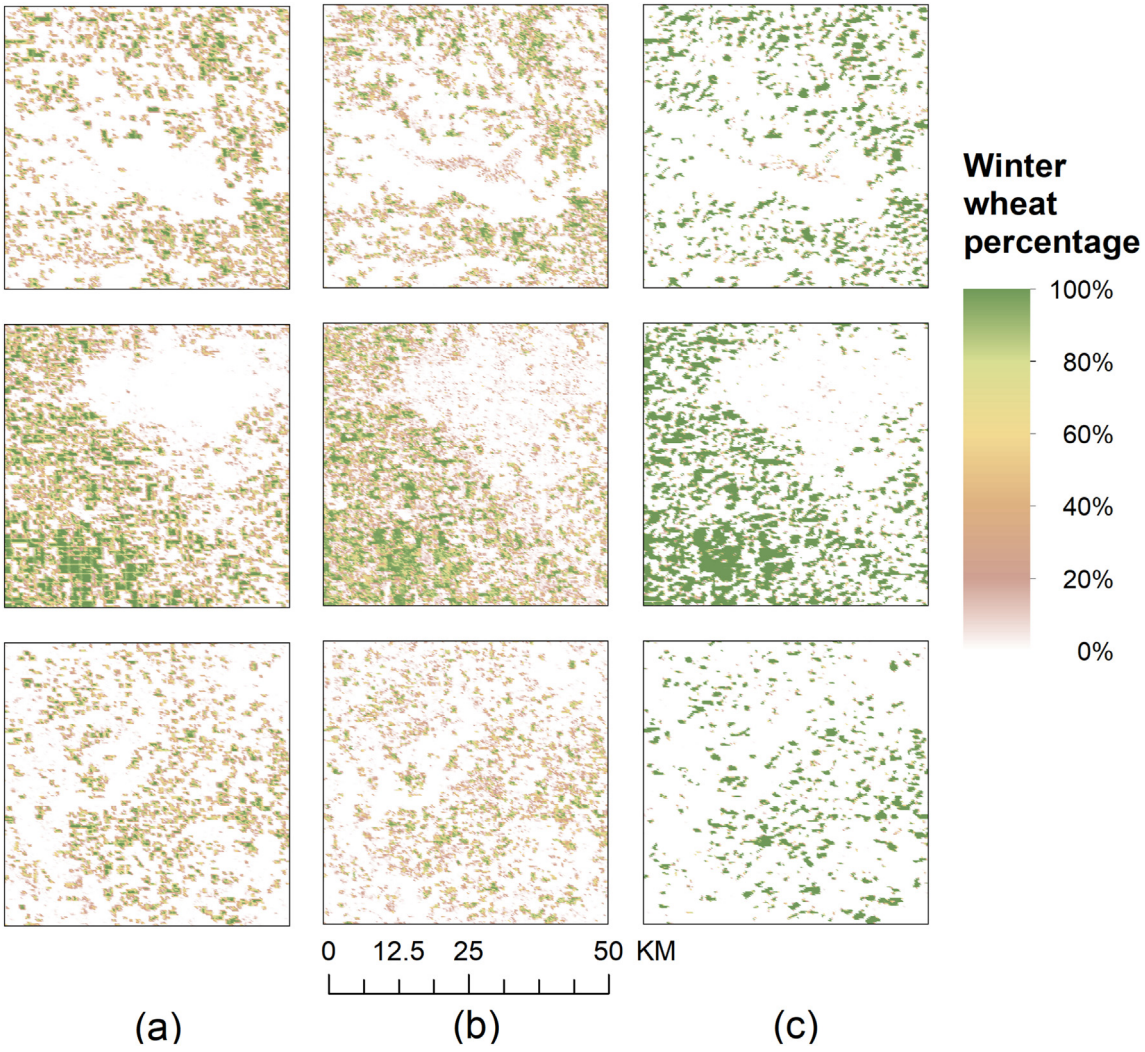

Fig. 5. Zoom views for the comparison between the CDL and the resultant maps of Kansas in 2016 in three sub-areas. Column (a) includes the CDL winter wheat percentage aggregated to 250. Columns (b) and (c) are 250 m winter wheat percentage maps by the temporal-only and the spatiotemporal models, respectively. The extents of the three sub-areas are highlighted by the squares in Fig. 4.

the NASS statistics (Fig. 9). The percentage difference of the CDL from the statistics could be as large as 248.6% (in 2013), and the MAPE is 155.2%. The temporal-only model resulted in 12.5% MAPE, the maximum difference is 33.2% (in 2008), and all differences in years other than 2008 are around or lower than 15%. The spatiotemporal model resulted in 14.9% MAPE, and the maximum difference is 27.0% (in 2011). Traditional classification algorithms are mostly trained and evaluated with pixel-level reference data (like the CDL using ground

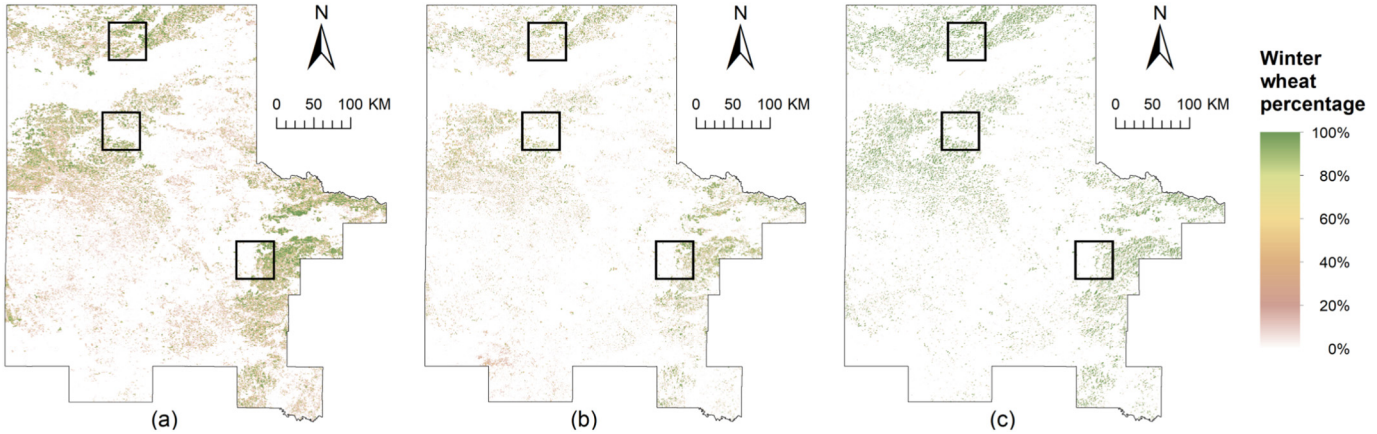

Fig. 6. Winter wheat cover of the CDL aggregated to 250 m (a) and winter wheat fuzzy maps produced by the temporal-only model (b) and the spatiotemporal model (c) for Northern Texas in year 2016. Squares show the extent of the zoom views in Fig. 7.

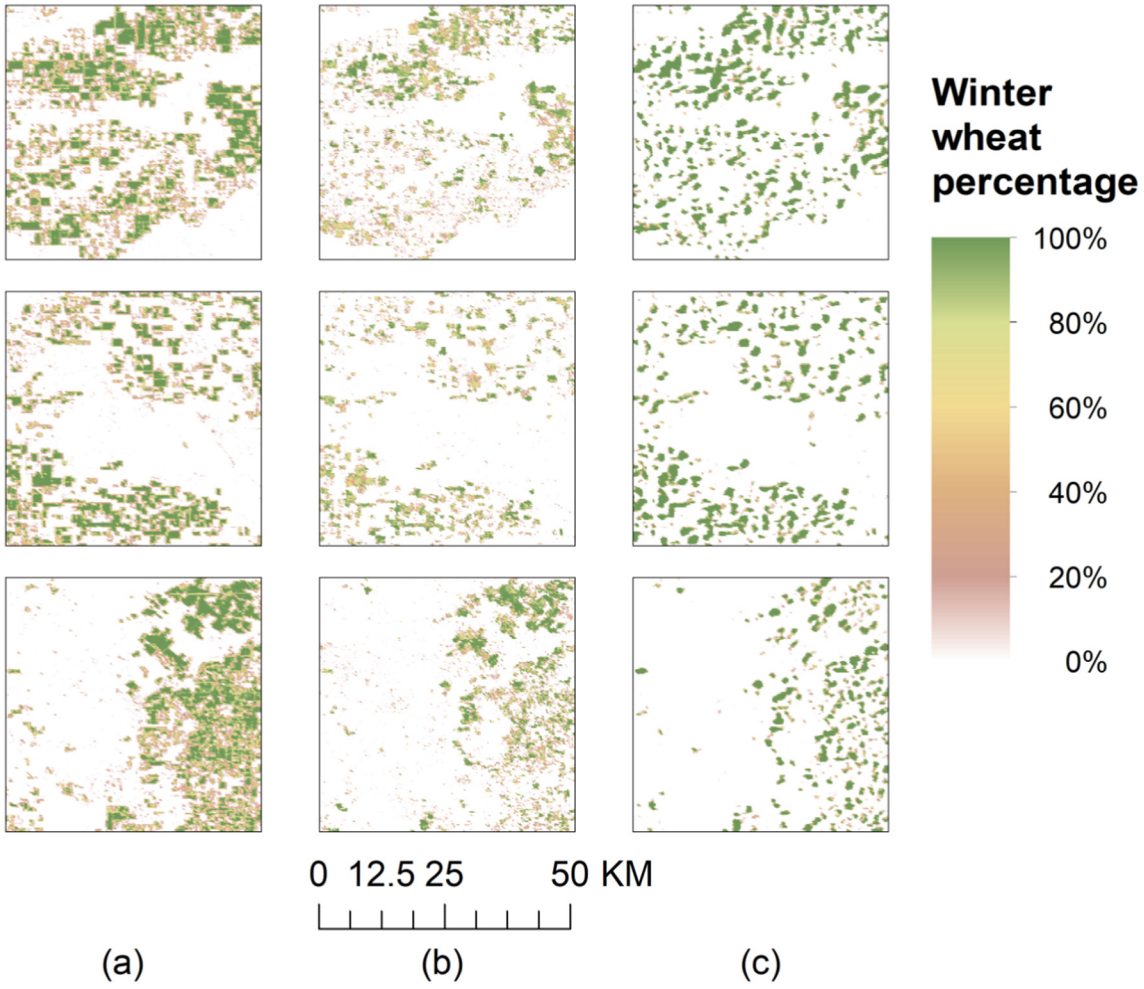

Fig. 7. Zoom views for the comparison between the CDL and the resultant maps of Northern Texas in 2016 in three sub-areas. Column (a) includes the CDL winter wheat percentage aggregated to 250. Columns (b) and (c) are 250 m winter wheat percentage maps by the temporal-only and the spatiotemporal models, respectively. The extents of the three sub-areas are highlighted by the squares in Fig. 6.

truths reported by farmers), but the classifiers hardly enforce constraints at a regional level. As a result, the ground samples used in classification may not fully represent the data distribution of the whole area and the total area from the classification products may differ from the regional statistics. By contrast, the proposed deep models are totally trained with the regional statistics and the resultant maps are supposed to be consistent with the statistical data by nature. The deep model framework is an effective approach to incorporate regional constraints into the classification process. The reason for the considerable discrepancy between the total acreage values of the CDL and the NASS statistics in Northern Texas has not been thoroughly investigated, and the maps produced by the deep models offer an opportunity to conduct local comparisons with the CDL to understand the spatial distribution of the differences.

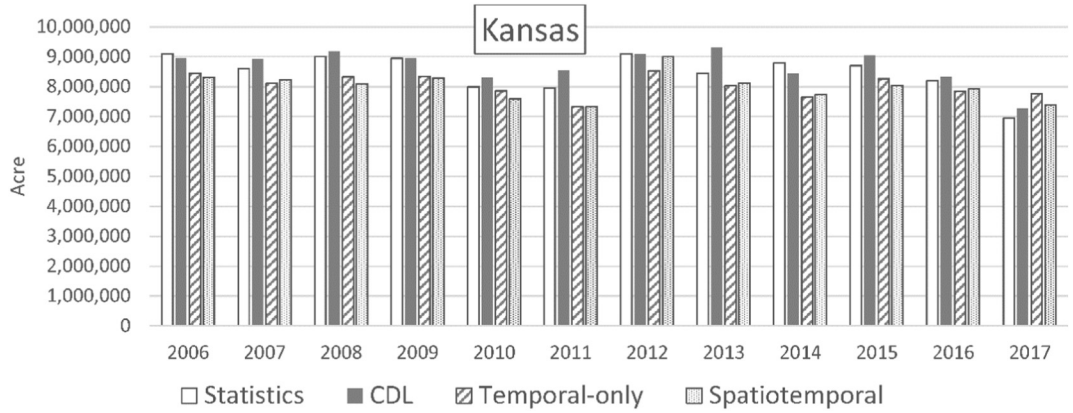

Fig. 8. Winter wheat acreage of Kansas from the USDA statistics compared with mapped acreage by the CDL, the temporal-only model, and the spatiotemporal model, in 2006–2017.

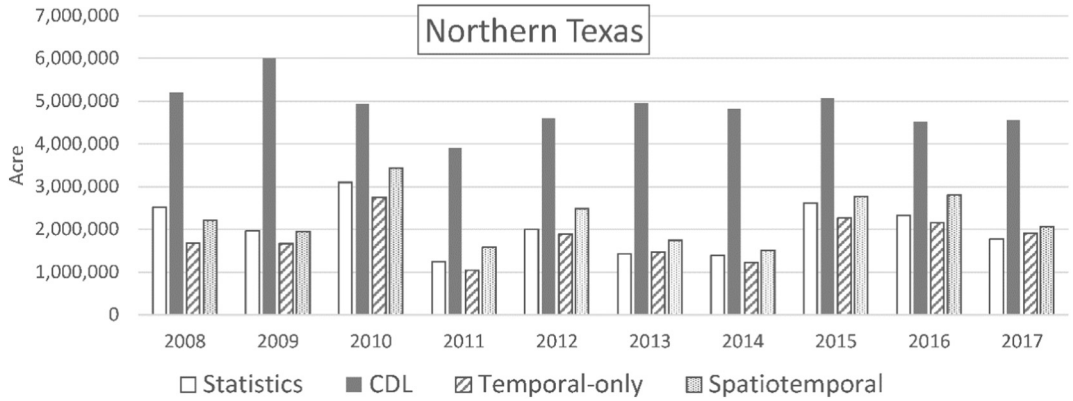

Fig. 9. Winter wheat acreage of Northern Texas from the USDA statistics compared with mapped acreage by the CDL, the temporal-only model, and the spatiotemporal model, in 2008–2017.

In summary, by conducting visual inspection, pixelwise accuracy assessment, and comparison with regional statistics, we see that the temporal-only model successfully produced winter wheat maps using only county-level statistics as training data. In Kansas where the winter wheat extent of the CDL is relatively accurate, the resultant maps show great agreement with the CDL. In Northern Texas where the CDL considerably differs from the NASS statistics, the use of the deep learning framework provides the possibility of generating winter wheat maps with decent consistency with the statistics. The spatiotemporal model resulted in reasonable patterns and regional estimates, but pixels in the map were subject to positional errors up to the size of the receptive field in the current settings.

4. Discussion

4.1. Interpretation of model behavior

In our mapping framework, the input of the model is 250-m NDVI image series directly from the MOD13Q1 product, and the output is county-level statistics. There is an extremely large difference in dimensionality between the input and the output, which results in a large parameter space in the model, a possible ill-posed problem, and a challenge to end-to-end learning. By designing a specialized deep architecture consisting of multiple fully-convolutional layers, constraints on feature representation are placed to reduce the negative effect of model complexity on generalization and overcome the problem of high dimensionality (Zhang et al., 2018). The deep model in this study automatically extracts features and builds a pathway to connect each of the *width*-by-*height*-by-*date* data cubes directly to a scaler value in the statistics. It is critical to examine the model details and investigate if features are learnt reasonably along the pathway.

One way to understand how a multi-layer convolutional model works is to inspect the activation patterns on each convolutional layer by different inputs. The temporal-only model is used as the example to demonstrate the internal model behavior because i) the model resulted in good pixelwise accuracy, and ii) visualization of only temporal convolution operations is relatively clear and straightforward without being affected by high-dimensional data structures and graphs. We visualized the activation of input NDVI series on various layers using deconvolution and guided back-propagation (Zeiler and Fergus, 2014), and the visualization of three pixels in 2017 is presented as an example (Fig. 10). Two of the NDVI profiles are from winter wheat pixels, and the third one is canola (Column "input series", Fig. 10). The land use types are from the CDL in 2017, which were correctly classified by the temporal-only model as winter wheat or non-winter-wheat. The series of the two winter wheat pixels are different, and their average (the dashed line in the third plot of input series) is very close to the canola series (solid line). Such intra-class heterogeneity and inter-class similarity are often the sources of mis-classification, so we chose the three pixels as an example to investigate how the deep model distinguishes varying temporal patterns.

By using the approach of guided back-propagation, the activation on each channel can be attributed to individual time steps in the input series to highlight which part of the series activates the channel. In the columns of convolutional layers in Fig. 10, the activations are shown by the brightness of the time steps. A bright dot means the channel is strongly activated by the time step, and a dark one means weak activation. In the first convolutional layer, all the three series activate 4 out of the 8 channels, and all the time steps contribute almost equally to the strength of the activation. This is expected because earlier layers in a deep model are used to capture local and basic patterns that might be present throughout the whole input. The activation pattern on the second convolutional layer is similar, except that the contributions by different dates start to vary and the three series start to show disparate activations (like Channel #6). In the third and the fourth convolutional layers, each channel responds to more complex patterns. Some channels are only activated by one series but not others, for example, Channels #3, #13, and #15 in the fourth layer. By recognizing complex patterns, the model starts to have the ability to distinguish the three series, but the activations by the same type are not necessarily similar. The last convolutional layer summarizes patterns learnt in the previous layer using a single channel. The increasing slope in the first series and the two peaks and one slope in the second series strongly activate the channel, and then the two series are identified as winter wheat. The activation by the third series is much weaker, and thus the canola pixel is not classified as winter wheat. Although the dual-peak profile of the canola pixel is very close to the average of the two winter wheat series, the activation pattern in some channels of later convolutional layers is completely different. By inspecting the activations on later layers, we see that similar series may follow distinct paths of channel activation through the model, which offers the ability to fully utilize complex patterns to capture seasonality information in the NDVI series. The patterns learnt by the deep architecture provide sufficient degree of model variance to account for the complexity in winter wheat growth, and all the patterns are completely trained by only county-level statistics.

When using the deep model for mapping, it is essential to understand whether the trained network recognizes different winter wheat patterns across years, or just relies on constant temporal patterns across years and captures inter-annual variability by coincidence under changing conditions like traditional manual approaches. We explored the model capability of recognizing varying patterns using the Kansas results by the temporal-only model, for which reliable CDL data are available for assessment and the pixelwise accuracy is high. The temporal NDVI profiles of classified winter wheat pixels could be quite different across years. For example, Fig. 11 presents the median time

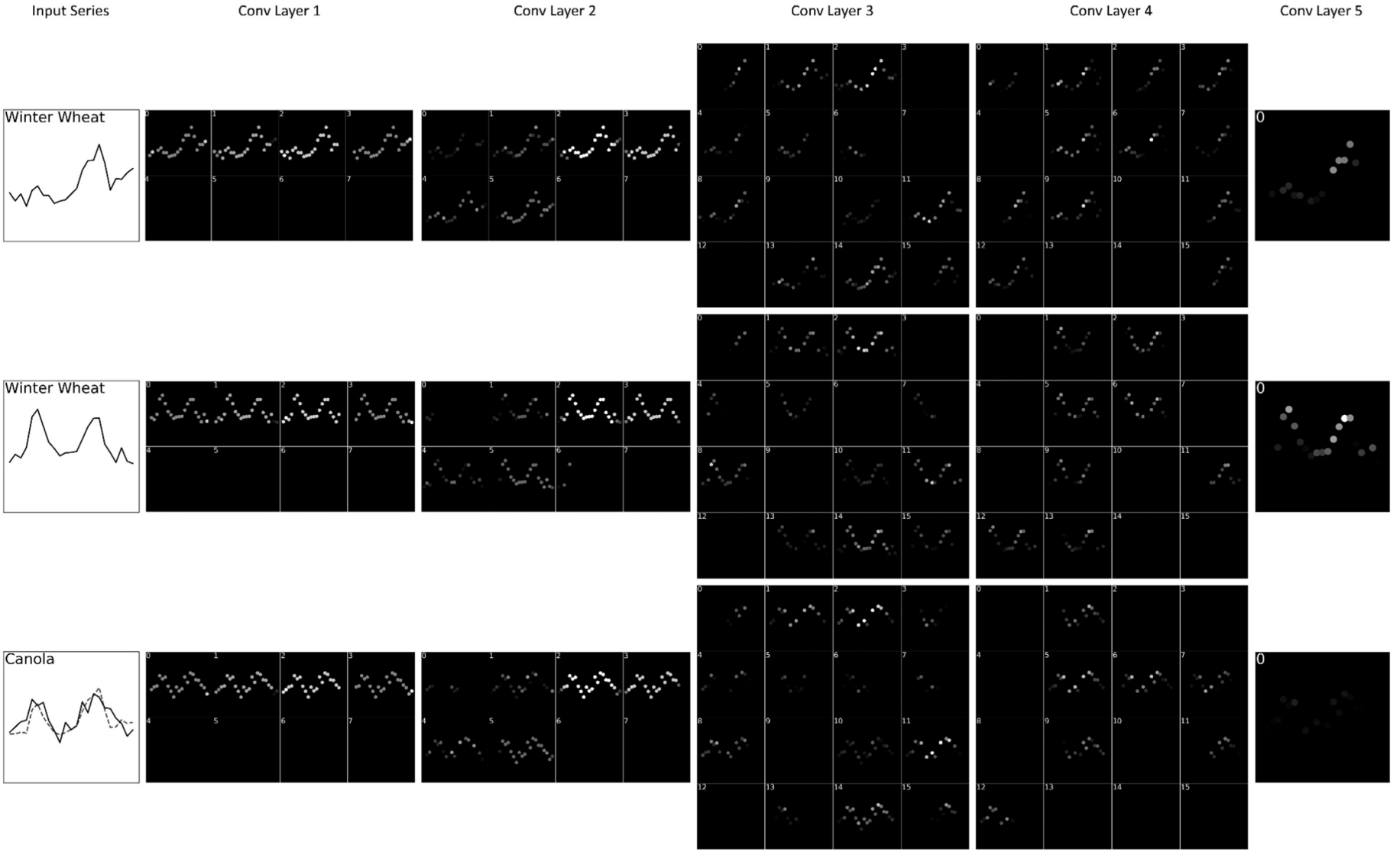

Fig. 10. Three NDVI series in solid lines and their activation patterns on all the convolutional layers of the temporal-only model. The first two NDVI profiles are from winter wheat pixels, and the third one is canola but is very similar to the average of the first two which is plotted in dashed line. The x-axis (date axis) is consistent with Fig. 11 and Fig. 12. The five convolutional layers can be found in the architecture of the temporal-only model (Fig. 2), with 8, 8, 16, 16, and 1 channel(s), respectively. Channel numbers starting from zero are labeled on the upper left corner of each channel. The activation by individual time steps on the five layers was calculated by deconvolution and guided back-propagation. NDVI values at different time steps are shown by dots, and the brightness of dots represent the strength of the activation by the current time step. Classifier adaptation for inter-annual variability.

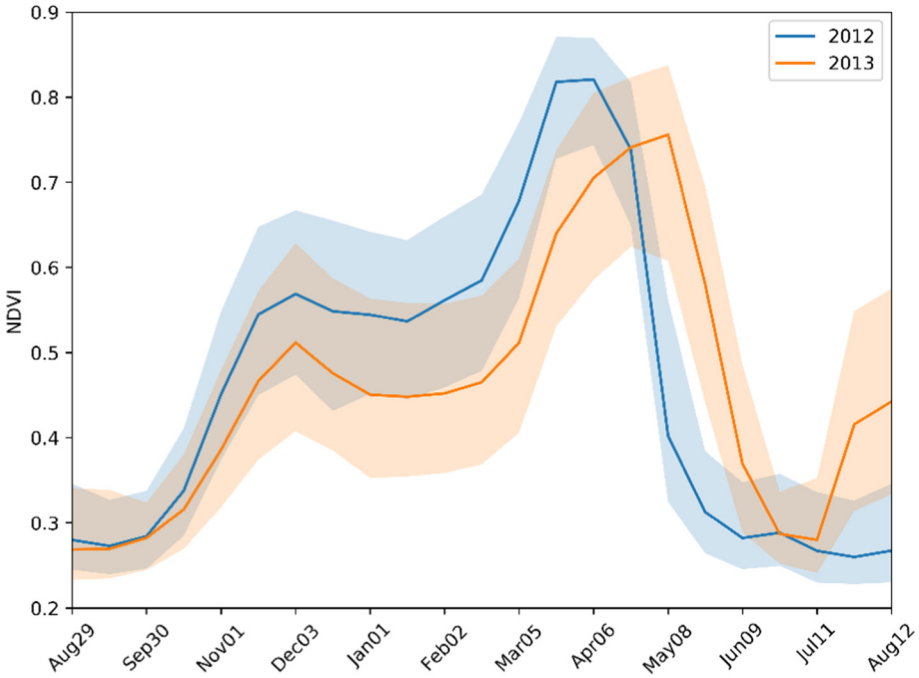

Fig. 11. Time series plots of winter wheat pixels classified by the temporal-only model in Kansas, 2012 and 2013. The winter wheat progress in 2012 was earlier than average while 2013 was later. Solid lines represent the median NDVI values of winter wheat pixels. Shaded areas show the ranges between the 80% and the 20% percentiles. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

series (blue and orange lines) and the 20%–80% ranges (blue and orange shaded areas) of winter wheat pixels mapped for years 2012 and 2013, respectively. According to the USDA weekly crop progress reports (WWW5, n.d.), in Kansas the year 2012 has the earliest winter wheat progress, and 2013 is one of the latest among recent years. The difference in the "headed" stage between the two years is as large as about one month, and the NDVI ranges have small overlaps. The dates of the NDVI peaks in Fig. 11 are consistent with the reports, showing that the distinct temporal patterns of the two years were effectively learnt by the statistics-based classifier. Similarly, Fig. 12 demonstrates the inter-annual variability resulted by crop conditions. NDVI in 2010 is much higher than 2006 during the peak season and lower in winter. According to the USDA weekly reports, the percentage of Kansas winter wheat that was in "good" or "excellent" conditions was over 55% at the harvest time in 2010, but only about 20% in 2006. Compared to the traditional approach of manually building and selecting features or rules to account for possible changes under various circumstances, end-to-end learning with deep models is a more efficient way to generalize the variability in phenology, crop conditions, and agricultural practices in the study area across years.

4.2. Applicability of the statistics-based mapping approach

The main objective of our study is to show the possibility of generating maps solely from regional statistics using deep neural networks. The method would be useful when ground samples are unavailable, for example, to fill the historical gaps of the CDL or to create maps for countries using published statistical data. While ground references are

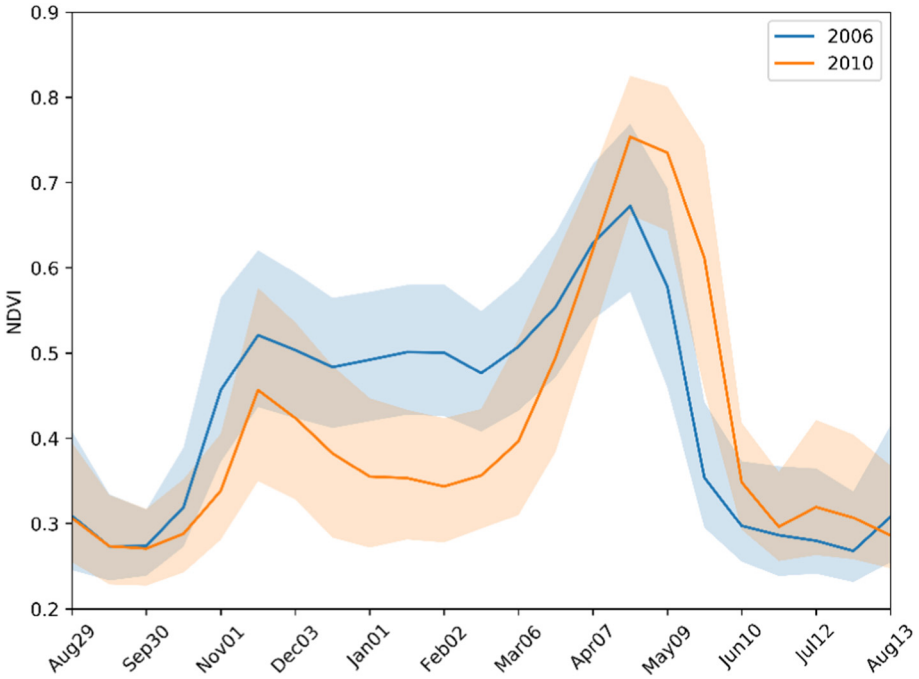

Fig. 12. Time series plots of winter wheat pixels classified by the temporal-only model in Kansas, 2006 and 2010. The crop condition around the harvest time in 2010 was much better than 2006. Solid lines represent the median NDVI values of winter wheat pixels. Shaded areas show the ranges between the 80% and the 20% percentiles.

Table 1
Evaluation metrics of winter wheat maps in Kansas.

	Wall-to-wall assessment								High-confidence-pixel-only assessment							
	Temporal-only model				Spatiotemporal model				Temporal-only model				Spatiotemporal model			
	Overall accuracy	Producer's accuracy	User's accuracy	F1 score	Overall accuracy	Producer's accuracy	User's accuracy	F1 score	Overall accuracy	Producer's accuracy	User's accuracy	F1 score	Overall accuracy	Producer's accuracy	User's accuracy	F1 score
2006	92.2%	65.6%	81.9%	0.729	87.8%	61.0%	61.9%	0.614								
2007	91.7%	61.6%	81.9%	0.703	86.3%	55.7%	57.2%	0.564								
2008	90.9%	54.2%	84.9%	0.661	87.0%	57.0%	61.1%	0.590	99.4%	92.1%	98.2%	0.950	99.3%	92.6%	95.7%	0.941
2009	91.3%	61.0%	79.8%	0.691	87.1%	58.8%	60.0%	0.594	99.6%	94.9%	98.5%	0.966	99.5%	94.4%	97.6%	0.960
2010	93.2%	63.9%	86.6%	0.735	88.6%	59.8%	62.0%	0.609	99.9%	96.0%	95.6%	0.958	99.8%	97.0%	88.3%	0.924
2011	92.2%	59.1%	84.7%	0.696	87.4%	54.2%	59.5%	0.567	99.6%	94.5%	97.6%	0.960	99.3%	92.1%	91.8%	0.920
2012	91.6%	65.4%	80.2%	0.720	86.0%	58.8%	57.5%	0.581	99.3%	87.7%	93.3%	0.904	98.5%	86.3%	77.4%	0.816
2013	90.8%	56.1%	83.4%	0.671	84.9%	50.5%	55.0%	0.527	99.6%	92.0%	96.9%	0.944	99.2%	90.2%	90.4%	0.903
2014	91.1%	54.0%	80.5%	0.646	87.4%	56.7%	58.6%	0.576	99.7%	92.4%	79.7%	0.856	99.7%	88.2%	78.9%	0.833
2015	90.9%	58.7%	79.9%	0.677	86.7%	55.8%	59.5%	0.576	99.5%	94.1%	96.7%	0.954	99.4%	92.9%	95.2%	0.941
2016	93.0%	67.6%	82.6%	0.743	87.4%	58.0%	58.0%	0.580	99.8%	97.3%	97.0%	0.972	99.4%	95.4%	89.4%	0.923
2017	93.9%	70.0%	80.6%	0.749	88.5%	59.3%	55.3%	0.572	99.8%	96.7%	94.3%	0.955	99.4%	95.9%	79.7%	0.870
Average	91.9%	61.4%	82.3%	0.702	87.1%	57.1%	58.8%	0.579	99.6%	93.8%	94.8%	0.942	99.3%	92.5%	88.5%	0.903

Table 2
Evaluation metrics of winter wheat maps in Northern Texas.

	Wall-to-wall assessment								High-confidence-pixel-only assessment							
	Temporal-only model				Spatiotemporal model				Temporal-only model				Spatiotemporal model			
	Overall accuracy	Producer's accuracy	User's accuracy	F1 score	Overall accuracy	Producer's accuracy	User's accuracy	F1 score	Overall accuracy	Producer's accuracy	User's accuracy	F1 score	Overall accuracy	Producer's accuracy	User's accuracy	F1 score
2008	91.7%	29.1%	88.3%	0.437	91.6%	35.8%	75.6%	0.486	98.2%	68.9%	99.8%	0.815	97.8%	61.9%	99.6%	0.763
2009	89.8%	23.4%	88.3%	0.370	89.7%	27.6%	75.8%	0.404	95.5%	54.8%	99.9%	0.708	95.4%	54.3%	99.8%	0.703
2010	93.4%	46.9%	83.7%	0.601	92.6%	53.8%	70.2%	0.609	99.0%	85.5%	99.1%	0.918	98.7%	81.6%	98.6%	0.893
2011	92.7%	20.1%	80.0%	0.322	92.6%	28.1%	66.4%	0.395	99.3%	43.7%	91.5%	0.592	99.3%	47.5%	88.1%	0.617
2012	91.4%	26.4%	70.3%	0.384	91.3%	35.9%	62.7%	0.457	98.2%	33.1%	95.9%	0.492	98.6%	50.4%	91.6%	0.651
2013	90.7%	22.0%	83.4%	0.348	90.4%	25.9%	70.8%	0.379	98.9%	48.5%	97.4%	0.648	98.8%	43.0%	95.9%	0.594
2014	90.7%	17.4%	82.9%	0.287	90.6%	22.8%	68.7%	0.342	98.7%	41.5%	98.9%	0.585	98.5%	39.4%	93.2%	0.554
2015	91.9%	35.2%	77.7%	0.484	91.4%	40.9%	67.0%	0.508	96.5%	68.2%	99.6%	0.810	96.1%	64.4%	99.6%	0.782
2016	93.1%	39.4%	80.9%	0.530	92.3%	45.0%	67.1%	0.539	99.0%	73.5%	96.4%	0.834	98.8%	67.5%	96.3%	0.794
2017	91.8%	28.3%	72.6%	0.407	91.5%	31.4%	64.5%	0.423	98.7%	49.8%	93.2%	0.649	98.7%	51.8%	95.0%	0.671
Average	91.7%	28.8%	80.8%	0.417	91.4%	34.7%	68.9%	0.454	98.2%	56.7%	97.2%	0.705	98.1%	56.2%	95.8%	0.702

limited by the difficulty and the cost of data collection, statistics published by many institutes have a long record length and broad spatial coverage. Although there are far fewer details in the county statistics than in the ground samples, in the experiment the maps generated from county statistics are consistent with the crop coverage developed by pixel-level supervised classification and many ground samples as long as the statistics agree well with the pixel-based classification map. When the statistics disagree with the existing crop maps as in the case of Northern Texas, it is often hard to locate and understand the source of discrepancy because the two datasets cannot be directly compared at a detailed spatial level. The training samples of the pixel-based classification maps may not represent the overall distribution of the whole region of statistics. The proposed method can produce maps consistent with the statistics without being affected by the disparate distribution of ground samples or existing classification maps, which provides a unique opportunity to inspect the differences between existing statistics and classification maps using the statistics-derived map as a bridge.

As an initial experiment based on MODIS imagery, the resultant maps are subject to the limitation of 250 m resolution and other factors, as indicated by existing MODIS-based crop classification efforts (Doraiswamy et al., 2007; Wardlow and Egbert, 2008; Shao et al., 2010). To simplify the model architecture and reduce the computational cost, only the time series of one index (NDVI) was used, which limited the multi-spectral information for crop classification. The statistics-based approach is applicable only when the input time series and bands can also identify the crop type of interest with pixel-level reference data.

The use of the global average pooling layer in the proposed model focuses on the estimation accuracy at the county level but not for each individual pixel. By training the model with many combinations of counties and years, the model has to successfully predict the crop cover at the pixel level to always agree with the statistics for all combinations rather than just manufacturing coincidences at the county level. Furthermore, we visualize the activations on different layers of the deep model to show how the proposed architecture captures seasonality in a hierarchical manner. The use of sigmoid activation between 0 and 1 on the last convolutional layer and the global average pooling layer ensures that the values in the resultant maps are physically reasonable. The mapping and inspection results suggest that the model trained with statistics really learns how to recognize the time series of winter wheat.

4.3. A tunable and extensible deep learning framework for mapping

The deep learning based approach proposed in this study creates a new type of mapping tasks that directly utilize statistics in supervised classification. The framework is highly flexible, which can be extended to incorporate more types of statistical data in crop mapping and other remote sensing applications. To improve crop mapping, commodity production and trade data from agricultural commissioners of individual counties can be linked to the fully convolutional network. Data related to agricultural activities such as water use (like delivery to irrigation districts) and pesticide use (an example is California Pesticide Information Portal, WWW6, n.d.) can formulate additional constraints on the network. Furthermore, the framework is potentially useful in a variety of mapping tasks for a combined use of regional statistics and pixelwise data, as long as the variable of interest is relevant to remote sensing observations to some extent. For example, the study that disaggregates population from administrative districts to pixels (Stevens et al., 2015) may benefit from the automated deep learning based solution.

When applying the deep network framework, regional statistics might be used as the sole source of training set, just like the present study, or as supplementary data to enforce additional constraints on the network. In the latter case, statistical data are combined with conventional pixelwise training samples to fully utilize all available data sources for model tuning. The framework is flexible to have multiple sets of regional and pixel-level information simultaneously contributing to the loss function of the neural network model, which accounts for various factors in a uniform way (Jia et al., 2018). Conventional classification efforts are often limited by the availability of ground samples, as ground data collection is expensive, time-consuming, or impossible retrospectively for historical periods. Meanwhile, statistical records are a rich source of accumulated domain knowledge, but it is a pity that statistics have rarely been employed effectively in classification. The use of statistics in the deep learning framework provides an opportunity to reduce the dependency on ground reference samples and extend mapping studies to a larger area, a longer period, or a broader scope of applications.

5. Conclusion

This study used per-county statistics in 2001–2017 to train deep models to produce winter wheat maps at the 250 m resolution of MODIS imagery. The training of the deep models did not require pixel-level reference data. The end-to-end framework of the fully-convolutional neural networks built a data pathway from NDVI image series to winter wheat coverage at the county level, and winter wheat maps were obtained from a middle layer of the network. The proposed mapping approach is highly automated and efficient because i) the approach does not rely on pixel-level references from ground data collection, and ii) the approach does not require manual feature engineering and selection but incorporates all inputs into a deep model to recognize complex and changing patterns. The model architecture that employed patterns in the temporal dimension successfully identified the seasonal dynamics of winter wheat. As the approach is able to produce maps using only statistical data as the training reference, the resultant maps possess high consistency with the statistics compared to classification maps trained by pixel-wise reference data. While this approach is valuable for areas or periods without pixel-wise ground samples, the resultant maps can also be employed to inspect the possible discrepancy between existing classification maps and regional statistics.

Land use mapping using statistics as reference data is a new type of remote sensing classification task. With the development of deep learning technology, it is possible to employ a highly-specialized deep network architecture to implement constraints from statistical data to regulate patterns learnt by the model and complete the task. This deep learning based classification framework is very extensible thanks to the flexibility of network architectures. The trained model can be tuned with more statistical data in the future or from other areas to improve generalization, reduce the need of ground sample collection, and alleviate the shortage of reference data in many tasks like historical mapping. Disparate data sources including pixelwise references and statistics of various human activities can be incorporated into a uniform framework to fully utilize all existing data. The new classification task and the deep learning based solution are worth further exploration.

Acknowledgement

We thank the anonymous reviewers for their careful reading of our manuscript and their insightful comments.

References

Allen, R.G., Tasumi, M., Trezza, R., 2007. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) — model. Journal of Irrigation and Drainage Engineering-Asce 133, 380–394.

Audebert, N., Le Saux, B., Lefèvre, S., 2018. Beyond RGB: very high resolution urban remote sensing with multimodal deep networks. ISPRS J. Photogramm. Remote Sens. 140, 20–32.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., et al., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: part 1: principles of the algorithm. Remote Sens. Environ. 110, 275–286.

Bastiaanssen, W.G.M., Menenti, M., Feddes, R.A., Holtslag, A.A.M., 1998. A remote sensing surface energy balance algorithm for land (SEBAL) — 1. Formulation. J. Hydrol.

213, 198–212.
Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, cropland data layer program. Geocarto International 26, 341–358.
Chen, X., Xiang, S., Liu, C.-, Pan, C.-, 2014b. Vehicle detection in satellite images by hybrid deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. 11, 1797–1801.
Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y., 2014a. Deep learning-based classification of hyperspectral data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 2094–2107.
Doraiswamy, P.C., Stern, A.J., & Akhmedov, B. (2007). Crop classification in the U.S. Corn Belt using MODIS imagery, 809–812.
Fisher, J.B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., et al., 2017. The future of evapotranspiration: global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. Water Resour. Res. 53, 2618–2626.
Guidici, D., Clark, M.L., 2017. One-dimensional convolutional neural network land-cover classification of multi-seasonal hyperspectral imagery in the San Francisco Bay Area, California. Remote Sens. 9, 629.
Harou, J.J., Pulido-Velazquez, M., Rosenberg, D.E., Medellín-Azuara, J., Lund, J.R., Howitt, R.E., 2009. Hydro-economic models: concepts, design, applications, and future prospects. J. Hydrol. 375, 627–643.
Howard, D.M., Wylie, B.K., 2014. Annual crop type classification of the US Great Plains for 2000 to 2011. Photogramm. Eng. Remote. Sens. 80, 537–549.
Hu, F., Xia, G., Hu, J., Zhang, L., 2015b. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sens. 7, 14680–14707.
Hu, W., Huang, Y., Wei, L., Zhang, F., Li, H., 2015a. Deep convolutional neural networks for hyperspectral image classification. Journal of Sensors 2015.
Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv preprint arXiv:1502.03167.
Irwin, E.G., Geoghegan, J., 2001. Theory, data, methods: developing spatially explicit economic models of land use change. Agric. Ecosyst. Environ. 85, 7–24.
Jakubauskas, M.E., Legates, D.R., Kastens, J.H., 2001. Harmonic analysis of time-series AVHRR NDVI data. Photogramm. Eng. Remote. Sens. 67, 461–470.
Jia, X., Karpatne, A., Willard, J., Steinbach, M., Read, J., Hanson, P.C., et al., 2018. Physics Guided Recurrent Neural Networks for Modeling Dynamical Systems: Application to Monitoring Water Temperature and Quality in Lakes. arXiv preprint arXiv:1810.02880.
Kampffmeyer, M., Salberg, A., Jenssen, R., 2016. Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks. pp. 1–9.
Kandasamy, S., Baret, F., Verger, A., Neveux, P., Weiss, M., 2013. A comparison of methods for smoothing and gap filling time series of remote sensing observations—application to MODIS LAI products. Biogeosciences 10, 4055–4071.
Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.
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, 778–782.
Li, W., Fu, H., Yu, L., Gong, P., Feng, D., Li, C., et al., 2016. Stacked Autoencoder-based deep learning for remote-sensing image classification: a case study of African land-cover mapping. Int. J. Remote Sens. 37, 5632–5646.
Li, W., Fu, H., Yu, L., Cracknell, A., 2017a. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. Remote Sens. 9.
Li, Y., Zhang, H., Shen, Q., 2017b. Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. Remote Sens. 9, 67.
Lyapustin, A., Wang, Y., Xiong, X., Meister, G., Platnick, S., Levy, R., et al., 2014. Scientific impact of MODIS C5 calibration degradation and C6 improvements. Atmospheric Measurement Techniques 7, 4353–4365.
Lyu, H., Lu, H., Mou, L., 2016. Learning a transferable change rule from a recurrent neural network for land cover change detection. Remote Sens. 8, 506.
Lyu, H., Lu, H., Mou, L., Li, W., Wright, J., Li, X., et al., 2018. Long-term annual mapping of four cities on different continents by applying a deep information learning method to landsat data. Remote Sens. 10, 471.
Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. High-resolution aerial image labeling with convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 55, 7092–7103.
Marcos, D., Volpi, M., Kellenberger, B., Tuia, D., 2018. Land cover mapping at very high resolution with rotation equivariant CNNs: towards small yet accurate models. ISPRS J. Photogramm. Remote Sens. 145, 96–107.
Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U., 2018. Classification with an edge: improving semantic image segmentation with boundary detection. ISPRS J. Photogramm. Remote Sens. 135, 158–172.
Marsden, C., Nouvellon, Y., Laclau, J., Corbeels, M., McMurtrie, R.E., Stape, J.L., et al., 2013. Modifying the G'DAY process-based model to simulate the spatial variability of Eucalyptus plantation growth on deep tropical soils. For. Ecol. Manag. 301, 112–128.
Massey, R., Sankey, T.T., Congalton, R.G., Yadav, K., Thenkabail, P.S., Ozdogan, M., et al., 2017. MODIS phenology-derived, multi-year distribution of conterminous US crop types. Remote Sens. Environ. 198, 490–503.
Monfreda, C., Ramankutty, N., Foley, J.A., 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Glob. Biogeochem. Cycles 22, GB1022.
Mou, L., Zhu, X.X., 2018. RiFCN: Recurrent Network in Fully Convolutional Network for Semantic Segmentation of High Resolution Remote Sensing Images. arXiv preprint arXiv:1805.02091.
Mou, L., Bruzzone, L., Zhu, X.X., 2018a. Learning Spectral-Spatial-Temporal Features Via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery. arXiv preprint arXiv:1803.02642.
Mou, L., Ghamisi, P., Zhu, X.X., 2018b. Unsupervised spectral-spatial feature learning via deep residual conv-deconv network for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 56, 391–406.
Penatti, O.A.B., Nogueira, K., Dos Santos, J.A., 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2015-October, pp. 44–51.
Phalke, A.R., Özdoğan, M., 2018. Large area cropland extent mapping with Landsat data and a generalized classifier. Remote Sens. Environ. 219, 180–195.
Ramankutty, N., Evan, A.T., Monfreda, C., Foley, J.A., 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. Glob. Biogeochem. Cycles 22, GB1003.
Rußwurm, M., Körner, M., 2017. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.
Rußwurm, M., Körner, M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. ISPRS Int. J. Geo Inf. 7, 129.
Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2014. Near real-time prediction of U.S. corn yields based on time-series MODIS data. Remote Sens. Environ. 147, 219–231.
Sayer, A., Hsu, N., Bettenhausen, C., Jeong, M., Meister, G., 2015. Effect of MODIS Terra radiometric calibration improvements on Collection 6 Deep Blue aerosol products: Validation and Terra/Aqua consistency. Journal of Geophysical Research: Atmospheres 120, 12,157–12,174.
Searchinger, T., Heimlich, R., Houghton, R.A., Dong, F., Elobeid, A., Fabiosa, J., et al., 2008. Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change. Science (New York, N.Y.) 319, 1238–1240.
Shao, Y., Lunetta, R.S., Ediriwickrema, J., Liames, J., 2010. Mapping cropland and major crop types across the Great Lakes Basin using MODIS-NDVI data. Photogramm. Eng. Remote. Sens. 76, 73–84.
Sherrah, J., 2016. Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery. arXiv preprint arXiv:1606.02585.
Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-scale Image Recognition. arXiv preprint arXiv:1409.1556.
Song, X., Potapov, P.V., Krylov, A., King, L., Di Bella, C.M., Hudson, A., et al., 2017. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. Remote Sens. Environ. 190, 383–395.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15, 1929–1958.
Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. PLoS One 10, e0107042.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al., 2015. Going deeper with convolutions. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 1–9.
Tang, Q., Peterson, S., Cuenca, R.H., Hagimoto, Y., Lettenmaier, D.P., 2009. Satellite-based near-real-time estimation of irrigated crop water consumption. J. Geophys. Res.-Atmos. 114, D05114.
Thenkabail, P.S., Wu, Z., 2012. An automated cropland classification algorithm (ACCA) for Tajikistan by combining Landsat, MODIS, and secondary data. Remote Sens. 4.
Volpi, M., Tuia, D., 2017. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 55, 881–893.
Wan, X., Zhao, C., Wang, Y., Liu, W., 2017. Stacked sparse autoencoder in hyperspectral data classification using spectral-spatial, higher order statistics and multifractal spectrum features. Infrared Physics and Technology 86, 77–89.
Wang, D., Morton, D., Masek, J., Wu, A., Nagol, J., Xiong, X., et al., 2012. Impact of sensor degradation on the MODIS NDVI time series. Remote Sens. Environ. 119, 55–61.
Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the US Central Great Plains. Remote Sens. Environ. 112, 1096–1116.
WWW1 United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) Quick Stats. https://www.nass.usda.gov/Quick_Stats/, Accessed date: 26 November 2018.
WWW2 Brazilian Institute of Geography and Statistics Data Download. https://downloads.ibge.gov.br/downloads_estatisticas.htm.
WWW3 Keras: The Python Deep Learning Library. https://keras.io/.
WWW4 Tensorflow: An Open Source Software Library for High Performance Numerical Computation. https://www.tensorflow.org.
WWW5 USDA NASS Crop Progress. http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1048, Accessed date: 26 November 2018.
WWW6 California Pesticide Information Portal. https://calpip.cdpr.ca.gov/main.cfm.
Xin, Q., Gong, P., Yu, C., Yu, L., Broich, M., Suyker, A.E., et al., 2013. A production efficiency model-based method for satellite estimates of corn and soybean yields in the Midwestern US. Remote Sens. 5, 5926–5943.
Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnelt, J., Congalton, R.G., et al., 2017. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. ISPRS J. Photogramm. Remote Sens. 126, 225–244.
Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: European Conference on Computer Vision, pp. 818–833.
Zhang, J., Liu, T., Tao, D., 2018. An Information-Theoretic View for Deep Learning. arXiv preprint arXiv:1804.09060.

Zhong, L., Gong, P., Biging, G.S., 2012. Phenology-based crop classification algorithm and its implications on agricultural water use assessments in California's Central Valley. Photogramm. Eng. Remote. Sens. 78, 799–813.
Zhong, L., Gong, P., Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: a multi-year experiment using Landsat imagery. Remote Sens. Environ. 140, 1–13.
Zhong, L., Hu, L., Yu, L., Gong, P., Biging, G.S., 2016a. Automated mapping of soybean and corn using phenology. ISPRS J. Photogramm. Remote Sens. 119, 151–164.
Zhong, L., Yu, L., Li, X., Hu, L., Gong, P., 2016b. Rapid corn and soybean mapping in US Corn Belt and neighboring areas. Sci. Rep. 6, 36240.
Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430–443.
Zhu, X.X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., et al., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine 5, 8–36.