Environmental Research Letters

LETTER

DeepCropNet: a deep spatial-temporal learning framework for county-level corn yield estimation

Tao Lin1, Renhai Zhong2,3, Yudi Wang4, Jinfan Xu1, Hao Jiang1, Jialu Xu1, Yibin Ying4,5, K C Ting6,7,8 and Haifeng Li1,2

1 College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, Zhejiang, 310058, People’s Republic of China
2 International Campus, Zhejiang University, Haining, Zhejiang, 314400, People’s Republic of China
3 China Academy of Electronic and Information Technology, Beijing, 100041, People’s Republic of China
4 Faculty of Agricultural and Food Science, Zhejiang A&F University, Hangzhou, Zhejiang, 311300, People’s Republic of China
5 Department of Agricultural and Biological Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States of America
6 School of Geosciences and Info-Physics, Central South University, Changsha, 410083, People’s Republic of China
7 Henan Laboratory of Spatial Information Application on Ecological Environment Protection, Zhengzhou, 450000, People’s Republic of China
8 Author to whom any correspondence should be addressed.
E-mail: lihaifeng@csu.edu.cn

Keywords: yield estimation, corn, LSTM, attention mechanism, multi-task learning, deep learning

Supplementary material for this article is available online

Abstract
Large-scale crop yield estimation is critical for understanding the dynamics of global food security. Understanding and quantifying the temporal cumulative effect of crop growth and spatial variances across different regions remains challenging for large-scale crop yield estimation. In this study, a deep spatial-temporal learning framework, named DeepCropNet (DCN), has been developed to hierarchically capture the features for county-level corn yield estimation. The temporal features are learned by an attention-based long short-term memory network and the spatial features are learned by the multi-task learning (MTL) output layers. The DCN model has been applied to quantify the relationship between meteorological factors and the county-level corn yield in the US Corn Belt from 1981 to 2016. Three meteorological factors, including growing degree days, killing degree days, and precipitation, are used as time-series inputs. The results show that DCN provides an improved estimation accuracy (RMSE = 0.82 Mg ha\(^{-1}\)) as compared to that of conventional methods such as LASSO (RMSE = 1.14 Mg ha\(^{-1}\)) and Random Forest (RMSE = 1.05 Mg ha\(^{-1}\)). Temporally, the attention values computed from the temporal learning module indicate that DCN captures the temporal cumulative effect and this temporal pattern is consistent across all states. Spatially, the spatial learning module improves the estimation accuracy based on the regional specific features captured by the MTL mechanism. The study highlights that the DCN model provides a promising spatial-temporal learning framework for corn yield estimation under changing meteorological conditions across large spatial regions.

1. Introduction

Providing sufficient food supply becomes a critical global issue under the increased population growth. Food supply, however, has become more vulnerable under the global warming phenomenon that causes the increased occurrence of extreme meteorological events (FAO 2018). Understanding the relationship between meteorological factors and crop growth is essential to quantify the crop yield estimation, which would provide decision support for national food production policy and local farming management (Liu et al 2016, Jin et al 2017, Jones et al 2017).

Process-based biophysical modeling (McCown et al 1996) and statistical modeling (Lobell and Burke 2010) are two mainstream approaches to quantify crop yield
based on meteorological factors. Biophysical modeling is able to represent the key processes of crop growth and yield formation based on mechanistic equations developed through theoretical derivation and experimental data. However, the validation of mechanistic equations is challenging, especially when scientific experiments and data are limited. Statistical modeling uses observed data of meteorology and crop yield to quantify their associations by statistical regression, which is convenient for large-scale applications. Statistical modeling has challenges in dealing with the complex interactions among meteorological variables and the nonlinear relationships among the dependent and independent variables. Biophysical modeling is more suitable for site-specific yield analysis, whereas statistical modeling is often adopted in large-scale spatial analysis. Some researchers look into the integration of process-based and statistical modeling in an effort to achieve higher accuracy (Michael et al. 2017). Under the rapid development of computing capabilities in recent years, artificial intelligence approaches, such as Random Forest (RF) (Saeed et al. 2017), artificial neural network (Alvarez 2009), Bayesian network (Gandhi et al. 2016), semiparametric neural network (Crane-Droesch 2018), and convolutional neural network (You et al. 2017, Yang et al. 2019), have gradually been applied to extract patterns for agricultural yield estimation. These studies, however, often simplify the temporal cumulative effect of crop growth and the spatial heterogeneity problem.

Crop growth is featured by its temporal cumulative growth and spatial explicit pattern. Temporally, crops grow and accumulate biomass over time. How to express the physiological process while building a machine learning model is not well understood. Spatially, meteorological conditions and their influence on crop yield vary across different regions (Butler and Huybers 2013, Zipper et al. 2016). Individual local models in each region ignore the inherent correlations among different regions. Most global models consider all regions as a whole and do not deal with the spatial differences of the meteorological conditions and their relationships with crop yield (Van Wart et al. 2013, Van Ittersum et al. 2013). How to develop a data-driven deep learning model to capture the spatial and temporal features for crop yield estimation is critical for an improved understanding of the relationship between crop yield and meteorological factors.

Long short-term memory (LSTM) model is a special type of recursive neural network, which can facilitate time-series analysis and handle complexity and nonlinearity by its unique structure (Hochreiter and Schmidhuber 1997). Previous studies have demonstrated that LSTM performs well in dealing with long sequential data in natural language modeling (Sundermeyer et al. 2012) and human trajectory prediction (Alahi et al. 2016). In agriculture, researchers have also begun to study the feasibility of LSTM, such as water table depth prediction (Zhang et al. 2018), cropland classification (Rußwurm and Körner 2017), and yield prediction (Jiang et al. 2019, Wang et al. 2018).

Attention (AT) mechanism has been inspired by human visual attention, and it can show how much attention the model is paid to the input sequence (Chaudhari et al. 2019). Attention mechanism has shown its state-of-the-art performance in concentration on different parts of a sequence data both in natural language processing (Luong et al. 2015, Wang et al. 2016) and in image processing (Ba et al. 2014, Mnih et al. 2014).

Multi-task learning (MTL) is an approach to build a model for multiple tasks to improve generalization. MTL can learn tasks in parallel while using a shared representation and what is learned for each task can help other tasks be learned better (Caruana 1997). MTL is able to study the commonalities and differences between different tasks and has shown good performance in rainfall prediction (Qiu et al. 2017), neuroimaging studies (Ma et al. 2018) and railway track inspection (Gibert et al. 2017).

Inspired by the successful applications of the deep learning technologies mentioned above, a deep spatial-temporal learning framework named Deep-CropNet (DCN) is developed in this study to understand the interactions between crop yield and meteorological factors through the integration of the temporal and spatial learning modules. For the temporal pattern, a temporal learning module that incorporates the attention mechanism into the LSTM model is applied to capture and visualize the temporal cumulative effect. To address the spatial heterogeneity problem, a spatial learning module is developed to facilitate MTL to take yield estimation in different meteorological regions as different tasks to learn both the common features and region-specific features. The US Corn Belt is selected as the region for a case study and the tasks are defined based on the spatial distributions of the meteorological conditions during the corn growth period. This study aims to address the three research questions as follows:

1. How does DCN perform for corn yield estimation compared with LASSO and RF?
2. How does DCN learn the temporal patterns between time-series meteorological information and corn yields?
3. How does DCN learn the regional specific features of corn yield estimation?

2. Model

2.1. DeepCropNet

The DCN model proposed in this study is a deep learning framework that uses multiple layers to hierarchically learn the spatial and temporal features from the meteorological and yield data. The DCN
consists of two parts: temporal learning module and spatial learning module (Figure 1). The input of DCN is a time-series of meteorological factors, including county-level weekly growing degree days (GDD), killing degree days (KDD) and precipitation (PRCP) from corn planted (week 1) to mature (week 20). The output is the corn yield anomaly at the county-level. In this study, an AT-LSTM network is trained with parameters shared for all regions to learn and visualize the general temporal pattern captured from the time-series of weekly meteorological factors. To learn the spatial features in each region and output the corn yield anomaly, region-specific output layers are developed based on MTL as the spatial learning module.

2.2. Temporal learning module

The general temporal features are learned by an AT-LSTM network, which consists of an LSTM Net and an Attention Net (Figure 1). The LSTM Net has five layers: input layer (one layer), hidden layers (three layers), and output layer (one layer). The input is a time sequence \( \{x_1, x_2, \ldots, x_t\} \) and \( x_t \) is a vector which includes meteorological factors [GDD\(_t\), KDD\(_t\), PRCP\(_t\)] in the period \( t \) during the corn growing season. The hidden layers are three LSTM layers composed of LSTM cells to transport and store information selectively. The dimension of features in each LSTM layer is set as 32 after selection range from 16 to 256. The multiple LSTM layers enable the LSTM Net to generate hierarchical features from layer by layer. Temporally, the feature vector and the cell state vector in each LSTM layer are transferred from the current period \( t \) to the next period \( t + 1 \). After processing the whole-time sequence data, the LSTM Net calculates the output, \( h_t \), the hidden feature vector of the LSTM in period \( t \). In this study, we define the vector of hidden features, \( h_t \), as the reference of the effect of meteorological factors on corn yield. The calculation process...
of $h_t$ is as follows:

$$g^f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$  \hspace{1cm} (1)
$$g^i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$  \hspace{1cm} (2)
$$S_t = g^f_t \ast S_{t-1} + g^i_t \ast \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$  \hspace{1cm} (3)
$$g^o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$  \hspace{1cm} (4)
$$h_t = g^o_t \ast \tanh(S_t),$$  \hspace{1cm} (5)

where $t$ represents the period $t$ during the corn growth season. $W_f, W_i, W_o, b_f, b_i, b_c$ are weight matrices and bias vector parameters of forget gate, input gate, output gate and cell state update, respectively. $S_t$ is the updated cell state vector at period $t$. $g^f_t, g^i_t$, and $g^o_t$ are the vectors generated by these gates. The three vectors consist of numbers range from zero to one, for the operation of filtering information. Zero represents totally removing, and one represents totally reserving. $S_t$ is the LSTM cell state vector to store information across the whole-time sequence. $h_t$ is the hidden feature vector of the LSTM. $\sigma$ and $\tanh$ are activation functions.

The Attention Net can generate different attention value $\alpha_t$ based on the $h_t$ to reflect the importance of hidden feature vector to corn yield in the corresponding period. The structure of the Attention Net is designed as one single-layer full-connected neural network in this study. The final feature vector ($H$) captured by the AT-LSTM is the combination of attention values of Attention Net and hidden feature vectors of LSTM Net calculated as follows:

$$\alpha_t = \text{softmax}(W_{AT} \cdot h_t + b_{AT})$$  \hspace{1cm} (6)
$$H = \alpha_1 h_1 + \alpha_2 h_2 + \ldots + \alpha_t h_t,$$  \hspace{1cm} (7)

where $\alpha_t$ is the attention value, $\text{softmax}$ is the activation function to limit the sum of attention value to one, $W_{AT}$ is the weight matrix of the Attention Net, and $b_{AT}$ is the bias vector.

As the sum of attention values during the corn growth period equals to one, the distribution of attention values can reflect the importance of hidden features at each period. The higher attention value represents that the hidden features in the corresponding period have a higher impact on the estimated yield. The attention distribution and its association with the corn growth process can provide insights to help us understand how the temporal module work for estimating yield.

2.3. Spatial learning module

MTL is applied to learn region-specific features in the relationships between corn yield and meteorological factors. Considering the meteorological conditions (extreme heat in particular) vary across latitudes (figure S1 is available online at stacks.iop.org/ERL/15/034016/mmedia), the nine states are divided into three regions and the yield estimation in each region is defined as an individual task to adjust the regional meteorological conditions. In the three regions, the LSTM layers and the Attention Net layer of the temporal learning module are shared while the final output layers of the spatial learning module are task-specific (or region-specific). The shared layers extract the common features across different regions, and the task-specific output layers generate weights based on local characteristics for final yield estimation in each region. The yield anomaly ($y$) is calculated as equation (8).

$$y = W_r \ast H + b_r,$$  \hspace{1cm} (8)

where $r$ is the index of the region, $W_r$ and $b_r$ are the weight matrix and bias vector of the responding regional specific output layer for region $r$. $H$ are the attention weighted features captured by the AT-LSTM network.
3. Data

3.1. Study area
This study focuses on county-level rainfed corn yield in the nine states of the US Corn Belt region from 1981 to 2016 (figure 2). These 743 counties accounted for 61% of the total corn production of the US in 2016 (USDA-NASS 2017). During the 36 years between 1981 and 2016, the annual county-level average yield in this area is 7.9 Mg ha$^{-1}$. The high-yielding regions are concentrated in southern Minnesota, Iowa, and northern Illinois. All counties in this study have an increasing trend of corn yield, with an average trend of 0.12 Mg ha$^{-1}$ yr$^{-1}$.

3.2. Data and preprocessing
The county-level rainfed corn yield data is obtained from USDA’s National Agriculture Statistics Service from 1981 to 2016 (USDA-NASS 2017). To capture the impact of meteorological factors on corn yield, the yield anomaly (yield after detrending) is calculated by linear regression (Lobell et al. 2011). To ensure the reliability of the data in our analysis, we design three criteria for data cleaning and processing:

(1) The county should have more than 12 year yield statistics, which is one-third of the whole research period.

(2) The county should have an average corn harvest area higher than 476 ha, which is 10 percentile value of the county-level average corn harvest area in the whole sample set.

(3) The county should have a yield trend that passes the t-test, with a two-side p-value less than 0.05.

There are 848 counties with 28,374 samples in the raw yield data. After the data cleaning and processing, 743 counties (87.6%) with 26,015 samples (91.7%) remain in our analysis. The state-level corn phenology data is based on USDA-NASS (2017). The growing season is defined as the average historical average length, 20 weeks, from the specific planting week recorded in each state to keep the periods consistent.

The daily county-level meteorological data is from the Applied Climate Information System Web Service (ACIS) (RCC-ACIS 2017). The meteorological data used in this study include maximum daily temperature, minimum daily temperature, and daily precipitation. Based on these meteorological data, cumulative meteorological factors are calculated including cumulative PRCP as an index to represent the water supply, GDD as an index to represent effective heat accumulation, and KDD as an index to represent extreme heat accumulation. The calculation of GDD and KDD is consistent with a previous study (Butler and Huybers 2015). During the corn growth period, the weekly meteorological factors are calculated and fed into the model as sequential vector inputs. To improve the model convergence speed, the input is standardized into standard normal distribution.

3.3. Baselines and performance evaluation
In this study, two approaches are chosen as baselines: LASSO regression and RF. For the LASSO regression model, the λ value is set to 0.0005 by the 10-fold cross-validations based on the training set. For RF, the tree number is set to 2000 and 20 of 60 input variables are randomly taken as candidates at each split. The maximum number of nodes for each tree is limited to 200. The gini net and randomForest R package are used to train LASSO and RF, respectively. The DCN is implemented on the Python platform using Pytorch library and performed on a Linux workstation (Ubuntu 14.04 LTS) with 128 GB of RAM, and an Nvidia GeForce GTX1080Ti graphics card with 11 GB of RAM.

To evaluate the models’ performance in yield estimation, all models are trained using the dataset during the period from 1981 to 2014 and tested using the 2015 and 2016 data. The estimated yield anomaly is added with the yield trend to get the final estimated yield. The RMSE and $R^2$ between actual yield and estimated yield in the test set are used as indicators to evaluate the estimation accuracy of these models. To analyze the benefit from the spatial learning module, the temporal learning module is applied individually to estimate yield as a comparison with DCN.

To evaluate the model performance in predicting the interannual variability of corn yield, 36 years are randomly partitioned into six subsets (table S1) to conduct 6-fold cross-validation. The DCN model is trained by the five subsets and tested in the remaining subset. The RMSE and $R^2$ of each subset are used to evaluate the model stability.

4. Results

4.1. Estimation accuracy of DCN at county-level
The DCN model shows its estimation performance at the county-level, with RMSE equals 0.83 Mg ha$^{-1}$ in 2015 and 0.81 Mg ha$^{-1}$ in 2016 (figure 3). The error maps indicate that DCN performs well in most counties, with an average absolute error lower than 0.8 Mg ha$^{-1}$ (figure 3). While in Missouri, DCN estimates less accurately, with the average absolute error equals 0.88 Mg ha$^{-1}$. These results can be explained by the high data missing rate (35%) of this state in the test set. A similar pattern that the number of county samples affects model’s performance has been reported in previous studies (Li et al. 2019). Although there exist few extreme estimation errors, the overall performance of DCN is satisfying which indicates that DCN is able to well capture the yield variability at the county-level.
4.2. Performance comparison among DCN, LASSO, and RF

DCN provides the best estimation results with RMSE equals 0.82 Mg ha\(^{-1}\) as compared with LASSO and RF (figure 4). LASSO provides the lowest estimation accuracy with RMSE equals 1.14 Mg ha\(^{-1}\), agreeing with the discovery in previous study that linear methods are limited in simulating the relationship between crop yield and meteorological factors relative to nonlinear approaches (Cai et al. 2018). RF provides a better result than LASSO with RMSE equals 1.05 Mg ha\(^{-1}\). Compared with LASSO and RF, DCN reduces the estimation RMSE by 0.23–0.32 Mg ha\(^{-1}\), which demonstrates the effectiveness of the spatial and temporal learning process by neural network based structure.

The results show a spatial pattern that all models perform better in northern states than in southern states (figure 4). The estimation RMSEs are correlated with the variances of yield anomaly at state-level, with Pearson’s correlation coefficient of 0.79 (figure S2). Similar results have been reported in a previous study (Li et al. 2019). The higher variabilities of the yield anomaly in southern states indicate that corn yield is more affected by meteorological stress in the south. For example, the historical meteorological data shows that the average cumulative KDD during the corn growth period in southern states (165 degree day) is 2.2 times higher than that in northern states (76 degree day) in past three decades. Therefore, it appears that the spatial distribution of meteorological stress is one of the driving factors for this spatial pattern.

The DCN model shows its higher estimation accuracy in most states compared with LASSO and RF, particularly in the southern states. In the states in which corn yield is more stable, such as Michigan and Iowa, different models perform similarly and DCN may not be the best approach. While in the southern states in which the corn yield has more variations due to the higher heat stress, DCN performs the best and

![Figure 3. The scatter plots and error maps between estimated and actual yield by DCN in 2015 (a), (c) and 2016 (b), (d). Positive values represent that the model yields an overestimated result.](image)
shows large improvements relative to LASSO and RF. DCN lowers the RMSE in these states by 13.1%–47.1% compared with LASSO and RF. The stable high performance of DCN indicates that it is a practical approach for yield estimation in the regions with more complex and stressful corn growth conditions.

### 4.3. Temporal pattern learning module by AT-LSTM

The AT-LSTM captures a general temporal pattern across all states that DCN provides higher attention weights on the features in reproductive stage than vegetative stage (figure 5). The attention values represent the relative importance of the cumulative information captured by the LSTM Net from the previous input data when estimating corn yields. The consistent pattern of attention values implies that the model has the ability to capture the general temporal information during corn growth across the geographical areas, which is essential for the model’s generalization in different regions. The weekly average attention value keeps to be 0.02 from week 1 to week 11 (vegetative stage), starts increasing at week 12 (silking stage), and achieves the highest value of 0.16 at week 20 (maturity stage). This result can be explained by the cumulative effect across time. From planted to mature, the effective information accumulates and drives the model to generate increased weights at a later stage. The temporal distribution of attention values shows that the DCN discovers the silking stage as the starting point for increased attention. Cumulative information becomes increasingly important to the yield estimation from silking stage. This indicates that DCN can recognize the key growth phases during the corn growing season through the temporal learning module. Previous studies have demonstrated that corn yield is highly correlated to the rainfall before silking and effective accumulated temperature after silking (Lu et al 2017). The variability of attention values starts increasing from week 15. The increasing variability during the later stage of corn growth possibly indicates that accumulated meteorological information have an increased impact on corn yield but may vary among different counties in different years even in the same state. The results illustrate that temporal learning module provides a promising approach to capture the

![Figure 4](image-url)
temporal pattern that relates to crop physiological process.

The temporal learning module (AT-LSTM) has already achieved a better estimation accuracy than LASSO and RF, with RMSE equals 0.89 Mg ha\(^{-1}\) and \(R^2\) equals 0.72 (figure S3). This demonstrates the function of the temporal feature learning in crop yield estimation. The results show relatively high estimation errors in Illinois, Indiana, and Ohio (figure 6). These overestimated results may result from the spatial variances in these states that are not captured by the temporal learning module.

4.4. Spatial pattern learning module by MTL

The spatial learning module improves the estimation performance relative to the temporal learning module,
where overall RMSE in the test set is reduced by 0.07 Mg ha$^{-1}$ (7.9%). All regions in 2015 and 2016 reach a better estimation accuracy with the improvements range from 1.7% to 11.2%, as the regional RMSEs are reduced by 0.02 to 0.10 Mg ha$^{-1}$ (figure 7). The most significant improvement is in the central region, with the RMSE reduced by 9.1% in 2015 and 11.2% in 2016. These improvements demonstrate that incorporating the spatial learning module can improve the model performance in regions with various meteorological conditions.

The spatial learning module brings improvements by reducing the relatively high estimation errors produced by the temporal learning module. The absolute estimation errors can be reduced by up to 1.15 Mg ha$^{-1}$ with a mean equaling 0.05 Mg ha$^{-1}$ at the county-level in 2015 and 2016 (figure 7). The largest improvements concentrate in counties of northern Indiana and northwest Ohio in 2015, with absolute estimation errors reduced by up to 1.15 Mg ha$^{-1}$ (figure 7). The individual temporal learning module overestimates the corn yield in these counties with estimation errors range from 0.75 to 2.92 Mg ha$^{-1}$. These overestimations can be explained by a localized pattern that these counties suffer excessive rainfall during the corn growing season in 2015, with their precipitation (669 mm) is 1.5 times the historical average (460 mm). These localized improvements indicate that the spatial learning module could possibly capture the local specific pattern (e.g. excessive rainfall; drought; extreme high temperature), to improve yield estimation.

4.5. Model robustness performance for interannual yield variability

The DCN model outperforms RF and LASSO at estimating interannual variability in corn yields, with the lowest RMSE and highest $R^2$ in every subset of the 6-fold cross-validation (table 1). Among these subsets, the RMSE of DCN ranges from 0.84 to 1.18 Mg ha$^{-1}$ and the $R^2$ ranges from 0.71 to 0.86. The DCN model shows the highest performance stability, with the lowest standard deviation of RMSE equals 0.13 Mg ha$^{-1}$, compared with RF (0.23 Mg ha$^{-1}$) and LASSO (0.16 Mg ha$^{-1}$). The DCN model shows the largest

![Figure 7. Comparison of estimation accuracy among DCN, temporal learning module, RF and LASSO. (a), (b) show RMSE barplots among four approaches and (c), (d) show changes of the absolute estimation errors between DCN and temporal learning module in 2015 and 2016. Negative values represent reduced estimation errors by the spatial learning module.](image-url)
performance improvement compared with RF and LASSO in subset 5, which includes the year 2012. The US Corn Belt suffered a heavy drought in 2012 with an average yield at 6.8 Mg ha$^{-1}$. These results indicate that DCN produces robust corn yield estimation across a wide range of environmental conditions.

### 5. Discussion

The DCN model provides improved yield estimation compared to the existing approaches using both meteorological data and remote sensing data (Kim and Lee 2016, Morell et al 2016, Li et al 2019). Li (2019) constructed a statistical model using climate and remote sensing data to predict corn yield in the Midwest US and achieved the best median RMSE of 0.90 Mg ha$^{-1}$ from 2003 to 2016. Morell (2016) developed a crop simulation model to estimate the county-level corn yield in the US Corn Belt with $R^2$ of 0.59. Kim (2016) proposed a machine learning approach using satellite images and climate data to estimate corn yield in Iowa with RMSE equals 0.76 Mg ha$^{-1}$. It is noted that different models have varied analysis scopes in both spatial and temporal scales and the dimensions of inputs. The DCN model in this study achieves satisfactory yield estimation accuracy with overall RMSE equals 0.82 Mg ha$^{-1}$ and state-level RMSE ranges from 0.61 to 1.08 Mg ha$^{-1}$ (figure 4). It is noteworthy that only three meteorological factors are used as input of the DCN model in this study. Additional data, such as vapor pressure deficit and vegetation indices derived from the remote sensing reflectance, can be added into the model to potentially improve its performance.

The DCN model provides interpretable information, which can provide some new insights in agricultural production through the combination of big data and deep learning. Low interpretability makes deep learning model a ‘black box’ and how to improving it is a research focus (Montavon et al 2018). In this study, the attention mechanism is applied to visualize the temporal patterns between time-series meteorological information and corn yields. The temporal patterns illustrate that DCN is able to recognize key growth phases, such as silking and kernel filling stages, which stages are essential to the corn yield as reported in related studies (Edreira et al 2011, Sánchez et al 2014). This information allows us to understand how deep learning based crop yield model work to improve its application.

Many possibilities remain to further refine the spatial and temporal learning framework, such as adjusting the temporal and spatial resolution, spatial clustering and model fusion as follows:

1. The temporal and spatial resolutions of DCN are scalable. The temporal resolution can be defined from daily to monthly, or by the crop growth phase. The DCN can also be applied at different spatial scales to estimate crop yield. The association between the performance of DCN and its spatial and temporal scales remains to be further studied.

2. An improved spatial clustering approach can be developed compared to the regional identification based on the state-level heat distribution in this study (figure S1). For example, unsupervised clustering approaches can be utilized to achieve a better clustering of spatial regions with more input dimensions such as soil and solar radiation.

3. Combining DCN with statistical and process-based crop model can possibly improve its performance. The knowledge obtained from the statistical analysis and crop model definition allows for the refining of the calculation process of DCN, such as improving the parameterization.

### 6. Conclusion

A deep spatial-temporal learning framework named DCN has been developed for corn yield estimation. The results highlight the potential of DCN in crop yield estimation because of the ability to capture the temporal general pattern and spatial specific features. The temporal learning module shows a consistent pattern in all states that DCN provides more attention to the reproductive stage rather than vegetative stage due to the cumulative effect. This result indicates that DCN can recognize the key growth phases during the corn growing season through the temporal learning module. The comparison between the temporal learning only and DCN proves that blending spatial learning module into DCN improves corn yield estimation in various regions. The result implies that it is advantageous to consider spatial variances while estimating meteorological impacts on corn yield. The association between the performance of DCN and its spatial and temporal scales remains to be further studied.

The study demonstrates that DCN is a promising approach for crop yield estimation and identifies potential for improvement. Although the model proposed here is for rainfed corn in the US Corn Belt, it can possibly be applied to other crops and in other regions.

---

**Table 1.** Comparison of estimation accuracy among different models by 6-fold cross-validation.

|               | RMSE (Mg ha$^{-1}$) | $R^2$ |
|---------------|---------------------|------|
|               | DCN | RF | LASSO | DCN | RF | LASSO |
| Subset 1      | 0.88 | 1.01 | 1.05 | 0.71 | 0.63 | 0.59 |
| Subset 2      | 1.18 | 1.23 | 1.33 | 0.74 | 0.71 | 0.67 |
| Subset 3      | 1.07 | 1.27 | 1.13 | 0.86 | 0.80 | 0.84 |
| Subset 4      | 0.84 | 0.87 | 0.94 | 0.77 | 0.76 | 0.72 |
| Subset 5      | 1.07 | 1.46 | 1.33 | 0.71 | 0.47 | 0.56 |
| Subset 6      | 0.90 | 0.97 | 1.05 | 0.76 | 0.73 | 0.68 |
| Overall       | 1.00 | 1.16 | 1.15 | 0.78 | 0.70 | 0.71 |
Acknowledgments

We thank anonymous reviewers for helpful comments and Yue Zhu for her assistance on model analysis. This work was supported by the National Natural Science Foundation of China (Grant Numbers 31701316, 41571397, 41871564) and China National Key Research and Development Plan (Grant Number 2017YFD0700605).

Data availability statements

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID iDs

Haifeng Li https://orcid.org/0000-0003-1173-6593

References

Alahi A et al 2016 Social LSTM: human trajectory prediction in crowded spaces 2016 IEEE Conf. Comput. Vis. Pattern Recognt pp 961–71
Alvarez R 2009 Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach Eur. J. Agron. 30 70–7
Ba J, Mnih V and Kavukcuoglu K 2014 Multiple object recognition with visual attention 1–10 arXiv:1411.7755
Butler E E and Huybers P 2013 Adaptation of US maize to temperature variations Nat. Clim. Change 3 68–72
Butler E E and Huybers P 2013 Variations in the sensitivity of US maize yield to extreme temperatures by region and growth phase Environ. Res. Lett. 10 014009
Cai Y et al 2018 Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches Agric. For. Meteorol. 274 144–59
Caruana R 1997 Multitask learning Mach. Learn. 28 41–75
Chaudhari S, Polatkan G, Ramanath R and Mithal V 2019 An attentive survey of attention models Mach. Learn. arXiv:1904.02874
Crane-Droesch A 2018 Machine learning methods for crop yield prediction and climate change impact assessment in agriculture Environ. Res. Lett. 13 114003
Edreza JR et al 2011 Heat stress effects around flowering on kernel set of temperate and tropical maize hybrids Field Crops Res. 123 62–73
FAO, IFAD, UNICEF, WFP, WHO 2018 The State of Food Security and Nutrition in the world 2018: Building Climate Resilience for Food Security and Nutrition (Rome: FAO)
Gilbert X, Patel V M and Chellappa R 2017 Deep multitask learning for railway track inspection IEEE Trans. Intell. Transp. Syst. 18 153–64
Gandhi N, Armstrong L I and Petkar O 2016 Predicting rice crop yield using Bayesian network 2016 Int. Conf. Adv. Comput. Commun. Informatics pp 795–9
Hohebeter S and Schmidhuber J 1997 Long short-term memory Neural Comput. 9 1735–80
Jiang H, Hu H, Zhong R, Xu J, Xu J, Huang J, Wang S, Ying Y and Lin T 2019 A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level Glob. Change Biol. accepted (https://doi.org/10.1111/gcb.14885)
Jin Z, Azzari G and Lobell D B 2017 Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches Agric. For. Meteorol. 247 207–20
Jones J W et al 2017 Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science Agric. Syst. 155 269–88
Kim N and Lee Y W 2016 Machine learning approaches to corn yield estimation using satellite images and climate data: A case of Iowa State J. Korean Soc. Surv. Geol. Photogramm. Cartogr. 34 383–90
Li Y et al 2019 Toward building a transparent statistical model for improving crop yield prediction: modeling rainfed corn in the US Field Crops Res. 234 35–65
Liu B et al 2016 Similar estimates of temperature impacts on global wheat yield by three independent methods Nat. Clim. Change 6 1130–6
Lobell D B and Burke M B 2010 On the use of statistical models to predict crop yield responses to climate change Agric. For. Meteorol. 150 1443–52
Lobell D B, Schlenker W and Costa-Roberts J 2011 Climate trends and global crop production since 1980 Science 333 616–20
Lu H, Xue J and Guo D 2017 Efficacy of planting date adjustment as a cultivation strategy to cope with drought stress and increase rainfed maize yield and water-use efficiency Agric. Water Manage. 179 237–35
Luong M T, Pham H and Manning C D 2015 Effective approaches to attention-based neural machine translation Proc. 2015 Conf. Empirical Methods Nat. Lang. Proc. (Lisbon: Association for Computational Linguistics) pp 1412–21
Ma Q et al 2018 Classification of multi-site MR images in the presence of heterogeneity using multi-task learning Neuroimage Clin. 19 476–88
McCrown R L, Hammer G L, Hargreaves J N G, Holzworth D P and Freebairn D M 1996 APSIM: a novel software system for model development, model testing and simulation in agricultural systems research Agric. Syst. 50 255–71
Michael J R, Noah O B, Thomas R S, David B L and Wolfram S 2017 Comparing and combining process-based crop models and statistical models with some implications for climate change Environ. Res. Lett. 12 195010
Mnih V, Heess N, Graves A and Kavukcuoglu K 2014 Recurrent models of visual attention Adv. Neural Inf. Process. Syst. 27 1–9 (www.papers.nips.cc/paper/5342-recurrent-models-of-visual-attention.pdf)
Montavon G, Samek W and Müller K R 2018 Methods for interpreting and understanding deep neural networks Digit. Signal Process. 73 1–15
Morell F J et al 2016 Can crop simulation models be used to predict local to regional maize yields and total production in the US Corn Belt Field Crops Res. 192 1–12
Qiu M et al 2017 A short-term rainfall prediction model using multi-task convolutional neural networks Proc.—IEEE Int. Conf. Data Mining, ICDM 2017 pp 395–404
RCC-ACIS 2017 RCC-ACIS web service (www.rcc-acis.org)
Rußwurm M and Körner M 2017 Multi-temporal land cover classiﬁcation with long short-term memory neural networks Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.—ISPRS Arch. 42 551–8 (www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXII-1-W1/551/2017/isprs-archives-XXII-1-W1-551-2017.pdf)
Saeedi U et al 2017 Forecasting wheat yield from weather data and MODIS NDVI using Random Forests for Punjab province, Pakistan Int. J. Remote Sens. 38 4831–54
Sánchez B, Rasmussen A and Porter J R 2014 Temperatures and the growth and development of maize and rice: A review Glob. Change Biol. 20 408–17
Sundermeyer M, Schl R and Ney H 2012 LSTM neural networks for language modeling Proc. Interspeech pp 194–7 (www.isca-speech.org/archive/interspeech_2012/i12_0194.html)
USDA-NASS 2017 Quick Stats 2.0. (Washington, DC: USDA-NASS) (www.nass.usda.gov/Quick_Stats)
Van Ittersum M K et al 2013 Yield gap analysis with local to global relevance—a review Field Crops Res. 143 4–17
Van Wart J et al 2013 Estimating crop yield potential at regional to national scales Field Crops Res. 143 34–43
Wang A X, Tran C, Desai N, Lobell D and Ermon S 2018 Deep transfer learning for crop yield prediction with remote sensing data Proc. 1st ACM SIGCAS Conf. Comput. Sustain. Soc.—COMPASS ‘18 pp 1–5
Wang Y, Huang M, Zhao L and Zhu X 2016 Attention-based LSTM for aspect-level sentiment classification Proc. 2016 Conf. Empir. Methods Nat. Lang. Process. pp 606–15
Yang Q, Shi L, Han J, Zha Y and Zhu P 2019 Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images Field Crops Res. 235 142–53
You J, Li X, Low M, Lobell D and Ermon S 2017 Deep Gaussian process for crop yield prediction based on remote sensing data 2017 Assoc. Adv. Articial Intell. (www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14435/14067)
Zhang J, Zhu Y, Zhang X, Ye M and Yang J 2018 Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas J. Hydrol. 561 918–29
Zipper S C, Qiu J and Kucharik C J 2016 Drought effects on US maize and soybean production: Spatiotemporal patterns and historical changes Environ. Res. Lett. 11 094021