

# Machine Learning Crop Yield Models Based on Meteorological Features and Comparison with a Process-Based Model

QINQING LIU,<sup>a</sup> MEIJIAN YANG,<sup>b,c</sup> KOUSHAN MOHAMMADI,<sup>b,c</sup> DONGJIN SONG,<sup>a</sup> JINBO BI,<sup>a</sup> AND GUILING WANG<sup>b,c</sup>

<sup>a</sup> *Department of Computer Science and Engineering, University of Connecticut, Storrs, Connecticut*

<sup>b</sup> *Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut*

<sup>c</sup> *Institute of the Environment, University of Connecticut, Storrs, Connecticut*

(Manuscript received 8 January 2022, in final form 12 July 2022)

**ABSTRACT:** A major challenge for food security worldwide is the large interannual variability of crop yield, and climate change is expected to further exacerbate this volatility. Accurate prediction of the crop response to climate variability and change is critical for short-term management and long-term planning in multiple sectors. In this study, using maize in the U.S. Corn Belt as an example, we train and validate multiple machine learning (ML) models predicting crop yield based on meteorological variables and soil properties using the leaving-one-year-out approach, and compare their performance with that of a widely used process-based crop model (PBM). Our proposed long short-term memory model with attention (LSTM<sub>att</sub>) outperforms other ML models (including other variations of LSTM developed in this study) and explains 73% of the spatiotemporal variance of the observed maize yield, in contrast to 16% explained by the regionally calibrated PBM; the magnitude of yield prediction errors in LSTM<sub>att</sub> is about one-third of that in the PBM. When applied to the extreme drought year 2012 that has no counterpart in the training data, the LSTM<sub>att</sub> performance drops but still shows advantage over the PBM. Findings from this study suggest a great potential for out-of-sample application of the LSTM<sub>att</sub> model to predict crop yield under a changing climate.

**SIGNIFICANCE STATEMENT:** Changing climate is expected to exacerbate extreme weather events, thus affecting global food security. Accurate estimation and prediction of crop productivity under extremes are crucial for long-term agricultural decision-making and climate adaptation planning. Here we seek to improve crop yield prediction from meteorological features and soil properties using machine learning approaches. Our long short-term memory (LSTM) model with attention and shortcut connection explains 73% of the spatiotemporal variance of the observed maize yield in the U.S. Corn Belt and outperforms a widely used process-based crop model even in an extreme drought year when meteorological conditions are significantly different from the training data. Our findings suggest great potential for out-of-sample application of the LSTM model to predict crop yield under a changing climate.

**KEYWORDS:** Agriculture; Climate variability; Time series; Forecasting; Crop growth; Neural networks

## 1. Introduction

Feeding the rapidly increasing global population is a grand challenge of the twenty-first century. A major obstacle to global food security is the crop yield loss caused by extreme weather events such as drought, flood, and heatwaves, which is expected to be exacerbated by the changing climate as warming increases the frequency and severity of weather extremes. Accurate estimation and prediction of regional and global crop productivity and crop response to climate variability, change, and extremes are crucial for developing agricultural policies, setting international food aid priorities, forecasting and analyzing global trade trends, and identifying effective strategies for climate change adaptation. Remote sensing of the actual crop growth status offers a powerful yield predictor and is the basis of numerous crop yield models using machine learning or other statistical approaches (Battude et al. 2016; Fieuzal et al. 2017; Geipel et al. 2014; Khaki et al. 2021; Shahhosseini et al. 2019). However, remote sensing data are not available for long-lead seasonal prediction and future projections. To address this challenge, here we explore

different approaches to crop yield prediction based on growing season meteorological forcing and soil properties, focusing on machine learning and its comparison with process-based modeling.

Most past studies estimating crop yield projection based on growing season meteorological variables made use of classical statistical approaches such as multivariate regression (e.g., Lobell et al. 2020; Yang et al. 2020). While many variables that influence crop growth and productivity were considered, temperature and precipitation are the most commonly used as yield predictors (e.g., Schlenker and Roberts 2009). Typically, these models used meteorological statistics (e.g., mean) during the growing season (or its part), which cannot capture the impact of extreme weather conditions during some critical stages of crop development that may have a disproportionate impact on crop yield. Another challenge facing conventional statistical models is the uncertain applicability to historically unprecedented events, as it is not clear whether the observed predictors and relationships can be extrapolated.

Process-based crop models offer a preferred alternative for studying climate change impact on crop yield, especially in the context of future projections (e.g., Bassu et al. 2014; Franke et al. 2020; Qian et al. 2019; Rosenzweig et al. 2018, 2014).

*Corresponding author:* Guiling Wang, guiling.wang@uconn.edu

DOI: 10.1175/AIES-D-22-0002.1 e220002

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](#)).

Brought to you by UNIVERSITY OF CONNECTICUT | Unauthenticated | Downloaded 10/03/22 04:33 PM UTC

Process-based models (PBMs) take meteorological variables (e.g., air temperature and humidity, precipitation, and solar radiation) and soil properties as inputs, parameterize crop physiological and phenological processes, and entail a large number of cultivar-specific parameters that have to be calibrated based on field experiments. Multimodel intercomparison studies have documented a large degree of model uncertainties in simulating both present-day yield and sensitivity to climate changes (e.g., [Asseng et al. 2015](#); [Franke et al. 2020](#)). An individual model may perform better than others in one region and be outperformed elsewhere ([Franke et al. 2020](#)); multimodel ensemble mean tends to perform well in multiple regions ([Bassu et al. 2014](#)). It is not clear which model may perform well for a given region of interest and running multiple PBMs over a large spatial domain often requires major international coordination involving experts from multiple modeling groups.

In recent years, machine learning (ML) including deep learning has emerged as an alternative approach to accelerate crop yield modeling and prediction ([Van Klompenburg et al. 2020](#)). ML algorithms identify structures from complex (often nonlinear) data and generate accurate predictive models without a priori assumptions on the relationship between drivers and response. Convolutional neural network (CNN) and long short-term memory (LSTM) methods are the most frequently used ML crop yield models ([Van Klompenburg et al. 2020](#)). Some of the ML crop yield predictions combined in-season imageries from remote sensing with meteorological features and ground surveying data (e.g., [Filippi et al. 2019](#); [Jiang et al. 2020](#); [Zhang et al. 2020](#)), while others rely on soil and meteorological features and can therefore be applicable in future projections (e.g., [Crane-Droesch 2018](#)). For example, [Zhang et al. \(2020\)](#) integrated multiple sources of satellite data with climate variables, soil properties, and irrigation information to predict county-level maize yield in China using random forest (RF), gradient boosting (XGBoost), and LSTM, among others, and found that satellite remote sensing of vegetation contributed the most information to the prediction skills. [Jiang et al. \(2020\)](#) built multiple ML crop yield models including least absolute shrinkage and selection operator (LASSO), CNN, and LSTM for the U.S. Corn Belt using meteorological forcing and cumulative exposure metrics during phenological stages derived from remote sensing vegetation index, and found the latter being a more important contributor to the ML prediction skills than the former. [Sun et al. \(2019\)](#) introduced a CNN-LSTM framework where CNN is used to process spatial features and LSTM is used to integrate different time steps of each year. [Khaki et al. \(2020\)](#) developed a CNN-recurrent neural network (RNN) framework that applied different CNN layers to process both weather components and soil conditions and employed RNN layers to capture the time dependencies of crop yield over a number of years. [Crane-Droesch \(2018\)](#) trained a semiparametric neural network (SNN) by combining an ordinary least squares model based on growing degree-days with a nonparametric deep neural network that used daily meteorological features to predict maize yield in the U.S. Midwest, and achieved improved efficiency and accuracy over either approach when applied to years withheld from

the modeling training. It was also found that prior knowledge about important phenomena and functional forms (in terms of growing degree-days for example) significantly improves the SNN performance. However, based on the performance metrics presented ([Crane-Droesch 2018](#)), it was not clear how well the model captured the spatiotemporal variability of crop yield especially during extreme years.

In this study, we explore various ML algorithms for crop yield prediction based on soil properties and daily meteorological features as a data-driven approach without the benefit of expert knowledge, and we compare the performance of the resultant ML models with a PBM driven by relevant daily meteorological forcing and soil properties that influence crop yield through well-understood mechanisms. Our main contribution toward the ML-based crop yield prediction is the development of the LSTM framework with the attention mechanism that is able to account for both the dynamic and static features. Our approach was inspired by the work of [He et al. \(2016\)](#), which introduced the shortcut connection in ResNet to skip neural network layers in the forward step of an input, and the work of [Guo et al. \(2019\)](#), which employed the hidden states of LSTM to calculate the attention to integrate all hidden states. Specifically, we incorporate the static features to calculate the attention, which allows the shortcut connections to pass the information from the early-stage meteorological forcing directly to the final state by skipping the following LSTM cells and distinguishes the contribution of growing periods according to static features. We use maize data in the U.S. Corn Belt as an example to demonstrate the performance of the ML and PBM models; the same approach can be applicable to other crops and other regions. [Section 2](#) describes the data, models, and experimental design. Results are presented in [section 3](#), with conclusions and discussion in [section 4](#).

## 2. Data, models, and methods

Multiple ML algorithms and a PBM are used in this study to estimate crop yield in the U.S. Corn Belt based on soil properties (static) and daily meteorological forcing (dynamic). We train multiple ML algorithms including RF, XGBoost, and LSTM using county-level yearly crop yield data and county-averaged daily meteorological data. For the PBM, we choose the decision support system for agrotechnology transfer (DSSAT) model, as it was found to be the top performer among the widely used models for simulating historical maize yield in the United States ([Franke et al. 2020](#)). We conduct the DSSAT simulation at fine grid level driven by daily meteorological forcing and aggregate the model output over maize-harvesting areas within each county to derive the county-level yield. Note that while the ML models can take all daily meteorological variables for which data are available, DSSAT uses only a subset of these variables. For a sensitivity test, we also experiment on training the ML model using the same subset of meteorological variables as the input to DSSAT.

Our study domain covers the region of 35.5°–49.5°N and 80°–102°W and includes a total of 1413 counties after excluding some bordering counties. To ensure that data of the best

quality are used, our study focuses on the period of 2000–18. Some of the counties may not have the complete record of all 19 years; years with yield observation missing are not counted in the calculation of model performance metrics.

#### a. Data

Observational data for the county-level yield and harvest area are available from the USDA National Agricultural Statistics Service (NASS; <https://www.nass.usda.gov>). For the subcounty level harvest areas (needed for the area-weighted averaging of yield within each county), we use data from the latest Spatial Production Allocation Model (SPAM 2010, version 1.0) at 5 arc-min resolution (International Food Policy Research Institute 2019) (which is highly consistent with the USDA Cropland Data Layer in the United States; <https://nassgeodata.gmu.edu/CropScape/>) (Yu et al. 2020). Data on nitrogen fertilizer type, application rate and timing are available at  $5 \text{ km} \times 5 \text{ km}$  resolution during 1979–2015 from PANGAEA of the International Science Council World Data System (Cao et al. 2018). Temporally, we extrapolate the data linearly to 2018; spatially, the data over the Corn Belt is aggregated to a 5-arc-min model resolution for use in the PBM and to each county for use in the ML models. Planting time information is from USDA Agricultural Handbook (National Agricultural Statistics Service 2010). Data on soil features were available at 5-arc-min resolution from the Global High-Resolution Soil Profile Database for Crop Modeling Applications (version 2.5) (International Research Institute for Climate and Society et al. 2015), including, among others, soil texture, bulk density, organic carbon, soil porosity, and soil water content at six depths up to 200 cm.

Data for surface meteorological variables are taken from the gridMET dataset (Abatzoglou 2013), which covers the contiguous United States at a  $1/24^\circ$  ( $\sim 4 \text{ km}$ ) resolution during 1979–2018 and provides daily data for 16 variables, including the specific humidity, mean vapor pressure deficit, precipitation, minimum relative humidity, maximum relative humidity, downwelling solar radiation, minimum air temperature, maximum air temperature, wind speed, wind direction, reference grass evapotranspiration, reference alfalfa evapotranspiration, energy release component, burning index, FM100 (100-h dead fuel moisture), FM1000 (1000-h dead fuel moisture). Only a subset of these variables (including precipitation, temperatures, humidity, and solar radiation), at its native resolution, is used as the input forcing for the PBM. For ML models, all 16 variables are used, and we aggregate the grid-level feature values into county level by using the mean value of all grids within that county. Considering that the study domain spans a large range of latitudes that differ in growing season starting date and length, our ML models make use of the 210-day gridMET data of each year from 10 April to 6 November for all counties, but sensitivity experiments are designed to test the impact of using data from just the early part of the growing season.

#### b. DSSAT—A process-based crop model

As one of the most widely used and highly rated crop models, the DSSAT model has been proved reliable in simulating a variety of crops, capturing crop phenology and yield, and

reproducing processes during crop growth in the soil–plant–atmosphere continuum (Jones et al. 2003). First released in 1986, the DSSAT-CERES-Maize model incorporates knowledge in photosynthesis, carbon and nitrogen dynamics, evapotranspiration, soil water infiltration, etc. to approach the actual field level conditions through empirical and theoretical equations (Jones et al. 1986). It takes environmental and practical information including weather, soil, management, and cultivar data as input to simulate a crop of interest in a uniform area of land (Attia et al. 2021). The weather data used as model input include daily precipitation, maximum and minimum air temperature, solar radiation, and wind speed from the beginning to the end of the growing season; the growing season length is determined by the accumulated growing degree-days. Here, in order to capture the spatial environmental heterogeneity, we employ the DSSAT model for the Corn Belt region at 5 arc-minute resolution ( $\sim 9 \text{ km}$ ) with prescribed environmental and practical conditions within each pixel. Six crop genetic coefficients for the maize cultivar type need to be calibrated in the model, including P1, P2, P5, and PHINT that reflect the required degree-days for crop emergence, anthesis, silking, and maturity, respectively, and G2 and G5 that determine the potential kernel number and size in a single plant (Tovihoudji et al. 2019). DSSAT has been calibrated for the Corn Belt region in an earlier study (M. Yang and G. Wang 2022, unpublished manuscript), and the same calibrated model is used here.

#### c. Feature engineering for ML models

Our ML crop prediction models make use of both dynamic features (meteorological variables) and static features as input. For each county, static features include the county centroid longitude, latitude, and area, 82 soil features (for all six soil layers combined), crop scope (which is the number of  $5 \text{ km} \times 5 \text{ km}$  grids with maize planting), fertilizer use, and year. Feature engineering was applied to some of the input variables prior to their usage in ML models, including normalization, transformation, and combination.

In *normalization*, different features have different mean and standard deviation; feature normalization converts the raw data to standardized anomalies so that the transformed features would all have a mean of 0 and a standard deviation of 1. That is,  $z_{ij} = (x_{ij} - \mu_i) / \sigma_i$ , where  $x_{ij}$  and  $z_{ij}$  are respectively the raw and transformed values for the  $j$ th record of the  $i$ th feature and  $\mu_i$  and  $\sigma_i$  are the raw-value mean and standard deviation of the  $i$ th feature.

In *transformation*, some of the features will become more meaningful after we apply a function to them. For example, for the wind direction feature,  $-179^\circ$  differs from  $179^\circ$  by only  $2^\circ$ , and for which the sin and cos functions can be used to preserve the continuity. Another reason to apply a transformation function is to convert a feature to normal distribution. For example, the crop scope of a county has a very skewed distribution, and a logarithmic transformation leads to a distribution that is close to normal.

In *combination*, sometimes the effect of one feature may depend on another feature, and these features can be

combined to create more informative features. For example, the difference between the maximum and minimum temperatures is used to derive the diurnal amplitude of temperature and is included as an additional feature.

#### d. ML algorithms and model setting

We built multiple ML models for crop yield prediction. For all ML models, we applied the leave-one-year-out approach to form the training and test sets. More specifically, during the parameter selection, we select one year for testing and the previous year for validation, and the remaining 17 years are for training. The hyperparameter set through grid search that performs the best on the union of 19 validation sets is selected and reported in Table 1. We retrain 19 models with 18-yr data as training and make predictions on the holdout testing year. Each data point appears once in the union of testing sets of 19 models and they are not shown in the parameter selection and training stage of each corresponding model to avoid sampling bias and data leakage.

##### 1) RF

Random forest is a machine learning algorithm for classification and regression tasks that consist of a large number of decision trees (Breiman 2001). Each decision tree is based on a bootstrap sample of the dataset and a subset of features, and it recursively splits the data space until predefined depth is reached, or the sample size of a node is small enough. Each partition is based on one feature and corresponding threshold that can minimize the entropy or Gini impurity. The average value for each leaf node is used for the prediction of the data points that fall into this node. After each regression tree is grown, the average prediction of all trees is used to do the final prediction of the random forest model.

In this study, we concatenate the dynamic features and static features for each county as the input vector, to train the RF model that predicts yield for the counties. The detailed parameters of the RF model are described in Table 1.

##### 2) XGBOOST MODEL

Gradient boosting (Friedman 2001) is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models. Unlike the random forest where each tree is independent of the other, Xgradient boosting combines weak learners into a single strong learner in an iterative fashion, and the latter iterations are trying to predict the residual, which is the difference between the predicted value and true label. XGBoost (Chen and Guestrin 2016) is a scalable machine learning system for tree boosting that improves the gradient boosting algorithm. It includes regularization items in the objective function, and shrinkage technique to prevent overfitting. Besides, several techniques such as second-order derivatives in the loss function and parallel learning substantially improve the efficiency.

Similar to the RF model, we combine the dynamic features and static features as inputs to train the XGBoost model. The parameters of XGBoost model are slightly different from RF model and are detailed in Table 1.

TABLE 1. The hyperparameters of different models.

	Hyperparameter	Value(s)
RF	No. of trees	1000
	Max tree depth	7
	Min sample split	6
	Min sample leaf	3
	Max feature no.	$\log_2(m)^a$
XGBoost	No. of trees	1000
	Max tree depth	7
	Column sample rate	0.8
	Learning rate	0.1
	Alpha	10
LSTMs	Dimension of LSTM states	64
	Dimension of FC layers $f_{\text{concat}}, f_{\text{init}}$	64, 1
	Dimension of FC layers $f_{\text{att}}$	$2T, T^b$
	Dimension of FC layers $f_{\text{pred}}$	256, 64, 1
	No. of epochs	50
	Batch size	128
	Learning rate	$1 \times 10^{-3},$ $1 \times 10^{-4},$ $1 \times 10^{-5c}$

<sup>a</sup>  $m$  is the total number of features.

<sup>b</sup>  $T$  is the total time steps.

<sup>c</sup> Learning rate decays at iteration 1000, 5000.

##### 3) LSTM

RNN is a class of artificial neural networks that can be used to model sequence data (Hopfield 1982). Connections between cells in an RNN form a directed graph along a temporal sequence and RNNs use the internal state (memory) of cells to capture temporal dynamics and dependencies. LSTM (Hochreiter and Schmidhuber 1997) is a special kind of RNN, capable of learning long-term dependencies. An LSTM cell includes two internal states: a cell state that is a vector designed to keep the long-term memory and a hidden state that is the output vector of the cell presenting the current working memory. At every time step, an LSTM cell takes the corresponding time series data and two states of the last LSTM cell as the inputs, then updates the states with an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. Figure 1 provides the detailed structures of the LSTM and its cells; the mathematical formulas are presented in Eqs. (1)–(6). For the *forget gate*,

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f). \quad (1)$$

For the *input gate*,

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \quad \text{and} \quad (2)$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c). \quad (3)$$

For the *update cell state*,

$$\mathbf{c}_t = \mathbf{f}_t * \mathbf{c}_{t-1} + \mathbf{i}_t * \tilde{\mathbf{c}}_t. \quad (4)$$



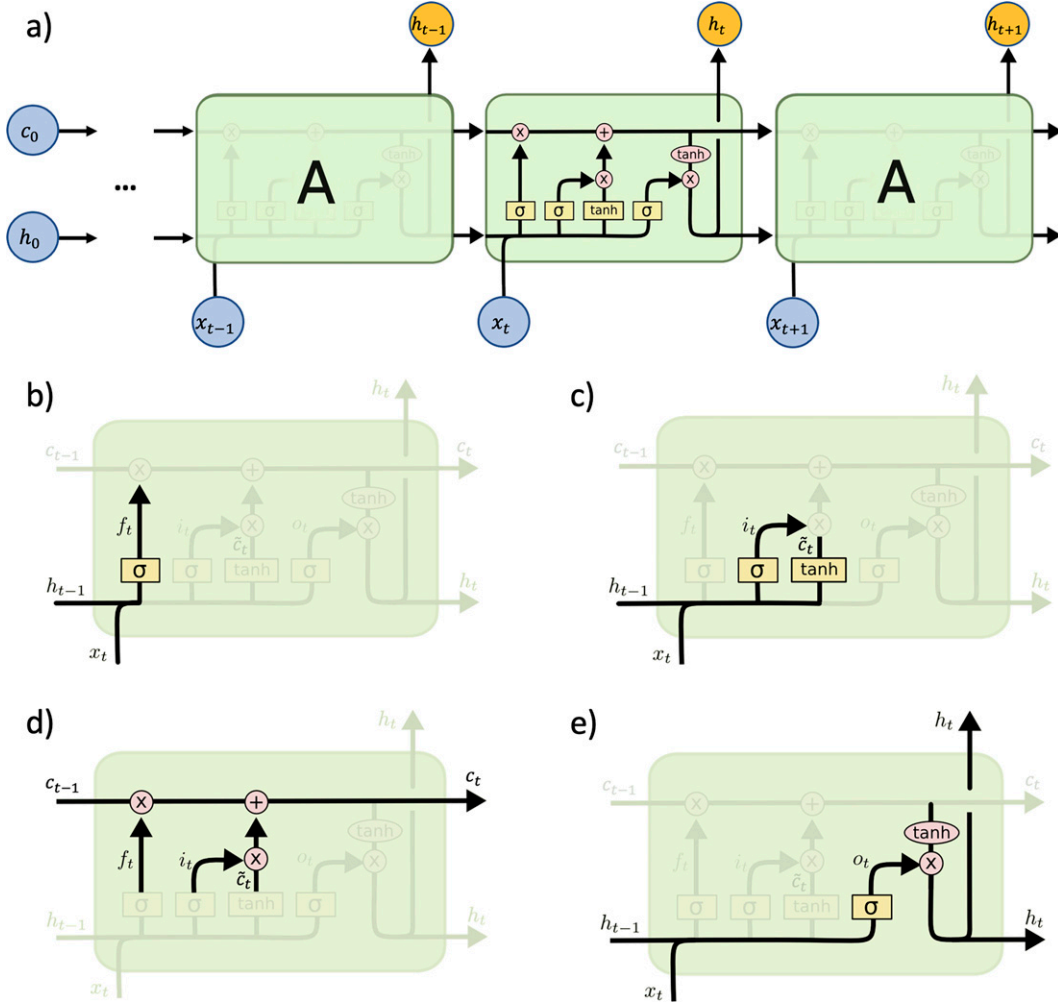


FIG. 1. Overall structure of LSTM network and detailed LSTM cell components [adapted from Olah (2015)]: (a) stack LSTM cells to build LSTM network and process time series  $(x_1, x_2, \dots, x_T)$ , (b) forget gate, (c) input gate, (d) update cell state, and (e) output gate.

For the *output gate*,

$$o_t = \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) \quad \text{and} \quad (5)$$

$$\mathbf{h}_t = \mathbf{o}_t * \tanh(\mathbf{c}_t). \quad (6)$$

Here, assume we have  $v$  time series features and set the hidden dimension as  $d$ . Then,  $\mathbf{x}_t$  is a vector of dimension  $v$ , and  $\mathbf{h}_{t-1}$  and  $\mathbf{h}_t$  are the hidden states at time steps  $t-1$  and  $t$ , with a dimension of  $d$ . The  $\mathbf{c}_{t-1}$ ,  $\tilde{\mathbf{c}}_t$ , and  $\mathbf{c}_t$  are also vectors with a dimension of  $d$ , and they stand for cell state at time step  $t-1$ , intermediate data, and cell state at time step  $t$ , respectively. Learnable parameters  $\mathbf{W}_f$ ,  $\mathbf{W}_i$ ,  $\mathbf{W}_c$ , and  $\mathbf{W}_o$  are matrices with a dimension of  $(d, d+v)$ , and  $\mathbf{b}_f$ ,  $\mathbf{b}_i$ ,  $\mathbf{b}_c$ , and  $\mathbf{b}_o$  are vectors with a dimension of  $d$ . Those parameters are randomly initialized and will be iteratively updated during neural network training with gradient descent algorithm. For Eqs. (1)–(8),  $[\cdot]$  is the concatenate function that takes two vectors as inputs and outputs one vector,  $\sigma(\cdot)$  is the sigmoid function, and  $\tanh(\cdot)$  is the

hyperbolic tangent function; the dot indicates matrix multiplication, and the asterisk is elementwise multiplication.

We developed three LSTM variations as depicted in Fig. 2, which correspond to three different ways to aggregate the static (non-time-series) features vector  $\mathbf{s}$  into modeling. In the first variant (LSTM<sub>concat</sub>), we process the static features with fully connected layers  $\mathbf{f}_{\text{concat}}(\cdot)$  and then concatenate the output with the last hidden state of LSTM layers to get final state  $\mathbf{z}$ :

$$\mathbf{l} = \mathbf{f}_{\text{concat}}(\mathbf{s}) \quad \text{and} \quad (7)$$

$$\mathbf{z} = [\mathbf{h}_T, \mathbf{l}]. \quad (8)$$

The second solution (LSTM<sub>init</sub>) applies fully connected layers  $\mathbf{f}_{\text{init}}(\cdot)$  to process the static features and uses the output to initialize the cell state and hidden state of LSTM layers:

$$\mathbf{c}_0, \mathbf{h}_0 = \mathbf{f}_{\text{init}}(\mathbf{s}), \quad \text{and} \quad (9)$$

$$\mathbf{z} = \mathbf{h}_T. \quad (10)$$

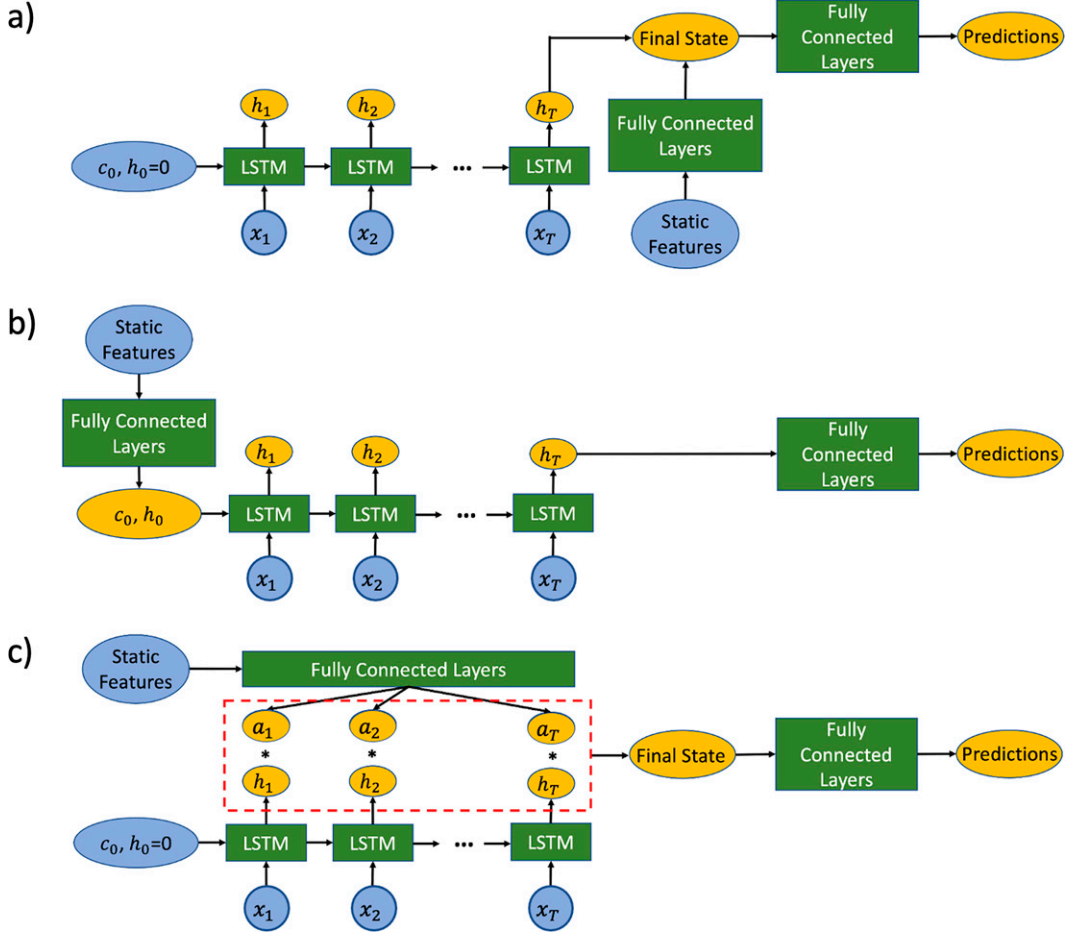


FIG. 2. Three variants of the LSTM model: (a) LSTM<sub>concat</sub> processes the static features with fully connected layers and concatenates the output with the last hidden state as final state, (b) LSTM<sub>init</sub> uses static features to initialize the cell state and hidden state, and (c) LSTM<sub>att</sub> uses static features to learn attention weights ( $a_1, a_2, \dots, a_T$ ) and calculates the weighted sum of all hidden states ( $h_1, h_2, \dots, h_T$ ) as final state. The rectangles (green) represent neural networks, and ovals represent data (blue for given features and orange for outputs of neural network layers).

The third approach (LSTM<sub>att</sub>) employs the attention mechanism after LSTM layers to provide a shortcut connection to aggregate all the hidden states and represents our new contribution toward ML model development. Instead of making predictions based on the last hidden state of LSTM, the attention allows neural networks to take the hidden states of all time steps as inputs. More precisely, we apply fully connected layers  $\mathbf{f}_{\text{att}}()$  to process the static features and learn a weight for each time step ( $a_1, a_2, \dots, a_T$ ). We then use the weighted sum of the hidden states as the final state:

$$(a_1, a_2, \dots, a_T) = \mathbf{f}_{\text{att}}(\mathbf{s}) \quad \text{and} \quad (11)$$

$$\mathbf{z} = \sum_{i=1}^T a_i * \mathbf{h}_i. \quad (12)$$

We further process the final state  $\mathbf{z}$  with fully connected layers  $\mathbf{f}_{\text{pred}}()$  to make the prediction  $\mathbf{y}_{\text{pred}}$  for all of these three variations:

$$\mathbf{y}_{\text{pred}} = \mathbf{f}_{\text{pred}}(\mathbf{z}). \quad (13)$$

We use data augmentation methods to alleviate model overfitting in LSTM. In the training of a neural network, one epoch consists of enough iterations so all examples in the training set have been passed through in backpropagation once. In each epoch, we generate a variation for each training example that differs from the original example in two ways: adding a Gaussian noise to time series variables, randomly shifting the time series variables, and filling in paddings at the beginning or end of the time series segment so that all records have the same length. In both ways, we keep the observed yield value unchanged as label. We fit a LSTM model with one LSTM layer. Then the final state from any variant is processed with fully connected layers. We include BatchNorm, rectified linear unit, and Dropout layer after fully connected layers except those that output the attention or prediction. More detailed network structure and training procedure are described in Fig. 1 and Table 1.

### e. Performance metrics

To evaluate the ML and DSSAT performance, we compare the predicted yield value  $y_{\text{pred}}$  with the observed value  $y_{\text{true}}$  and estimate four performance metrics, the Pearson correlation coefficient  $r$ , the coefficient of determination  $R^2$ , root-mean-square error (RMSE), and the mean absolute error (MAE).

The Pearson correlation coefficient  $r$  measures the linear correlation between two variables with a value between  $-1$  and  $+1$ . It can be calculated as (here  $y_{\text{true},i}$  and  $y_{\text{pred},i}$  are the observed and predicted values for the  $i$ th data record, and  $\bar{y}_{\text{true}}$  and  $\bar{y}_{\text{pred}}$  are their corresponding averages)

$$r = \frac{\sum_{i=0}^n (y_{\text{true},i} - \bar{y}_{\text{true}})(y_{\text{pred},i} - \bar{y}_{\text{pred}})}{\sqrt{\sum_{i=0}^n (y_{\text{true},i} - \bar{y}_{\text{true}})^2} \sqrt{\sum_{i=0}^n (y_{\text{pred},i} - \bar{y}_{\text{pred}})^2}}. \quad (14)$$

The coefficient of determination  $R^2$  measures how much variance in the predictand  $\bar{y}_{\text{true}}$  can be explained by the predictor  $\bar{y}_{\text{pred}}$  in a linear regression fit, and it can be estimated as

$$R^2 = 1 - \frac{\sum_i [y_{\text{true},i} - f(y_{\text{pred},i})]^2}{\sum_i (y_{\text{true},i} - \bar{y}_{\text{true}})^2}. \quad (15)$$

In the best-case scenario ( $R^2 = 1$ ), the modeled values exactly match the observed values. A baseline model, which always predicts  $\bar{y}_{\text{true}}$  will have  $R^2 = 0$ . Models that have worse predictions than the baseline will have a negative  $R^2$ .

RMSE and MAE are two metrics measuring the differences in magnitude between predicted and observed values. They are respectively estimated as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{\text{true},i} - y_{\text{pred},i})^2} \quad \text{and} \quad (16)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_{\text{true},i} - y_{\text{pred},i}|. \quad (17)$$

## 3. Results

Accounting for the spatiotemporal data of the crop yield, the performance metrics for each model are listed in Table 2. Among the multiple ML models trained, the LSTM<sub>att</sub> performs the best in capturing both the spatiotemporal variability (with the highest correlation coefficient with the observed data) and the magnitude of the yield (with the lowest RMSE and MAE). The XGBoost performs slightly worse than LSTM<sub>att</sub>. The RF model is notably less skillful than LSTM<sub>att</sub> and XGBoost but produces better prediction than the other two types of LSTM (LSTM<sub>concat</sub> and LSTM<sub>init</sub>).

All machine learning models trained in this study outperform the process-based model DSSAT. The DSSAT yield error is 2–3 times as large as the ML yield errors. Based on the spatiotemporal correlation, DSSAT prediction explains only

TABLE 2. The overall performance of different models; the values within parentheses indicate the standard deviation.

	$r$	$R^2$	RMSE	MAE
RF	0.70 (0.00)	0.46 (0.00)	26.79 (0.04)	20.91 (0.04)
XGBoost	0.78 (0.00)	0.60 (0.00)	23.03 (0.14)	17.40 (0.10)
LSTM <sub>concat</sub>	0.55 (0.00)	0.27 (0.01)	31.12 (0.14)	24.67 (0.14)
LSTM <sub>init</sub>	0.61 (0.01)	0.37 (0.02)	28.90 (0.42)	21.97 (0.13)
LSTM <sub>att</sub>	0.85 (0.00)	0.73 (0.00)	18.96 (0.16)	14.14 (0.08)
LSTM <sub>subset</sub>	0.85 (0.00)	0.72 (0.01)	19.15 (0.24)	14.42 (0.15)
DSSAT	0.40	0.16	63.20	49.57

16% of the observed yield variance, while the ML models explain up to 73% of the observed variance. Although the ML models do include more meteorological variables as input than DSSAT, that is not the cause for the performance difference between the two. In fact, in an LSTM<sub>att</sub> sensitivity test reducing the input variables to the same subset as used by DSSAT (LSTM<sub>subset</sub>), the model performance metrics show no statistically significant changes according to a  $t$  test we conducted on the RMSE with a  $p$  value of 0.40. This indicates that the additional input variables to ML models are redundant and do not carry new information.

Because of the presence of redundant variables in the ML models, removing one or more presumably important input does not cause performance deterioration, suggesting a high degree of equifinality. For example, interpretable LSTM experiments identified wind direction as the most important feature. This finding is unexpected since wind direction does not influence the surface water and energy fluxes or plant physiological processes, but it is also understandable since wind direction reflects the regional atmospheric circulation pattern that heavily influences the meteorological variables (e.g., precipitation, solar radiation, and temperature) that control plant physiological processes. Removing wind direction from the model input does not reduce the prediction accuracy, because the same information is likely covered by the other variables. The rest of the analysis focuses on a more detailed comparison between DSSAT and ML models using LSTM<sub>att</sub> as an example.

Figure 3 shows the models' performance in capturing the long-term mean yield in maize-growing counties with sufficient observational data. Counties where less than 10% of the land is used to grow maize and counties with less than 10 years of observed yield data are excluded from this comparison. The ML-predicted yield is remarkably similar to observations and shows no clear spatial pattern of model biases. The DSSAT model shows systematic biases with a dipole pattern, underestimating the long-term mean yield on the southwestern side of the Corn Belt and overestimating on the northeastern side; for most counties, the relative error is within 20%.

Averaged across the Corn Belt, maize yield has been increasing with time (Fig. 4a), due primarily to increased fertilizer applications and seed technology advancement. The ML models learn this trend from data and can therefore accurately predict the year-to-year variation and trend of yield; DSSAT is biased high in the early stage of the study period and biased low in the later stage, producing a much weaker

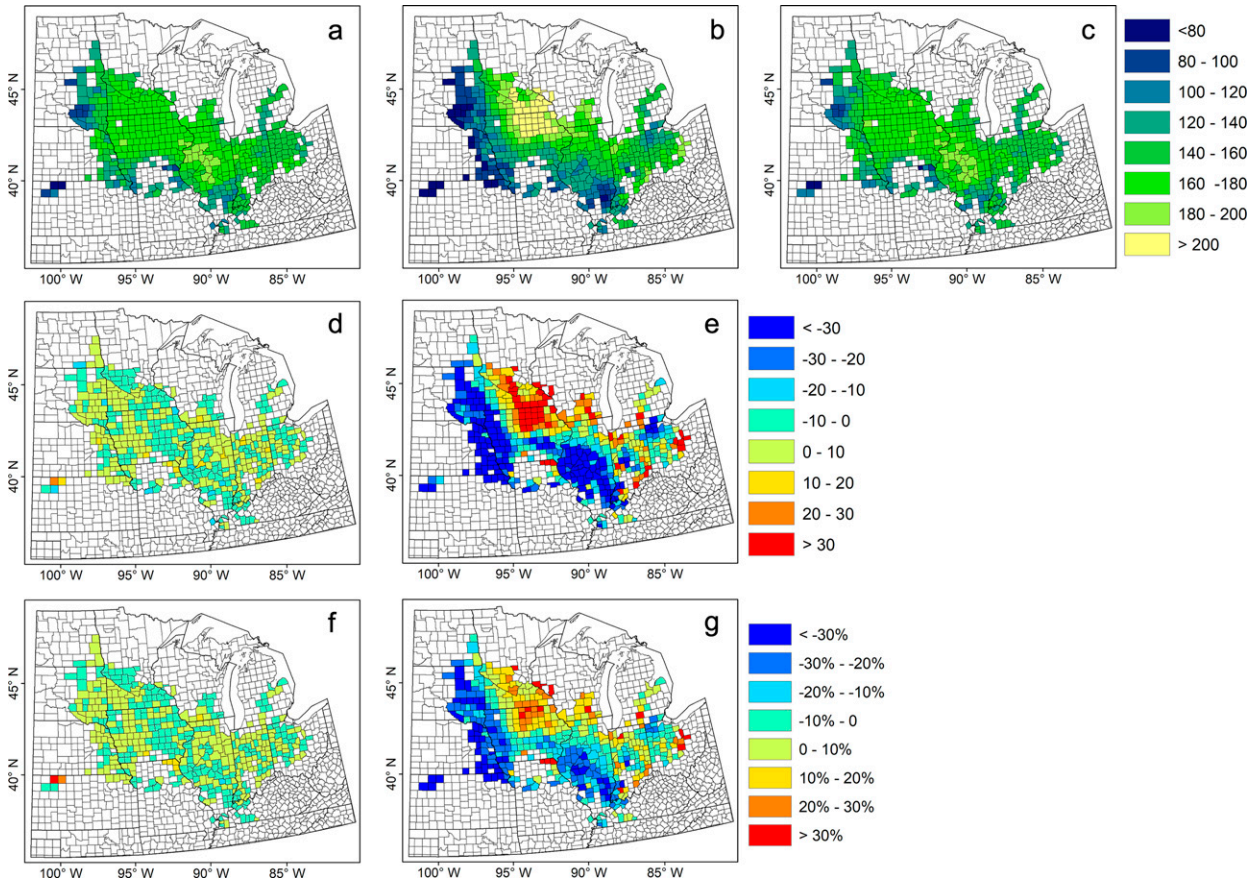


FIG. 3. Long-term mean of county-level maize yield ( $\text{bushels acre}^{-1}$ ;  $1 \text{ bushel acre}^{-1} = 67.25 \text{ kg ha}^{-1}$ ): (a) ML model yield, (b) DSSAT model yield, (c) observed yield, (d) ML model absolute bias, (e) DSSAT model absolute bias, (f) ML model relative bias, and (g) DSSAT model relative bias. Counties with less than 10 years of observed yield data or counties where maize is grown over less than 10% of its area are omitted from the plotting.

trend than observed. To examine the yield response to the interannual variability of meteorological forcing, we compare the detrended anomalies (Fig. 4b) between observation and model predictions. It is evident that the ML-predicted yield response is close to observations, while DSSAT overestimates the sensitivity of crop yield to hydrometeorological variation. The greatest disparity is found in the year 2012 when a severe drought caused widespread crop failure in the Midwest and Great Plains. The DSSAT-predicted yield loss is  $\sim 80\%$  more severe than the observed.

The spatially distributed Pearson correlation coefficient provides an objective assessment for the model skill in reproducing the interannual variability of crop yield (Fig. 5). Estimated based on the raw time series (which lumps the trend and interannual variability together), the correlation coefficient for most counties is in the range of 0.25–0.75 for DSSAT and larger than 0.75 for LSTM<sub>att</sub> (Figs. 5a,c). Removing the trend in both observation and model prediction leads to an increase of the correlation coefficient for DSSAT and a decrease for LSTM<sub>att</sub>, but ML still outperforms the process-based model in simulating the interannual variability (Figs. 5b,d). Interestingly, northern Iowa and southern Minnesota are

identified by both models as a region where yield is challenging to predict.

Figure 6 shows the models' skill in capturing the yield spatial variability and how the skill varies from year to year, which confirms the great advantage of the ML model over DSSAT. Detrending the data reduces the RMSE and MAE (especially for DSSAT in 2012) but has negligible impact on the spatial correlation. In most of the years, the spatial correlation coefficient for LSTM<sub>att</sub> is in the range of 0.75–0.9 and fluctuates in the range of 0.2–0.7 for DSSAT; the RSME and MAE in DSSAT are 2–4 times as large as those in LSTM<sub>att</sub>. The extreme drought year 2012 offers an opportunity to test the model performance in scenarios from which little data are available. Both RMSE and MAE suggest a clear performance drop for both the ML and DSSAT models in 2012. Because the ML model training leaves one year out for validation, in predicting the 2012 yield the LSTM<sub>att</sub> (trained based on the 18 years other than 2012) was applied to an extreme year that has no counterpart in the training data. This outlier application explains the dip in the ML prediction skill in 2012. Although process-based models are equally applicable to all scenarios, the DSSAT RMSE and MAE in 2012 are clearly higher (before



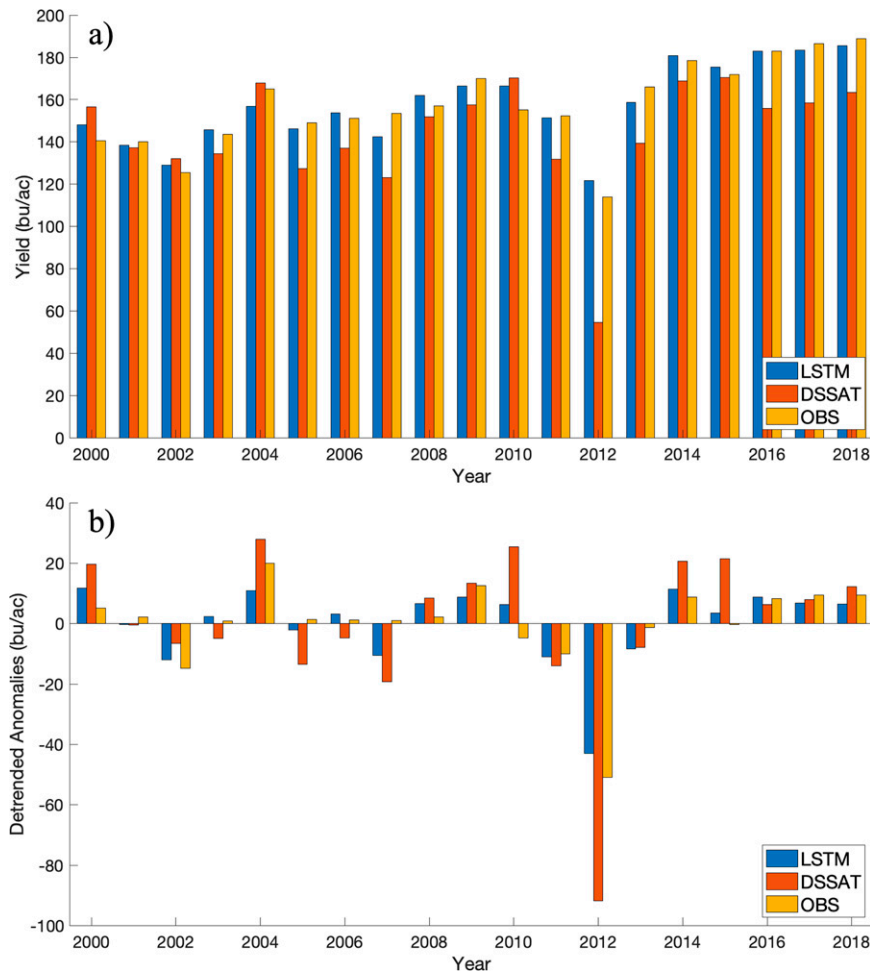


FIG. 4. Planting area-weighted average of yield across the Corn Belt from LSTM, DSSAT, and observations: (a) raw yield and (b) detrended anomalies.

detrending) than other years, which is a result of the model being overly sensitive to the interannual variability of meteorological forcing (Franke et al. 2020; M. Yang and G. Wang 2022, unpublished manuscript). Interestingly, the spatial correlation coefficient in 2012 is relatively high when compared with other years for both DSSAT and ML models. For predicting crop yield during an extreme drought, DSSAT and LSTM<sub>att</sub> show a similar level of skills in capturing the yield spatial variability, but the magnitude of model errors in DSSAT is larger.

For applications in climate impact assessment, meteorological forcing for the entire growing season is available as input to the yield prediction models. For real-time yield forecasts, observed meteorological forcing data are available for only part of the growing season. To test the ML model performance ahead of the USDA production estimates, we use different lengths of the meteorological forcing data starting from 10 April in all five ML models (Fig. 7). The LSTM<sub>att</sub> model still performs the best among all five models, and the prediction skill increases as the growing season progresses up to mid-August (18 weeks after 10 April); meteorological data from the later stage of the growing season do not bring additional

benefit. This statement also holds for the LSTM<sub>concat</sub> model. In contrast, the XGB, RF, and LSTM<sub>init</sub> models rely more on county features and show little sensitivity to the length of the meteorological forcing data or the forecast time within the growing season. Another observation is that the RMSE of LSTM<sub>init</sub> increases as more data from the growing season are added. This is due to LSTM<sub>init</sub> applying the static features to initialize the hidden state; as the record of dynamic features (meteorological data) gets longer, additional noise is added and the signal from the static features becomes relatively weaker. This also highlights the importance of well-designed deep learning models such as our LSTM<sub>att</sub> that are capable of properly handle both time series features and static features. The LSTM<sub>att</sub> model's robust performance at forecast times well ahead of the harvest season suggests that the model can support both real time forecast and long-term climate impact assessment.

#### 4. Conclusions and discussion

In this study, we build multiple ML models to predict maize crop yield in the U.S. Corn Belt based on meteorological

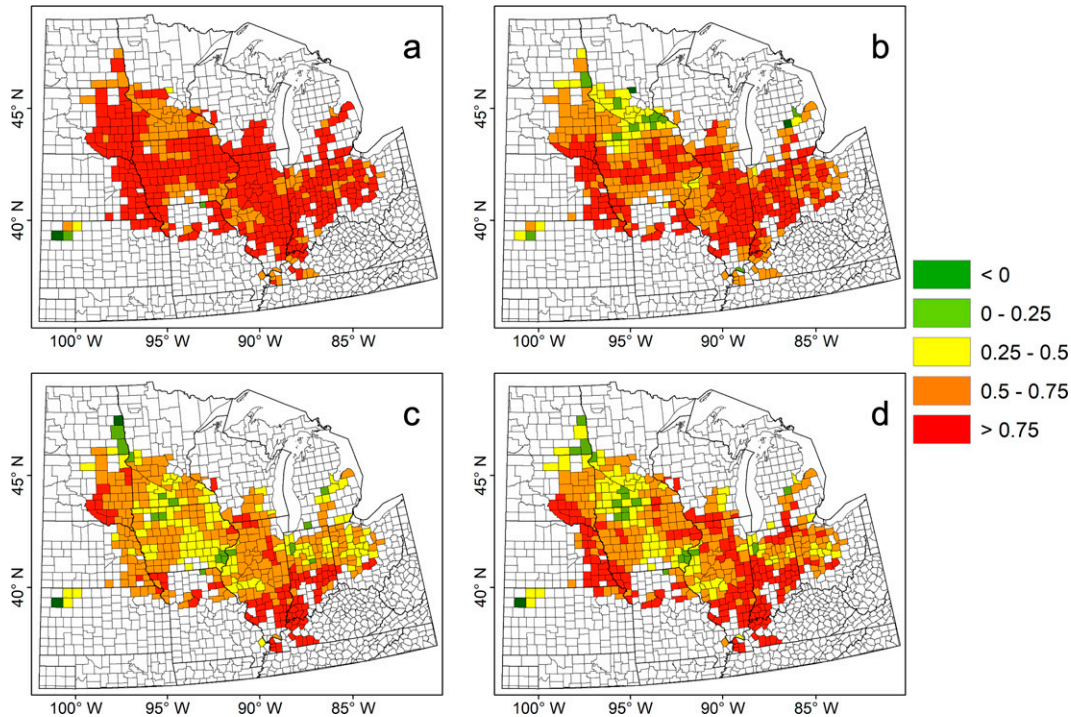


FIG. 5. The Pearson's correlation coefficient  $r$  between the true label and predicted yield of 19 years in each county, which indicates that the ML model captures the interannual variation better than DSSAT: (a) ML model before detrending, (b) ML model after detrending, (c) DSSAT model before detrending, and (d) DSSAT model after detrending. Counties with less than 10 years of observed yield data or counties where maize is grown over less than 10% of its area are omitted from the plotting.

forcing and soil properties, assess the ML prediction skills using the leave-one-year-out training and validation approach, and compare the ML performance with a process-based crop model DSSAT. Overall, the ML models outperform DSSAT by a large margin in reproducing the long-term mean, trend, and spatiotemporal variability of maize yield. Results from sensitivity experiments indicate that the superb prediction skills are intrinsic to the ML algorithms and not due to the intake of more meteorological variables. The process-based model DSSAT, although considered the best of its kind for simulating crops in the United States, extracts only a small fraction of useful information embedded in the soil and meteorological variables.

When applied to an extreme drought year that has no counterpart in the training data, the ML performance drops, but still has a clear advantage over DSSAT. The considerable prediction skill under the extreme drought scenario serves as strong evidence for the potential applicability of ML models in projecting crop yield in a future climate that may be outside the natural range of variability observed in the past.

In comparison with previous existing deep learning methods, our proposed LSTM<sub>att</sub> method is able to integrate both time series features and static features more effectively and has fewer assumption constraints. For example, the CNN-LSTM model of Sun et al. (2019) considered dynamic features (time series) only. We consider it important to also include static

features such as soil hydraulic properties, because crop yield under the same meteorological forcing can differ significantly between sandy soil and clayish soil, for example. In our study, we integrate the static features into the LSTM mode using the novel LSTM<sub>att</sub> model, which achieved the best overall prediction performance and is also robust under extreme weather conditions. The CNN-RNN framework of Khaki et al. (2020) relied on the average yield of the current year as an input variable in the training stage and used the average yield of the previous year as a substitute to make predictions in the test stage. This substitution assumption would be problematic for years of extreme conditions that may cause major crop loss or even failure such as 2012. Our proposed LSTM<sub>att</sub> model does not include recent yield as an input variable, which makes it more robust and suitable for application to studies of climate change and extremes.

This study, especially the performance comparison between the ML models and DSSAT, is subject to a few uncertainties. Most process-based models such as DSSAT were developed based on plot-level field data (some decades old) but were applied to larger spatial scales. In this study, the DSSAT model for maize was run on a 5-arc-min (9 km by 9 km) grid system, and the county level yield was then derived (as the average of the grid-level yields weighted by the area of cropland within each grid cell) for use in model calibration and verification against county-level observations. While the effect of the scale

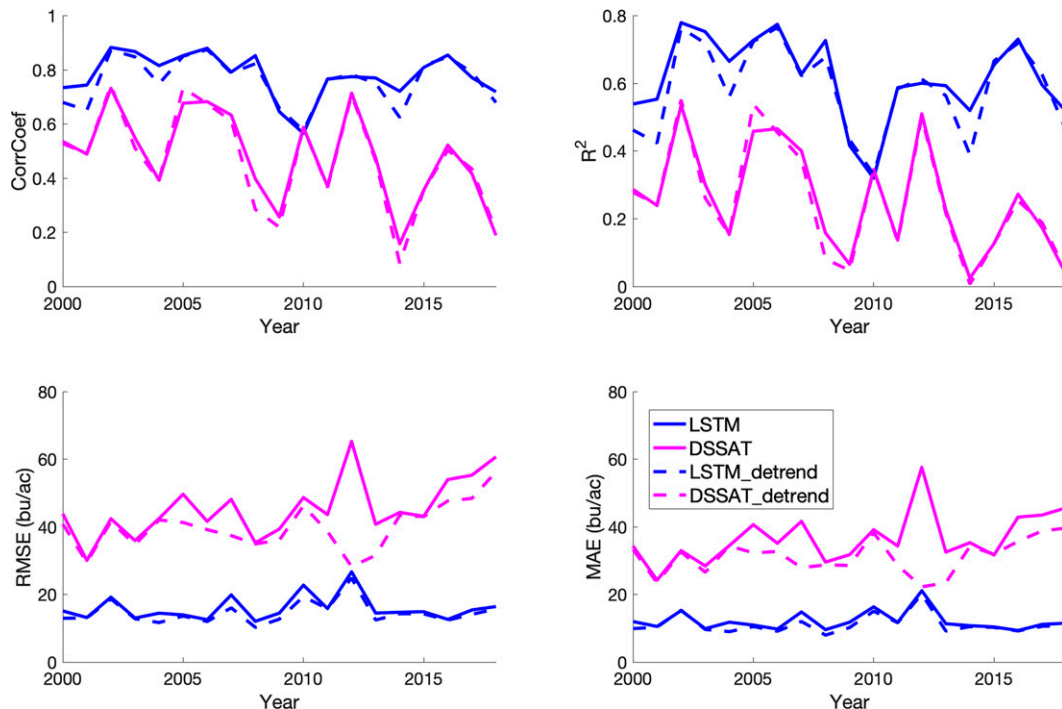


FIG. 6. LSTM<sub>att</sub> and DSSAT model performance metrics in different years, calculated on the basis of comparison between model and observed county level maize yield across the Corn Belt in counties where maize is grown over at least 10% of its area.

mismatch (field vs grid) may be partially compensated for by calibration, it remains a limitation of the DSSAT application in this study. Another source of uncertainty for DSSAT modeling has to do with agricultural practices such as planting date, fertilizer application, and irrigation that tend to vary substantially across fields at scales much smaller than the size of the 5-arc-min grid used here.

Our study focuses on rain-fed agriculture and did not account for the effect of irrigation in both the DSSAT and ML models.

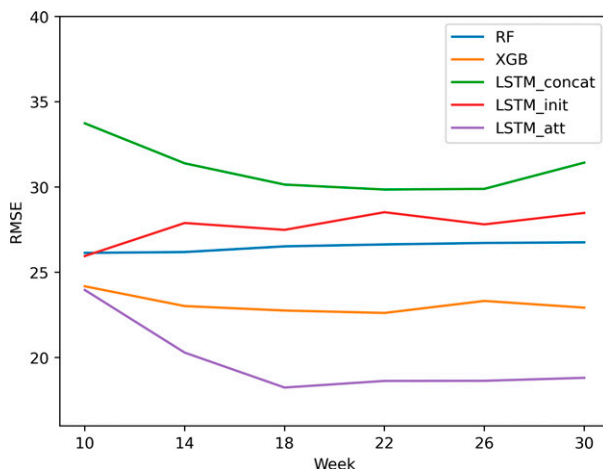


FIG. 7. Model performance at different forecast times (in weeks since 10 Apr) during the growing season. Note that week 10 corresponds to 19 Jun.

Although irrigation can have a great influence on crop yield, the percentage of irrigated farmland in the Corn Belt region is fairly small according to surveys from USDA–NASS and other fine-resolution irrigation map products such as Xie et al. (2021). Despite some increases in the past decades, currently irrigation measures in the Corn Belt region are primarily deployed in Nebraska and Kansas. The relatively low model performance in the counties located in Kansas may be due to the lack of consideration for irrigation in the models. In eastern Nebraska, despite the high irrigation extent, irrigation frequency is low (Xie et al. 2021), so our models can still perform well in reproducing the observed yield. Elsewhere in the Corn Belt, irrigation measures are extremely sparse and rain-fed agriculture is therefore representative.

Another potential source of uncertainty is the planted area information. We used the planted area in the year 2010 as the grid weight to derive the county-level yield from the grid-level DSSAT simulations, and as a static input feature (at the county level) for the ML models. However, agricultural land use is a dynamic process, and the Corn Belt experienced episodes of expansion and reduction of maize planting including for example the biofuels boom (Motamed et al. 2016). Agricultural expansion may lead to planting over marginal land with low productivity, so average yields may decrease with the planted area (Cabas et al. 2010), which complicates our assessment of the model performance. To separate the impact of this factor, we conducted sensitivity analyses including removing the weight from the spatial averaging of DSSAT yield and removing the planted area information from the

ML model input variables. However, the model performance metrics showed negligible changes. Therefore, agricultural expansion or abandonment is likely not a major factor influencing the variability of county-average yield.

Our study adds to a rapidly growing body of literature on ML models for crop yield prediction. The development of the innovative LSTM<sub>att</sub> algorithm sets our study apart from other studies using LSTM models (e.g., Jiang et al. 2020; Luo et al. 2022; Schwalbert et al. 2020; Wang et al. 2020). Based on soil features and meteorological forcing variables and without relying on vegetation remote sensing, our LSTM<sub>att</sub> models are applicable to not only real time yield forecast but also long-term climate impact assessment. In contrast, most existing ML-based crop yield forecast models make use of both meteorological forcing data and vegetation remote sensing (Jeong et al. 2022; Kuwata and Shibasaki 2015; Paudel et al. 2022; Sun et al. 2020; Zhou et al. 2022), and therefore cannot be used for future impact assessment when remotely sensed vegetation index is not available. For example, Sun et al. (2020) proposed a multilevel deep learning model coupling RNN and CNN to extract spatiotemporal features of remote sensing data, weather data, and soil property data for county-level corn yield estimation in the U.S. Corn Belt. Jiang et al. (2020) developed a LSTM model that integrates heterogeneous crop phenology, meteorology, and remote sensing data to estimate county-level corn yields in the U.S. Corn Belt. Zhou et al. (2022) utilized climate variables and remote sensing-derived metrics as inputs for three machine learning methods to predict wheat yield in 1582 counties across China. Jeong et al. (2022) proposed a combined 1D CNN and LSTM model for early prediction of rice yield at pixel scale in South and North Korea using satellite vegetation indices and meteorological and geospatial data as predictors. It was found that remotely sensed vegetation data may be a more important contributor to yield forecast skills than meteorological forcing data (Jiang et al. 2020; Zhang et al. 2020). However, with the attention channel, our LSTM<sub>att</sub> already performs well even without the benefit of remote sensing data; our follow-up research will examine how incorporating remote sensing data into the LSTM<sub>att</sub> model may further improve its performance in real time yield forecast.

**Acknowledgments.** This work was supported by the University of Connecticut (UConn) Civil and Environmental Engineering Department Research Initiative and the UConn Center for Biological Risks summer grant. We thank the three anonymous reviewers for their constructive comments on an earlier version of this paper.

**Data availability statement.** Observational data for the county-level yield and harvest area are available from the USDA National Agricultural Statistics Service (NASS; <https://quickstats.nass.usda.gov/>). Subcounty-level harvest areas are from the latest Spatial Production Allocation Model (SPAM 2010 v1.0; <https://doi.org/10.7910/DVN/PRFF8V>), as cited in International Food Policy Research Institute (2019). Data for nitrogen fertilizer are available from PANGAEA of the International Science Council World Data System (<https://doi.org/10.1594/PANGAEA.883585>), as cited in Cao et al. (2018). Data on soil features are from the Global High-Resolution Soil Profile Database for Crop Modeling Applications (v2.5; <https://doi.org/10.7910/DVN/1PEEY0>), as cited in Han et al. (2015). Data for surface meteorological variables are openly available from the gridMET dataset (<https://www.northwestknowledge.net/metdata/data/>), as cited in Abatzoglou (2013).

## REFERENCES

- Abatzoglou, J. T., 2013: Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.*, **33**, 121–131, <https://doi.org/10.1002/joc.3413>.
- Asseng, S., and Coauthors, 2015: Rising temperatures reduce global wheat production. *Nat. Climate Change*, **5**, 143–147, <https://doi.org/10.1038/nclimate2470>.
- Attia, A., and Coauthors, 2021: Sensitivity of the DSSAT model in simulating maize yield and soil carbon dynamics in arid Mediterranean climate: Effect of soil, genotype and crop management. *Field Crop. Res.*, **260**, 107981, <https://doi.org/10.1016/j.fcr.2020.107981>.
- Bassu, S., and Coauthors, 2014: How do various maize crop models vary in their responses to climate change factors? *Global Change Biol.*, **20**, 2301–2320, <https://doi.org/10.1111/gcb.12520>.
- Battude, M., A. Al Bitar, D. Morin, J. Cros, M. Huc, C. M. Sicre, V. Le Dantec, and V. Demarez, 2016: Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sens. Environ.*, **184**, 668–681, <https://doi.org/10.1016/j.rse.2016.07.030>.
- Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Cabas, J., A. Weersink, and E. Olale, 2010: Crop yield response to economic, site and climatic variables. *Climatic Change*, **101**, 599–616, <https://doi.org/10.1007/s10584-009-9754-4>.
- Cao, P., C. Lu, and Z. Yu, 2018: Historical nitrogen fertilizer use in agricultural ecosystems of the contiguous United States during 1850–2015: Application rate, timing, and fertilizer types. *Earth Syst. Sci. Data*, **10**, 969–984, <https://doi.org/10.5194/essd-10-969-2018>.
- Chen, T., and C. Guestrin, 2016: XGBoost: A scalable tree boosting system. *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, New York, NY, Association for Computing Machinery, 785–794, <https://doi.org/10.1145/2939672.2939785>.
- Crane-Droesch, A., 2018: Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ. Res. Lett.*, **13**, 114003, <https://doi.org/10.1088/1748-9326/aae159>.
- Fieuzal, R., C. M. Sicre, and F. Baup, 2017: Estimation of corn yield using multi-temporal optical and radar satellite data and artificial neural networks. *Int. J. Appl. Earth Obs. Geoinf.*, **57**, 14–23, <https://doi.org/10.1016/j.jag.2016.12.011>.
- Filippi, P., and Coauthors, 2019: An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precis. Agric.*, **20**, 1015–1029, <https://doi.org/10.1007/s11119-018-09628-4>.
- Franke, J. A., and Coauthors, 2020: The GGCMI Phase 2 experiment: Global gridded crop model simulations under uniform changes in CO<sub>2</sub>, temperature, water, and nitrogen levels (protocol version 1.0). *Geosci. Model Dev.*, **13**, 2315–2336, <https://doi.org/10.5194/gmd-13-2315-2020>.



- Friedman, J. H., 2001: Greedy function approximation: A gradient boosting machine. *Ann. Stat.*, **29**, 1189–1232, <https://doi.org/10.1214/aos/1013203451>.
- Geipel, J., J. Link, and W. Claupein, 2014: Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. *Remote Sens.*, **6**, 10 335–10 355, <https://doi.org/10.3390/rs61110335>.
- Guo, T., T. Lin, and N. Antulov-Fantulin, 2019: Exploring interpretable LSTM neural networks over multi-variable data. *Proc. 36th Int. Conf. on Machine Learning*, Long Beach, CA, ICML, 2494–2504, <http://proceedings.mlr.press/v97/guo19b/guo19b.pdf>.
- Han, E., A. Ines, and J. Koo, 2015: Global high-resolution soil profile database for crop modeling applications. International Food Policy Research Institute Working Paper, 37 pp., <http://ebrary.ifpri.org/cdm/singleitem/collection/p15738coll2/id/129734>.
- He, K., X. Zhang, S. Ren, and J. Sun, 2016: Deep residual learning for image recognition. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, IEEE, 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
- Hochreiter, S., and J. Schmidhuber, 1997: Long short-term memory. *Neural Comput.*, **9**, 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hopfield, J. J., 1982: Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, **79**, 2554–2558, <https://doi.org/10.1073/pnas.79.8.2554>.
- International Food Policy Research Institute, 2019: Global spatially-disaggregated crop production statistics data for 2010 version 1.1. Harvard Dataverse, V3, accessed 18 September 2020, <https://doi.org/10.7910/DVN/PRFF8V>.
- International Research Institute for Climate and Society, Michigan State University, and HarvestChoice IFPRI, 2015: Global high-resolution soil profile database for crop modeling applications. Harvard Dataverse, accessed 16 August 2020, <https://doi.org/10.7910/DVN/1PEEY0>.
- Jeong, S., J. Ko, and J.-M. Yeom, 2022: Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea. *Sci. Total Environ.*, **802**, 149726, <https://doi.org/10.1016/j.scitotenv.2021.149726>.
- Jiang, H., and Coauthors, 2020: 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. *Global Change Biol.*, **26**, 1754–1766, <https://doi.org/10.1111/gcb.14885>.
- Jones, C. A., J. R. Kiniry, and P. T. Dyke, 1986: *CERES-Maize: A Simulation Model of Maize Growth and Development*. Texas A&M University Press, 198 pp.
- Jones, J. W., and Coauthors, 2003: The DSSAT cropping system model. *Eur. J. Agron.*, **18**, 235–265, [https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7).
- Khaki, S., L. Wang, and S. V. Archontoulis, 2020: A CNN-RNN framework for crop yield prediction. *Front. Plant Sci.*, **10**, 1750, <https://doi.org/10.3389/fpls.2019.01750>.
- , H. Pham, and L. Wang, 2021: Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. *Sci. Rep.*, **11**, 11132, <https://doi.org/10.1038/s41598-021-89779-z>.
- Kuwata, K., and R. Shibasaki, 2015: Estimating crop yields with deep learning and remotely sensed data. *2015 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Milan, Italy, IEEE, 858–861, <https://doi.org/10.1109/IGARSS.2015.7325900>.
- Lobell, D. B., J. M. Deines, and S. Di Tommaso, 2020: Changes in the drought sensitivity of US maize yields. *Nat. Food*, **1**, 729–735, <https://doi.org/10.1038/s43016-020-00165-w>.
- Luo, Y., and Coauthors, 2022: Accurately mapping global wheat production system using deep learning algorithms. *Int. J. Appl. Earth Obs. Geoinf.*, **110**, 102823, <https://doi.org/10.1016/j.jag.2022.102823>.
- Motamed, M., L. McPhail, and R. Williams, 2016: Corn area response to local ethanol markets in the United States: A grid cell level analysis. *Amer. J. Agric. Econ.*, **98**, 726–743, <https://doi.org/10.1093/ajae/aav095>.
- National Agricultural Statistics Service, 2010: Field crops usual planting and harvesting dates. USDA NASS Rep. 628, 51 pp., [https://www.nass.usda.gov/Publications/Todays\\_Reports/reports/fcdate10.pdf](https://www.nass.usda.gov/Publications/Todays_Reports/reports/fcdate10.pdf).
- Olah, C., 2015: Understanding LSTM networks. GitHub, <http://colah.github.io/posts/2015-08-Understanding-LSTMs>.
- Paudel, D., and Coauthors, 2022: Machine learning for regional crop yield forecasting in Europe. *Field Crop. Res.*, **276**, 108377, <https://doi.org/10.1016/j.fcr.2021.108377>.
- Qian, B., and Coauthors, 2019: Climate change impacts on Canadian yields of spring wheat, canola and maize for global warming levels of 1.5°C, 2.0°C, 2.5°C and 3.0°C. *Environ. Res. Lett.*, **14**, 074005, <https://doi.org/10.1088/1748-9326/ab17fb>.
- Rosenzweig, C., and Coauthors, 2014: Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proc. Natl. Acad. Sci. USA*, **111**, 3268–3273, <https://doi.org/10.1073/pnas.1222463110>.
- , and Coauthors, 2018: Coordinating AgMIP data and models across global and regional scales for 1.5°C and 2.0°C assessments. *Philos. Trans. Roy. Soc.*, **376A**, 20160455, <https://doi.org/10.1098/rsta.2016.0455>.
- Schlenker, W., and M. J. Roberts, 2009: Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proc. Natl. Acad. Sci. USA*, **106**, 15 594–15 598, <https://doi.org/10.1073/pnas.0906865106>.
- Schwalbert, R. A., T. Amado, G. Corassa, L. P. Pott, P. V. V. Prasad, and I. A. Ciampitti, 2020: Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agric. For. Meteorol.*, **284**, 107886, <https://doi.org/10.1016/j.agrformet.2019.107886>.
- Shahhosseini, M., R. A. Martinez-Feria, G. Hu, and S. V. Archontoulis, 2019: Maize yield and nitrate loss prediction with machine learning algorithms. *Environ. Res. Lett.*, **14**, 124026, <https://doi.org/10.1088/1748-9326/ab5268>.
- Sun, J., L. Di, Z. Sun, Y. Shen, and Z. Lai, 2019: County-level soybean yield prediction using deep CNN-LSTM model. *Sensors*, **19**, 4363, <https://doi.org/10.3390/s19204363>.
- , Z. Lai, L. Di, Z. Sun, J. Tao, and Y. Shen, 2020: Multilevel deep learning network for county-level corn yield estimation in the US Corn Belt. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **13**, 5048–5060, <https://doi.org/10.1109/JSTARS.2020.3019046>.
- Tovihoudji, P. G., P. B. Akponikpè, E. K. Agbossou, and C. L. Bièlders, 2019: Using the DSSAT model to support decision making regarding fertilizer microdosing for maize production in the sub-humid region of Benin. *Front. Environ. Sci.*, **7**, 13, <https://doi.org/10.3389/fenvs.2019.00013>.
- van Klompenburg, T., A. Kassahun, and C. Catal, 2020: Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.*, **177**, 105709, <https://doi.org/10.1016/j.compag.2020.105709>.

- Wang, X., J. Huang, Q. Feng, and D. Yin, 2020: Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches. *Remote Sens.*, **12**, 1744, <https://doi.org/10.3390/rs12111744>.
- Xie, Y., H. K. Gibbs, and T. J. Lark, 2021: Landsat-based Irrigation Dataset (LANID): 30 m resolution maps of irrigation distribution, frequency, and change for the US, 1997–2017. *Earth Syst. Sci. Data*, **13**, 5689–5710, <https://doi.org/10.5194/essd-13-5689-2021>.
- Yang, M., and Coauthors, 2020: The role of climate in the trend and variability of Ethiopia's cereal crop yields. *Sci. Total Environ.*, **723**, 137893, <https://doi.org/10.1016/j.scitotenv.2020.137893>.
- Yu, Q., and Coauthors, 2020: A cultivated planet in 2010—Part 2: The global gridded agricultural-production maps. *Earth Syst. Sci. Data*, **12**, 3545–3572, <https://doi.org/10.5194/essd-12-3545-2020>.
- Zhang, L., Z. Zhang, Y. Luo, J. Cao, and F. Tao, 2020: Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in China using machine learning approaches. *Remote Sens.*, **12**, 21, <https://doi.org/10.3390/rs12010021>.
- Zhou, W., Y. Liu, S. T. Ata-Ul-Karim, Q. Ge, X. Li, and J. Xiao, 2022: Integrating climate and satellite remote sensing data for predicting county-level wheat yield in China using machine learning methods. *Int. J. Appl. Earth Obs. Geoinf.*, **111**, 102861, <https://doi.org/10.1016/j.jag.2022.102861>.