

ORIGINAL ARTICLE

Biometry, Modeling, and Statistics

In-season crop phenology using remote sensing and model-guided machine learning

George Worrall¹ | Jasmeet Judge¹ | Kenneth Boote² | Anand Rangarajan³

¹Center for Remote Sensing, Agricultural and Biological Engineering Department, Univ. of Florida, FL 32601, USA

²Agricultural and Biological Engineering Dep., Univ. of Florida, Gainesville, FL 32601, USA

³Computer and Information Science and Engineering Dep., Univ. of Florida, Gainesville, FL 32601, USA

Correspondence

George Worrall, Center for Remote Sensing, Agricultural and Biological Engineering Department, University of Florida, FL 32601, USA.

Email: gworral@ufl.edu

Assigned to Associate Editor Yao Zhang.

Funding information

University of Florida Center for Remote Sensing; University of Florida Informatics Institute

Abstract

Accurate in-season crop phenology estimation (CPE) using remote sensing (RS)-based machine-learning methods is challenging because of limited ground-truth data. In this study, a biophysical crop model was used to guide neural network (NN)-based, in-season CPE. Using the Decision Support System for Agrotechnology Transfer (DSSAT), we conducted uncalibrated simulations for corn (*Zea mays* L.) across Iowa and Illinois in the U.S. Midwest with in-season weather and historical information for planting and harvest. We investigated guiding the NN CPE method with growth stage (GSTD) and water stress factor (WSF) outputs from these simulations. Results show that guided NNs are able to estimate onset and progression of phenological stages more accurately than an unguided baseline and a crop model-only method. GSTD guidance improved CPE during seasons when progress deviated from a regional average because of temperature but was detrimental during seasons of delayed planting and harvest. WSF guidance improved CPE during seasons when planting and harvest were delayed by heavy rainfall but performed less well during grainfill and mature stages. Neural network-based CPE guided by both GSTD and WSF provided the most accurate estimates for pre-emergence, emerged, silking, and grainfill stages as well as lower RMSE for the median stage transition date than reported in three full-season CPE studies. An accurate RS method for estimating planting could link DSSAT simulations to the current planting window and improve upon these results. This model-guided approach can be extended to other crops and regions to unlock in-season crop risk assessments that are directly linked to crop phenology.

1 | INTRODUCTION

CPE, crop phenology estimation; CPR, crop progress report; CS, cosine similarity; DSSAT, Decision Support System for Agrotechnology Transfer; DSSAT-RR, DSSAT-only ridge regression; EWSD, excess water stress; FPAR, fraction of photosynthetically active radiation; GDD, growing degree days; GSTD, growth stage; LSTM, long short-term memory; NDVI, normalized difference vegetation index; NN, neural network; nRT, near real time; NSE, Nash–Sutcliffe efficiency; RS, remote sensing; WOY, week of year; WSF, water stress factor.

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Agronomy Journal* published by Wiley Periodicals LLC on behalf of American Society of Agronomy.

Knowledge of crop phenology, particularly for important commodity crops that are grown globally, is vital to ensure food systems remain stable. With intercrop synchrony increasing, there is greater likelihood of extreme weather events disrupting global food supply (Mehrabi & Ramankutty, 2019). Real-time knowledge of crop phenology within a region can help to quantify crop vulnerability (Kamali et al.,

2018). When paired with accurate weather forecasting, this knowledge can aid anticipation and mitigation of harvest impact in a more synchronous global food chain.

Tracking crop phenology throughout a growing season is a task commonly done via ground surveys, such as the USDA crop progress reports (USDA-NASS, 2021a). In this study we use the terms ‘crop phenology’, common in remote sensing, and “crop growth stages”, common in crop modeling, to refer to both physiological stages of crop growth and planting and harvest processes. The USDA collectively groups these under the term “crop progress”. These surveys are labor intensive, as they are needed weekly to ensure timely tracking of regional crop progression through phenological stages. The large geographical areas that must be covered to produce comprehensive ground surveys are often prohibitive, and several top grain-producing countries do not have publicly available survey data. Major growing regions, however, can be observed using remote sensing (RS) platforms, which offer data globally and provide information in remote regions. With recent advances in within-season crop classification (Kerner et al., 2022), as well as the continuous expansion of freely available, high-resolution satellite data (Zhang et al., 2020), it is now possible to monitor cropland in near real time (nRT).

Remote sensing has been used to track crop phenology (Zeng et al., 2016; Liu et al., 2018), but RS-based crop phenology estimation (CPE) remains challenging because not all crop growth stages produce detectable canopy changes (Gao & Zhang, 2021) such as the silking to grainfill transition in corn (*Zea mays* L.). Additionally, data availability is a major limitation because crop phenology field trials are localized and costly (Gao & Zhang, 2021). Unlike remotely sensed crop mapping, where millions of examples are available each year, regional scale crop phenology data produced by ground surveys is correlated for adjoining regions so data must be grouped by growing season.

Remote sensing methods often target heuristic growth stages, such as “green-up” and “senescence” (e.g., Zeng et al., 2016; Liu et al., 2018), rather than phenology-based growth stages such as emergence and maturity. The relationships between green-up and senescence periods and phenological crop growth stages, however, differ between crop types and methods used (Gao & Zhang, 2021). Studies that do target phenology-based growth stages, such as Diao et al. (2021), aim to predict the median transition date for a given growth stage. While this simplifies the problem, it obscures the nature of growth stage transition and crop phenology within a region. Given that adverse conditions during some growth stages make the crop more vulnerable to yield loss than during other stages (e.g., silking in corn), accurate knowledge of the proportion of a region’s crop in a particular growth stage at any given time within a season is more important than the date at which 50% of the crop has transitioned. Moreover, the ± 10 -d reference performance window often cited in these studies

Core Ideas

- Remote sensing-based CPE with machine learning is challenged by limited training data.
- DSSAT crop model outputs provide additional guidance to NN-based CPE methods.
- An NN guided with GSTD and WSF from DSSAT outperforms unguided methods.
- GSTD improves CPE during seasons with abnormal temperatures; WSF improves CPE during rainfall-driven delays.
- More accurate planting date estimates will improve crop model-guided CPE during delayed seasons.

is itself as large as the most critical yield loss window for moisture deficit in crops such as corn (Claassen & Shaw, 1970).

Most existing RS-based crop phenology methods use either curve-based or trend-based approaches, which use satellite-derived vegetation indices that are related to crop growth (Gao & Zhang 2021). Curve-based studies have shown success in relating normalized difference vegetation index (NDVI) to corn phenological stages. Seo et al. (2019) and Diao (2020) used double logistic curve fitting methods to relate start-of-season, peak-of-season, and end-of-season reference points to corn phenology. Recently, Diao et al. (2021) used a hybrid matching model with both curve and shape-fitting methods to estimate crop phenology by transforming seasonal NDVI curves to match a known reference curve. Curve-based approaches, however, require a full season of observations and cannot provide nRT estimates. Trend-based methods, which can provide nRT estimates, often focus on specific growth stages such as emergence (Gao, Anderson, Daughtry, et al., 2020) or end-of-season (Gao, Anderson, & Hively, 2020). Others predict phenological timing with lag in excess of 7–24 d (Liu et al., 2018). There are few studies that produce nRT estimates for all growth stages, and existing trend-based nRT algorithms are reliant on RS data availability (Gao & Zhang, 2021). Given that some of the most critical periods of vulnerability in many crops last only a few days (Claassen & Shaw, 1970; Doss et al., 1974), nRT algorithms that are dependent upon RS data availability or produce a lagged estimate are not reliable enough to produce actionable CPE.

To resolve this RS data dependency, some studies have proposed nRT CPE methods that incorporate additional information and do not rely on curve- or trend-based techniques. For example, hidden Markov models track crop progress via weekly accumulated growing degree days (GDD) and remotely sensed NDVI (Shen et al., 2013). Recently, a neural network (NN) approach was implemented using weather and

soil information along with crop growth driver interactions and RS data for nRT CPE (Worrall et al., 2021).

Another approach to nRT CPE that is completely independent of RS data is to use crop models. Crop modeling suites, such as the Decision Support System for Agrotechnology Transfer (DSSAT) and the Agricultural Systems Production Simulator (Jones et al., 2003; Holzworth et al., 2014) partition crop life cycles into constituent growth stages to simulate crop sensitivity and response to meteorological conditions such as yield loss from drought or delay in flowering because of extended photoperiod. These models are driven by soil and weather information and provide estimates of crop growth stage timing and development. Crop models, however, must be calibrated to produce accurate simulations of in situ growth, and cultivar parameters typically have a 2–3 yr lifespan (Akhavizadegan et al., 2021). Moreover, in-season CPE using crop models requires planting and harvest dates that are not readily available in regions that lack ground survey data.

A hybrid approach to integrating biophysical models and NN methods, known as physics-guided machine learning, has been used in fields such as hydrology and materials science (Willard et al. (2020)). While studies have shown that NN methods are effective predictors of agronomic targets, such as gross primary production (Peng et al., 2019) and yield (Barbosa dos Santos et al., 2021), studies that guide NNs with crop models are rare. One recent study has shown that machine-learning-based yield prediction can be improved with guidance from existing crop models (Shahhosseini et al., 2021). The authors of this study found that presenting a variety of machine-learning methods with outputs from Agricultural Systems Production Simulator decreased the error of in-season yield prediction for corn in the US Midwest by up to 20%.

Crop model guidance may be particularly beneficial in NN-based nRT CPE, given that physics-guided machine learning is most effective in problems with limited training data (e.g., Jia et al., 2020). In these instances, biophysical models that simulate growth and development processes are able to provide intermediate variables and limit the complexity of the functions that NNs are required to learn in order to achieve the same performance. Though crop models require calibration and planting and harvest dates to produce accurate in-season CPE on their own, their outputs may provide additional guidance useful in NN-based estimation of crop phenology. Using a broad range of default cultivars, crop model simulations may provide in-season guidance on correlations between crop phenology, water stress, and weather variables within season, thus removing the onus on NN-based CPE of learning these relationships directly. In addition, because uncalibrated crop model cultivars cannot be overfit to in-sample (training) data, crop model guidance may play a moderating role in NN-based CPE during growing seasons of abnormal crop phenological progress. Given the increased likelihood of extreme weather

events affecting global food supply (Mehrabi & Ramankutty, 2021), developing nRT CPE methods that are more robust during abnormal crop phenology patterns is important. The goal of this study is to address this knowledge gap and develop an NN-based framework for nRT CPE that incorporates information from RS, crop model, and other available data. The objective of this study is to develop a method for integrating information from biophysical crop model simulations into an NN method for improved nRT CPE. Specifically, the method is implemented and evaluated for corn, one of the world's major commodity crops, grown in the Corn Belt the United States.

2 | METHODS

2.1 | Study area and datasets

This study was conducted for corn fields in Iowa and Illinois in the Corn Belt of the United States over the period 2003 to 2017. In 2017, these states, which both consist of nine USDA-defined agricultural statistical districts, hereafter districts (see Figure 1), accounted for 27% of all corn acres planted in the United States (USDA-NASS, 2021c). Corn planting in the region typically spans April and May (week of year [WOY] 15–24), silking begins around the beginning of July (WOY 27), and harvest takes place between early September and late November (WOY 36–50). Weekly USDA crop progress reports (CPRs) are produced from ground surveys, which track the progress of corn throughout the season. Such CPRs track USDA-defined planted, emerged, silking, dough, dent, mature, and harvested stages. This study used CPR data and commentary from agricultural statistics documents for Iowa (USDA-NASS, 2005b, 2010b, 2013b, 2015b) and Illinois (USDA-NASS, 2005a, 2010a, 2013a, 2015a), and the USDA Weekly Weather and Crop Bulletin (U.S. Department of Commerce, NOAA, & USDA, 2021).

2.2 | Remote sensing, weather data, and soil properties

Table 1 shows the resolutions and sources of the data used in this study. Corn field locations in the study region were identified using the corn–soy data layer (CSDL) (Wang et al., 2020) (2003–2008) and the USDA Cropland Data Layer (CDL) (USDA-NASS, 2021b) (2008–2017). An example subset of the USDA CDL for 2017 is shown in Supplemental Figure S1. Daily weather data and photoperiod from DayMet were extracted for the centroid of each field (Thornton et al., 2020). The fraction of photosynthetically active radiation (FPAR) product from the NASA Moderate Resolution Imaging Spectroradiometer (MODIS) two-satellite constellation

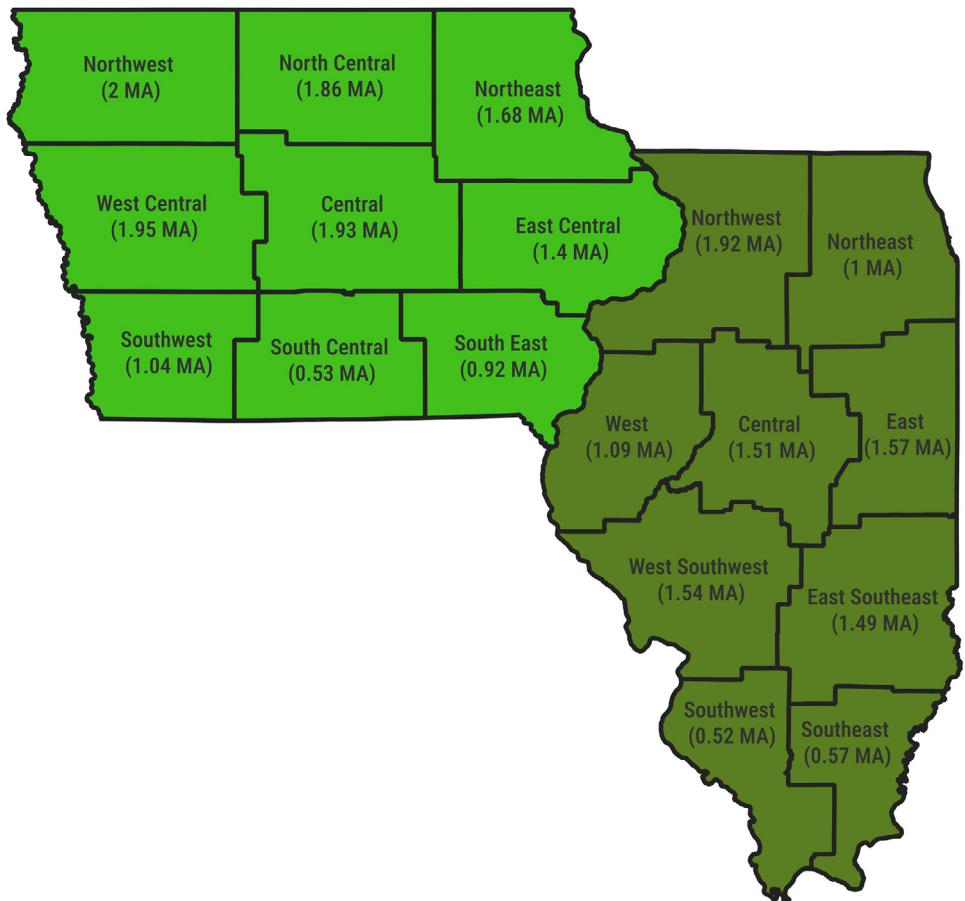


FIGURE 1 Planted area of corn for each agricultural statistical district in the U.S. states of Iowa and Illinois for 2017 in million acres (MA).

was used to provide field-scale remote sensing information (Myndeni et al., 2015). The 250-m FPAR product is derived from leaf canopy reflectances captured by MODIS satellite optical sensors. To ensure fields were large enough to encompass a 250-m MODIS FPAR pixel, corn fields of area $<0.81 \text{ km}^2$ or with a field area-to-bounding rectangle area ratio of <0.9 were discarded following the method outlined in Worrall et al. (2021). The FPAR data were filtered and smoothed to daily values using a modified version of the Savitsky–Golay filtering method outlined in Sakamoto et al. (2011). To recreate the effects of in-season crop phenology monitoring, FPAR values were masked beyond each CPR release date before filtering to emulate nRT data availability. Soil information was acquired from a precomputed DSSAT soil database derived from HarvestChoice soils and SoilGrids data (Han et al., 2019), which provides soil texture and hydraulic properties to a depth of 2 m on a global 10-km resolution grid. Soil data was saved for each 10-km grid square within a district that contained corn.

This study was conducted at district level. Weather data, including air temperature, solar radiation, and rainfall, were aggregated by calculating the mean and standard deviation over each district. Assuming a log-normal distribution for

rainfall rate (Kedem & Chiu, 1987), rainfall data were converted to log scale by applying a $\log(x + 1)$ transform before aggregation. The FPAR data was binned to form histogram distributions of FPAR values across each district. Soil texture and hydraulic properties were aggregated to district level by calculating the means of these properties for each 10-km grid square within a district. Supplemental Figure S2 shows the spatial distribution of sand, silt, and clay contents at 5-cm depth for these grid squares in Iowa. To recreate in-season CPE data availability, year-long data was cropped to variable lengths using CPR release days for WOY 13–51 of each year in the study. Zero values were added to the start of these variable length time series to form fixed-length input tensors.

2.3 | Target phenological stages

In this study, we aimed to estimate the timing of five different crop phenological stages: emerged, silking, grainfill, mature, and harvested, with a sixth stage, pre-emergence, used during estimation to represent the preplanting and planted period. To generate crop phenology distributions, USDA CPRs were

TABLE 1 Remotely sensed fraction of photosynthetically-active radiation (FPAR), meteorological, and soil data used as inputs to the neural network methods.

Parameter	Source	Wavelength	Resolution		Units	Input	Standardization	Reference
			Spatial	Temporal				
Meteorological								
Solar Radiation	ORNL DayMet	—	1 km	Daily	W m^{-2}	MJ m^{-2}	Min.–max.	Thornton et al. (2020)
Photoperiod				h	h	h	Min.–max.	
Temperature				°C	°C	GDD	Min.–max.	
Rainfall				mm d^{-1}	mm d^{-1}	Lognormal-Gaussian transform	min.–max.	
Remote Sensing								
FPAR	NASA MODIS Aqua/Terra	VIS/NIR	250 m+	4 d	—	—	Density histogram	Myneni et al. (2015)
Soil								
5 arc min. gridded soils	Soilgrids/Harvest Choice	—	5 arc min (10 km)	Constant	—	—	None	Han et al. (2019)

TABLE 2 Comparison of the timing of corn phenological stages in this study with those defined by the Decision Support System for Agrotechnology Transfer (DSSAT) CERES-IXIM-Maize crop model and the USDA.

Study crop phenology stage	USDA crop progress stage	DSSAT crop growth stage
Pre-emergence	—	Preplanting
—	Planted	Planted
Emerged	Emerged	Emerged
—	—	Late juvenile
—	—	Tasseling
Silking	Silking	Silking
—	—	Grainfill
Grainfill	Dough	—
—	Dent	—
Mature	Mature	Mature
Harvested	Harvested	Harvested

converted from cumulative to fractional crop phenology for each stage, with the fraction of a district's crop across all stages summing to one for any given week. Table 2 provides a comparison of the stage definitions for this study—those defined by the USDA, and those defined by the crop model used in this study for NN guidance.

2.4 | Crop model

Crop model simulations for this study were conducted using the CSM-IXIM maize crop model within DSSAT, a leading crop modeling suite (Lizaso et al., 2011; Jones et al., 2003). These simulations provided biophysically based crop growth and development for the study. The DSSAT inputs include daily rainfall, solar radiation, minimum and maximum air temperature at 2 m, as well as cultivar parameters that define characteristics such as growth rate and leaf development. Eight generic DSSAT corn cultivars with varying rates of growth stage progression were used (Table 3). These cultivars are provided with the DSSAT installation and were not calibrated for the study period.

Planting and harvest dates for each grid were randomly sampled from the in-sample historical planting and harvest windows of each district. Table 4 shows which years formed the composite historical planting and harvest windows during validation and evaluation. The DSSAT automatic planting management functionality was set to plant on first suitable day between the sampled date and the last day in the historical planting window. The criteria for a suitable planting day were average soil moisture to a depth of 30 cm $\geq 40\%$ of the drained upper limit–wilting point range, and surface

TABLE 3 Decision Support System for Agrotechnology Transfer CERES-IXIM maize cultivars used during crop model data generation.

Cultivar code	P1	P2	P5	G2	G3	PHINT
PC0001	160	0.750	780	750.0	8.50	49.00
PC0002	185	0.750	850	800.0	8.50	49.00
PC0003	212	0.750	850	800.0	8.50	49.00
PC0004	240	0.750	850	800.0	8.50	49.00
PC0005	260	0.750	850	800.0	8.50	49.00
990001	320	0.520	940	620.0	6.00	38.90
990002	200	0.300	800	700.0	8.50	38.90
990003	110	0.300	680	820.4	6.60	38.90

Note: Parameters are as follows: P1, growing degree days (GDD) from emergence to end of juvenile phase; P2, extent of developmental delay (in days) as a function of photoperiod above 12.5 h for a single day; P5, GDD from silking to plant maturity; G2, maximum number of kernels of topmost ear; G3, kernel filling rate under optimum conditions (mg d^{-1}); PHINT, interval in GDD between successive leaf appearances.

soil temperature between 10 and 40 °C. Simulation harvest date was set to the harvest date randomly sampled from the historical distribution.

In this study, we used DSSAT outputs related to growth stage (GSTD) as well as water stress factors related to growth (WSGD), photosynthesis (WSPD), and excess water (EWSD), hereafter referred to collectively as water stress factors (WSFs). Table 5 shows the DSSAT codes and valid ranges for these model outputs. Within CSM-IXIM, WSFs are used to calculate yield loss and plant tissue death because of soil water deficit or excess. The GSTD progression within the model is unaffected by WSFs, with crop growth stage progression driven purely by accumulation of GDD, which are calculated using base 8 °C and maximum 34 °C temperature.

2.5 | Long short-term memory-based networks

The group of NN structures used in this study use long short-term memory (LSTM) (Graves, 2012), a method that has found success in time-series classification and natural language processing applications. Long short-term memory is capable of handling time-series data with variable length gaps between key events. We used LSTM to encode in-season weather and FPAR data, which were then presented to a fully connected (dense) network to produce nRT CPE. Long short-term memory encoder-fully connected NN architectures have been used in both crop mapping (Xu et al., 2020) and CPE (Worrall et al., 2021). In this study, the NN structures built upon our previous research, which found that NNs that separate crop growth drivers in a branched structure outperform sequential structures in NN-based CPE (Worrall et al., 2021). As shown in Figure 2, the NN structures included three

branches. The first branch was presented with solar radiation and photoperiod inputs, the second branch was presented with FPAR and GDD inputs, while the third branch received rainfall and soil properties information. In addition, coordinates of the centroid of each district were passed to the first dense layer to provide spatial context for each time series. A softmax output layer was used to enforce the constraint that the sum of the estimates of the fraction of the crop in each stage on any given week must sum to one.

2.6 | Experimental setup

2.6.1 | Crop model simulations

Simulations were conducted on 10-km grid squares across the study region, matching the spatial resolution of the soil data from Han et al. (2019). Only grid squares containing corn fields with area 0.27 km² or greater were used. These simulations were conducted using the preformatted DSSAT soil grid files from Han et al. (2019) and DayMet daily weather data spatially averaged over each 10-km grid. For each grid in a given year, eight DSSAT simulations were conducted, one for each cultivar, resulting in eight GSTD outputs and eight sets of three WSFs. Simulation outputs were combined to district level for each cultivar. Mean and standard deviation values for each WSF–cultivar combination were calculated, resulting in three mean and three standard deviation values per cultivar for a total of 48 WSF time series. The GSTD simulation outputs were combined to district-level binned distributions by forming histograms of the GSTD values for each cultivar across all grids in a district for each day of a growing season. Because crop growth stage information is given by GSTD, a ridge regression was conducted over these outputs to obtain in-season crop phenology estimates from DSSAT. Figure 3 is a flow diagram of the gridded DSSAT simulation to ridge-regression-based CPE. These in-season crop phenology estimates established a performance baseline for comparison with the NN approaches used in this study. Because simulation planting and harvest dates are sampled from in-sample USDA CPR planting and harvest distributions, accuracy of the DSSAT-only ridge regression (DSSAT-RR) on pre-emergence and harvest is correlated with the in-sample average for these stages.

2.6.2 | Model-guided LSTM

To investigate how best to integrate crop model outputs, an unguided NN method (LSTM_{UG}; Figure 2a) was evaluated alongside three different methods for using crop model simulation outputs. The three model-guided methods were as follows: (a) a method that used binned time series GSTD

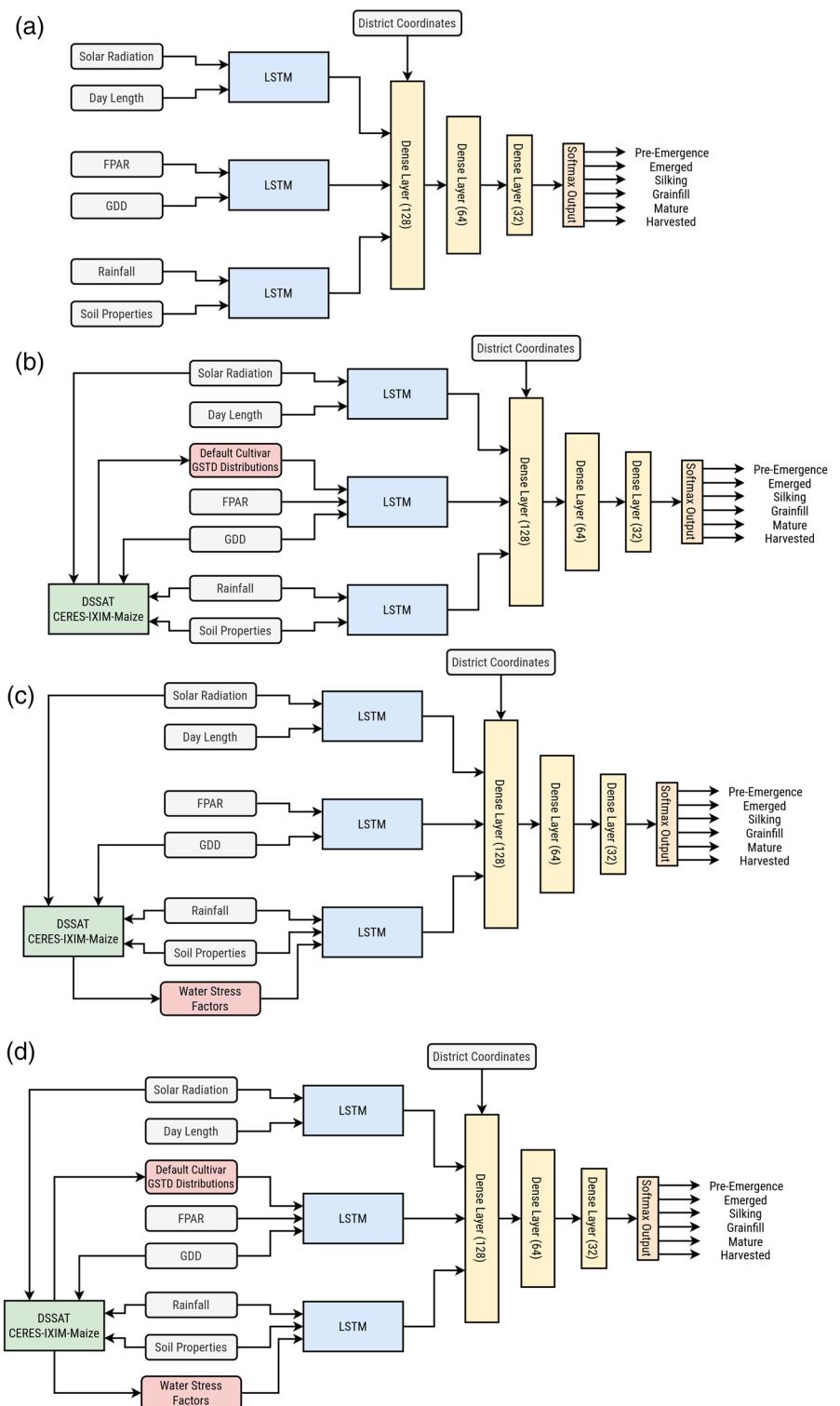
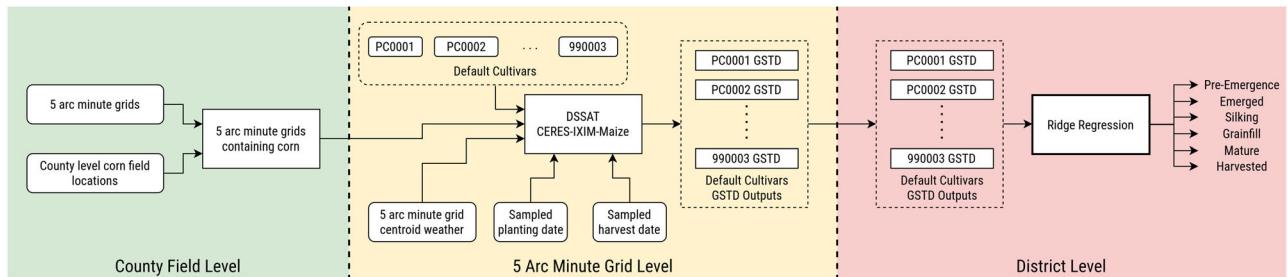


FIGURE 2 Neural network (NN) methods used in this study. (a) Unguided NN (long short-term memory [LSTM]). (b) Neural network with crop growth stage (GSTD) distribution time series generated by Decision Support System for Agrotechnology Transfer (DSSAT) (LSTM_{GSTD}). (c) Neural network with water stress factor (WSF) time series distributions generated by DSSAT (LSTM_{WSF}). (d) Neural network with crop growth stage distribution and water stress factor time series generated by DSSAT (LSTM_{WSF+GSTD}). All NN methods use LSTM and receive masked, in-season time series inputs that are front-padded to fixed length tensors. FPAR, fraction of photosynthetically-active radiation; GDD, growing degree days.

TABLE 4 Separation of years into in and out-of-sample and early stopping years during model validation and evaluation.

Data split	In-sample years	Early stopping year	Out-of-sample years
Cross-validation Fold 1	2005, 2006, 2007, 2008, 2010, 2011, 2016, 2017	2013	2003, 2015
Cross-validation Fold 2	2003, 2008, 2010, 2011, 2013, 2015, 2016	2007	2005, 2006, 2017
Cross-validation Fold 3	2005, 2006, 2008, 2010, 2013, 2015, 2016, 2017	2003	2007, 2011
Cross-validation Fold 4	2003, 2005, 2007, 2011, 2013, 2015, 2016, 2017	2006	2008, 2010
Cross-validation Fold 5	2003, 2005, 2006, 2007, 2010, 2011, 2015, 2017	2008	2013, 2016
Model evaluation	2003, 2005, 2006, 2007, 2008, 2010, 2011, 2013, 2015, 2017	2016	2004, 2009, 2012, 2014

**FIGURE 3** Method for producing crop phenology estimates using generic Decision Support System for Agrotechnology Transfer (DSSAT) CSM-IXIM maize cultivars. Weights for each cultivar and stage are computed via ridge regression. Planting and harvest dates are sampled from historical crop progress reports distributions. GSTD, growth stage.**TABLE 5** Decision Support System for Agrotechnology Transfer (DSSAT) CSM-IXIM maize model outputs used in this study.

Model output	DSSAT code	Scale
Crop growth stage	GSTD	1–6
Water stress factor – photosynthesis	WSPD	0–1
Water stress factor – growth	WSGD	0–1
Excess water stress factor	EWSD	0–1

Abbreviations: EWSD, excess water stress; GSTD, growth stage; WSGD, water stress factor related to growth; WSPD, water stress factor related to photosynthesis.

distributions from DSSAT for each of the eight used cultivars ($LSTM_{GSTD}$; Figure 2b), (b) a method that used district-level mean and standard deviation time series of the three WSFs from DSSAT for each of the eight used cultivars ($LSTM_{WSF}$; Figure 2c), and (c) a method that used both GSTD distribution and WSF time series ($LSTM_{WSF+GSTD}$; Figure 2d). The first two methods were used to assess the value of DSSAT GSTD and WSF outputs independently, and the third method was used to determine the impact of complementarities in the DSSAT GSTD and WSFs on NN performance.

2.6.3 | Validation and evaluation years

Four years—2004, 2009, 2012, and 2014—were selected from the study period (2003–2017) as evaluation years and

kept separate during method development. They were then used to evaluate methods once NN hyperparameters and preferred DSSAT simulation output combinations had been established. The year 2004 was randomly sampled from the study period years, while years 2009, 2012, and 2014 were chosen to represent a broad range of crop phenological scenarios. Table 6 shows the deviations of crop phenological progress from the average at 50% complete during the study period for both Iowa and Illinois. During 2004, planting in both Iowa and Illinois occurred earlier than the study period average. Cooler temperatures during the summer months in Iowa led to slower crop maturity and harvest, while in Illinois, corn progressed close to the study period average because of optimal growing conditions. In 2009, planting progressed normally in Iowa but was delayed in Illinois by heavy rain. Both states experienced slow phenological progress that year because of cooler-than-normal temperatures, and harvest was significantly delayed by heavy rainfall in October. During 2012, warm conditions in May helped planting to proceed quickly, and phenological progress was further sped up by hot and dry conditions during the summer months, leading to earlier-than-normal maturity and harvest. In 2014, planting was delayed in both states because of low soil temperatures and late frost, but warmer conditions in May and June brought phenological progress in line with the study period average. Cooler temperatures in July and August once again slowed progress and the season ended with a delayed harvest. Figures 4 and 5 show the phenological progress of corn in

TABLE 6 Deviation of evaluation year phenological stage timing to 50% complete vs. the study period average for Iowa (IA) and Illinois (IL).

Year	Planted		Emerged		Silking		Grainfill		Mature		Harvested	
	IA	IL	IA	IL	IA	IL	IA	IL	IA	IL	IA	IL
d												
2004	-2	-7	-3	-6	5	-4	8	-4	9	3	7	0
2009	-3	24	1	19	7	16	13	17	12	27	24	39
2012	3	-6	0	-5	-4	-4	-5	-1	-9	-10	-18	-11
2014	7	6	5	3	1	0	-3	0	9	6	9	15

Note. Positive values indicate a delay.

Iowa and Illinois for the evaluation years and the study period average progress.

2.6.4 | Network loss, cross-validation, and evaluation

Neural network method hyper-parameters were selected using fivefold cross-validation over the in-sample years, with fold year splits outlined in Table 4. Each LSTM module in the NN structures contained 32 hidden units. A learning rate of 1×10^{-5} , a batch size of 64, and a dropout rate between layers was set to 0.2. During training, early stopping with weight restoration was used to prevent NN overfitting, with 1 yr from each training set kept separate and assessed after each epoch. An initial 5,000 epochs was used, with early stopping tolerance of 500 epochs without improvement. Neural networks in this study used Kullback–Leibler divergence (D_{KL}) as a loss function, which is a measure of the difference between two probability distributions. The D_{KL} term is common in regression problems with targets that are distributions that sum to one. For two distributions, $P(x)$ and $Q(x)$, D_{KL} is calculated as follows:

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) \log \left(\frac{P(x)}{Q(x)} \right)$$

where D_{KL} is used in this study as a measure between estimated and observed crop phenology across all stages for a given week. Loss was evaluated at the district level.

2.7 | Metrics

Three metrics were used to assess the performance of CPE methods. The first metric, Nash–Sutcliffe efficiency (NSE), was used in this study as a metric for how well an NN describes the time series values of a given phenological stage;

NSE is defined as follows:

$$NSE = 1 - \frac{\sum_{t=1}^T (Q_m^t - Q_o^t)^2}{\sum_{t=1}^T (Q_o^t - \bar{Q}_o)^2}$$

where Q_m^t is the model estimate at time t , Q_o^t is the observed value at time t , and \bar{Q}_o is the mean value of the observed time series. The NSE values range from $-\infty$ to a maximum of 1, where 1 means the model perfectly describes the observed time series. An NSE value less than zero implies the model is worse at describing the observed time series than the mean of the observed time series. The NSE was computed over the time series of the fraction of corn in a district currently in the target stage over an entire growing season. The NSE values are thus affected by a method's ability to estimate timing of crop progression both into and out of a phenological stage.

The second metric is cosine similarity (CS), which is a measure of the similarity of a method's crop phenology estimate across all stages for a given week to those reported in the CPRs. Cosine similarity is a measure of the difference between two vectors. For two vectors, \mathbf{A} and \mathbf{B} , CS is calculated as follows:

$$CS = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

In this study, CS was used to provide information on a method's ability to estimate fractional crop phenology across all stages for a given WOY.

In addition to the above two metrics, we also assessed each method on RMSE of the median date (date to 50%) of cumulative progression into a phenological stage. This metric is calculated as follows:

$$RMSE_{stage} = \sqrt{\frac{\sum_{i=1}^N (date_{est}^{stage} - date_{obs}^{stage})^2}{N}}$$

Where $RMSE_{stage}$ is the RMSE for a phenological stage expressed in days, and N is the total number of district–year combinations in the evaluation years. While this metric only

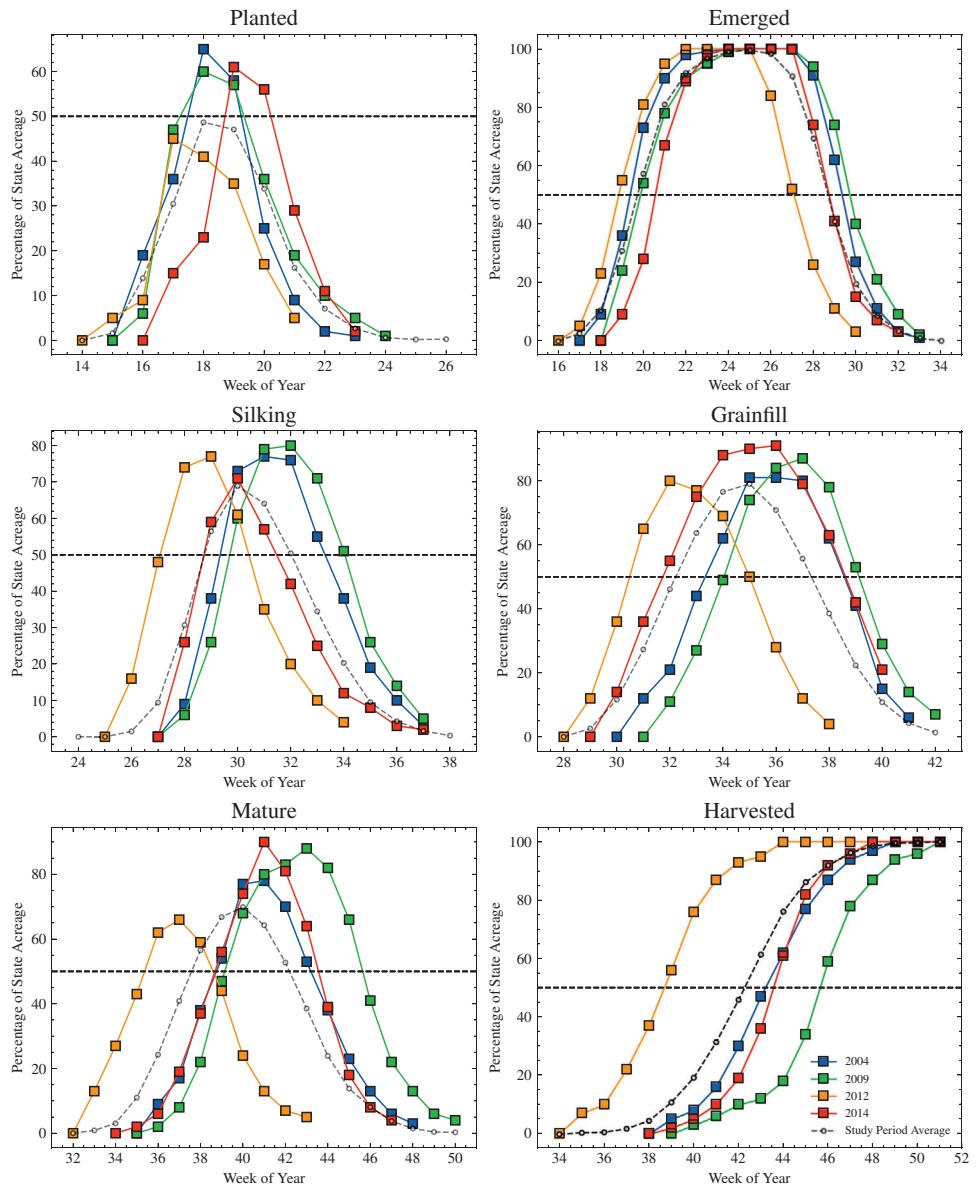


FIGURE 4 Progress of corn in Iowa through each of the study-defined phenological stages during the evaluation years 2004, 2009, 2012, and 2014.

describes performance at the median point in each stage, it is commonly used in other crop phenology studies. The RMSE was used in this study to compare methods with full-season CPE methods from three recent CPE studies: Seo et al (2019), with results reported for districts in Iowa and Illinois 2003 to 2015 and Diao (2020) and Diao et al. (2021), with results reported for districts in Illinois from 2002 to 2017. We refer the reader to Diao et al. (2021) for graphical representations of their results that were not reported in tabulated form.

The CPE methods in this study were tasked with estimating fractional crop phenology of corn for a given district on the CPR release date, d , with the estimates for all stages on day d summing to one. Graphical representations of fractional crop phenology show state-level estimates to aid visual comparison of the differences between Iowa and Illinois phenological

progress. State-level estimates of crop phenology were produced by weighting district-level estimates of each method by the number of acres planted in each district for a given year (USDA-NASS, 2021c). These graphical representations differ from USDA CPR data, which are reported as cumulative progress into a phenological stage.

3 | RESULTS AND DISCUSSION

3.1 | Overall performance

Tables 7, 8, and 9 show the NSE of the three NN-based and the DSSAT-RR methods across the four evaluation years for the combined Iowa and Illinois districts, respectively.

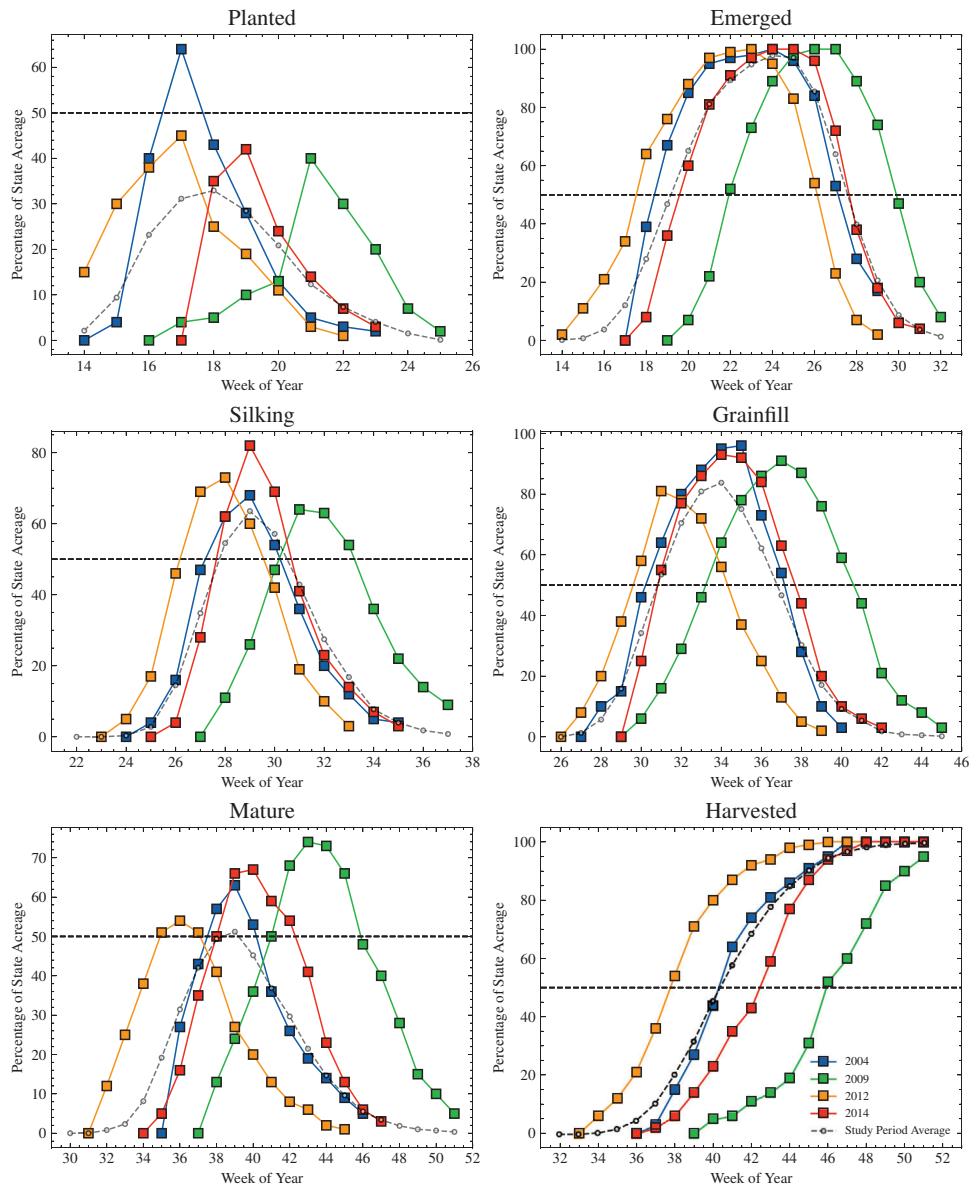


FIGURE 5 Progress of corn in Illinois through each of the study-defined phenological stages during the evaluation years 2004, 2009, 2012, and 2014.

Figure 6 shows the CS box and whisker plots for each method across all districts for the evaluation years. The best method for each phenological stage was a model-guided NN, with the $\text{LSTM}_{\text{GSTD}}$ and $\text{LSTM}_{\text{WSF+GSTD}}$ methods producing better NSE than the unguided method for all stages. The $\text{LSTM}_{\text{WSF+GSTD}}$ method produced the best NSE for all stages up to grainfill, with the most pronounced differences in performance on the silking and grainfill stages when corn is most vulnerable to environmental stressors. Cosine similarity ranges for the $\text{LSTM}_{\text{WSF+GSTD}}$ method in Figure 6 are also much smaller during the middle of the season (WOY 25–40) than for all other methods. Accuracy of NN methods was lowest for the silking and mature stages, while the DSSAT-RR method was worse for later phenological stages. The DSSAT-

RR method CS across all stages was significantly lower later in the season as shown in Figure 6.

3.2 | GSTD guidance

The GSTD outputs from DSSAT were found to improve in-season CPE when compared with the unguided LSTM_{UG} method. For the 2004 Iowa season, the $\text{LSTM}_{\text{GSTD}}$ produced similar NSE for the mature stage, both noticeably higher than the unguided and DSSAT methods, as shown in Table 10. The LSTM_{UG} estimates were increasingly early with respect to the observed data as phenological progression through grainfill and maturity and was slowed by the cooler

TABLE 7 Median (Med), mean, and standard deviation (Std) of Nash–Sutcliffe efficiency (NSE) values across all districts for the evaluation years.

Model	Pre-emergence			Emerged			Silking			Grainfill			Mature			Harvested		
	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std
DSSAT-RR	0.969	0.932	0.091	0.957	0.909	0.130	0.882	0.806	0.184	0.834	0.809	0.132	0.495	0.452	0.319	0.948	0.792	0.345
Unguided	0.972	0.954	0.067	0.954	0.938	0.068	0.826	0.803	0.142	0.875	0.820	0.149	0.848	0.665	0.610	0.961	0.920	0.123
GSTD	0.972	0.948	0.061	0.958	0.929	0.089	0.879	0.841	0.127	0.905	0.898	0.069	0.882	0.793	0.212	0.975	0.931	0.101
WSF	0.959	0.950	0.044	0.938	0.936	0.055	0.886	0.845	0.123	0.905	0.872	0.095	0.829	0.802	0.207	0.976	0.938	0.087
WSF+GSTD	0.984	0.961	0.045	0.974	0.948	0.069	0.930	0.897	0.088	0.942	0.911	0.073	0.854	0.805	0.173	0.968	0.930	0.087

Note. DSSAT-RR, Decision Support System for Agrotechnology Transfer -only ridge regression; GSTD, crop growth stage; WSF, water stress factor. Bold values indicate the method with the highest median NSE for each stage.

TABLE 8 Median (Med), mean, and standard deviation (Std) of Nash–Sutcliffe efficiency (NSE) values for all methods on Iowa districts for the evaluation years.

Model	Pre-emergence			Emerged			Silking			Grainfill			Mature			Harvested		
	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std
DSSAT-RR	0.979	0.977	0.013	0.970	0.966	0.018	0.911	0.863	0.103	0.863	0.833	0.110	0.576	0.508	0.248	0.944	0.853	0.177
Unguided	0.984	0.979	0.015	0.966	0.961	0.026	0.867	0.821	0.115	0.881	0.830	0.126	0.857	0.793	0.159	0.961	0.931	0.102
GSTD	0.986	0.980	0.017	0.971	0.965	0.022	0.886	0.855	0.100	0.892	0.883	0.074	0.906	0.825	0.142	0.975	0.936	0.087
WSF	0.952	0.951	0.045	0.936	0.941	0.042	0.876	0.819	0.133	0.877	0.851	0.101	0.828	0.842	0.097	0.965	0.936	0.081
WSF+GSTD	0.991	0.977	0.033	0.980	0.971	0.027	0.931	0.914	0.067	0.941	0.903	0.084	0.890	0.815	0.190	0.966	0.933	0.066

Note. DSSAT-RR, Decision Support System for Agrotechnology Transfer -only ridge regression; GSTD, crop growth stage; WSF, water stress factor. Bold values indicate the method with the highest median NSE for each stage.

TABLE 9 Median (Med), mean, and standard deviation (Std) of Nash–Sutcliffe efficiency (NSE) values for all methods on Illinois districts for the evaluation years.

Model	Pre-emergence			Emerged			Silking			Grainfill			Mature			Harvested		
	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std
DSSAT-RR	0.932	0.888	0.111	0.920	0.853	0.165	0.832	0.750	0.225	0.825	0.786	0.147	0.489	0.395	0.368	0.948	0.731	0.447
Unguided	0.961	0.928	0.086	0.938	0.915	0.087	0.806	0.784	0.162	0.874	0.811	0.169	0.821	0.537	0.829	0.966	0.908	0.139
GSTD	0.944	0.916	0.071	0.947	0.893	0.113	0.863	0.826	0.147	0.939	0.913	0.059	0.854	0.760	0.259	0.972	0.926	0.112
WSF	0.959	0.949	0.044	0.941	0.931	0.066	0.908	0.872	0.106	0.918	0.892	0.084	0.840	0.763	0.270	0.985	0.940	0.092
WSF+GSTD	0.962	0.944	0.050	0.961	0.925	0.088	0.928	0.880	0.102	0.942	0.920	0.057	0.832	0.794	0.155	0.970	0.927	0.104

Note. DSSAT-RR, Decision Support System for Agrotechnology Transfer -only ridge regression; GSTD, crop growth stage; WSF, water stress factor. Bold values indicate the method with the highest median NSE for each stage.

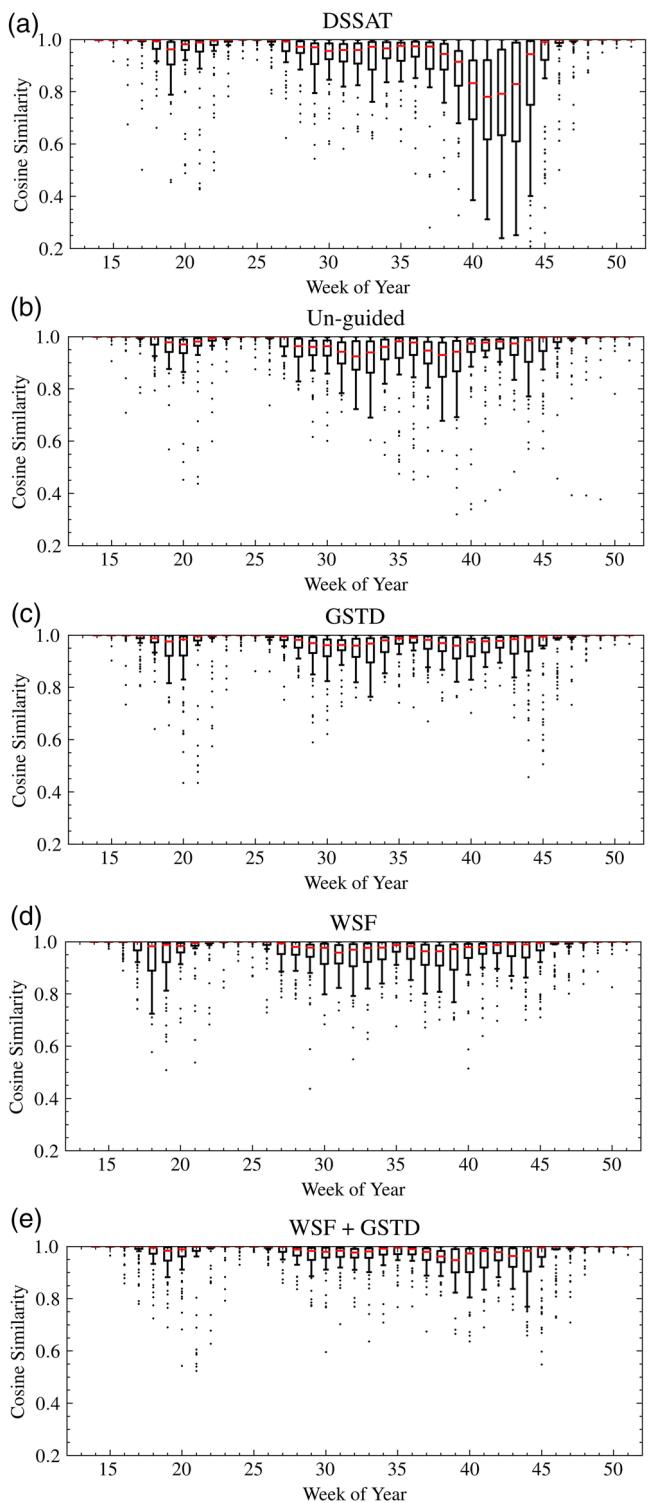


FIGURE 6 Week-to-week cosine similarity between the estimated and observed crop phenology across all stages during the evaluation years for the each method. DSSAT, Decision Support System for Agrotechnology Transfer; GSTD, growth stage; WSF, water stress factor.

TABLE 10 Median (Med), mean, and standard deviation (Std) of Nash-Sutcliffe efficiency (NSE) values for all methods for Iowa districts.

Model	Pre-emergence			Emerged			Silking			Grainfill			Mature			Harvested			
	Med		Mean	Med		Mean	Med		Mean	Med		Mean	Med		Mean	Med		Mean	Std
	2004	2009	2012	2014															
DSSAT-RR	0.967	0.965	0.012	0.958	0.010	0.903	0.883	0.057	0.745	0.707	0.107	0.574	0.499	0.241	0.979	0.982	0.982	0.011	
Unguided	0.984	0.976	0.014	0.975	0.016	0.910	0.917	0.029	0.921	0.922	0.041	0.889	0.825	0.181	0.967	0.913	0.913	0.162	
GSTD	0.991	0.986	0.011	0.977	0.961	0.027	0.926	0.880	0.098	0.948	0.906	0.099	0.938	0.923	0.052	0.986	0.982	0.982	0.015
WSF	0.917	0.916	0.021	0.912	0.903	0.037	0.758	0.741	0.122	0.730	0.742	0.106	0.795	0.780	0.072	0.949	0.943	0.943	0.038
WSF+GSTD	0.996	0.993	0.006	0.985	0.978	0.025	0.933	0.913	0.058	0.954	0.926	0.081	0.941	0.937	0.067	0.993	0.975	0.975	0.040
DSSAT-RR	0.980	0.977	0.010	0.962	0.953	0.022	0.879	0.851	0.090	0.862	0.855	0.053	0.403	0.405	0.111	0.594	0.569	0.569	0.112
Unguided	0.991	0.984	0.014	0.985	0.973	0.023	0.845	0.839	0.054	0.809	0.803	0.068	0.831	0.796	0.091	0.935	0.877	0.877	0.094
GSTD	0.986	0.980	0.016	0.973	0.967	0.018	0.881	0.887	0.045	0.867	0.871	0.061	0.757	0.747	0.109	0.800	0.811	0.811	0.090
WSF	0.994	0.973	0.054	0.990	0.987	0.010	0.900	0.945	0.045	0.877	0.878	0.058	0.827	0.835	0.069	0.869	0.855	0.855	0.119
WSF+GSTD	0.995	0.992	0.006	0.983	0.985	0.008	0.878	0.872	0.074	0.832	0.831	0.093	0.813	0.802	0.058	0.854	0.853	0.853	0.062
DSSAT-RR	0.981	0.977	0.012	0.981	0.974	0.013	0.929	0.938	0.022	0.942	0.933	0.032	0.695	0.540	0.364	0.902	0.907	0.907	0.048
Unguided	0.972	0.971	0.014	0.959	0.957	0.013	0.728	0.767	0.114	0.713	0.683	0.137	0.601	0.621	0.097	0.960	0.956	0.956	0.017
GSTD	0.987	0.984	0.013	0.979	0.976	0.013	0.891	0.885	0.052	0.860	0.861	0.062	0.709	0.687	0.132	0.966	0.961	0.961	0.015
WSF	0.941	0.927	0.030	0.925	0.919	0.016	0.886	0.855	0.091	0.868	0.862	0.080	0.823	0.807	0.095	0.962	0.961	0.961	0.026
WSF+GSTD	0.931	0.932	0.037	0.950	0.941	0.028	0.943	0.946	0.035	0.867	0.882	0.047	0.618	0.577	0.216	0.921	0.922	0.922	0.032
DSSAT-RR	0.988	0.988	0.004	0.977	0.978	0.006	0.748	0.778	0.131	0.833	0.835	0.082	0.638	0.585	0.163	0.954	0.952	0.952	0.014
Unguided	0.991	0.986	0.009	0.934	0.942	0.032	0.732	0.762	0.144	0.898	0.912	0.031	0.945	0.929	0.047	0.988	0.979	0.979	0.020
GSTD	0.979	0.971	0.023	0.963	0.954	0.023	0.737	0.770	0.127	0.887	0.895	0.058	0.937	0.943	0.021	0.993	0.991	0.991	0.007
WSF	0.991	0.991	0.004	0.967	0.955	0.033	0.824	0.773	0.170	0.935	0.923	0.041	0.955	0.945	0.053	0.993	0.983	0.983	0.029
WSF+GSTD	0.997	0.993	0.009	0.983	0.980	0.015	0.943	0.923	0.072	0.975	0.972	0.012	0.958	0.946	0.030	0.985	0.983	0.983	0.011

Note. DSSAT-RR, Decision Support System for Agrotechnology Transfer -only ridge regression; GSTD, crop growth stage; WSF, water stress factor. Bold values indicate the method with the highest median NSE for each stage.

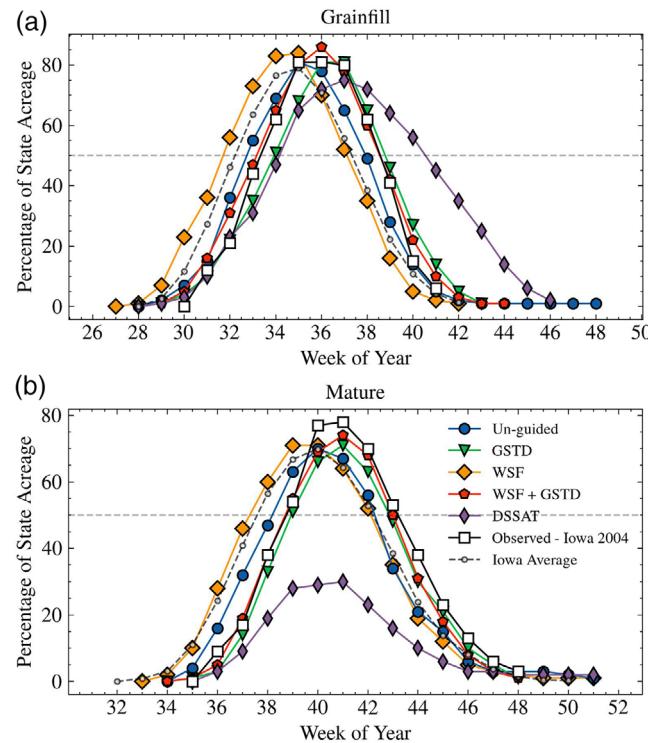


FIGURE 7 Estimated vs. observed phenological progress of corn through the (a) grainfill and (b) mature stages in Iowa during 2004. DSSAT, Decision Support System for Agrotechnology Transfer; GSTD, growth stage; WSF, water stress factor.

mid-summer weather (see Figures 7a,b). The $LSTM_{GSTD}$ method performance was higher on these stages, GSTD outputs from DSSAT simulations improve CPE during the height of the season, when satellite-observable change in canopy appearance is at a minimum. While GSTD-based DSSAT-RR method accuracy was reduced, the $LSTM_{GSTD}$ method was able to use GSTD information in conjunction with canopy appearance and weather inputs for improved CP estimates.

Although the $LSTM_{GSTD}$ method performed well for Iowa in 2004, the method had lower NSE values versus the other guided methods during 2004 season in Illinois, as shown in Table 11. The $LSTM_{GSTD}$ estimates were delayed for silking, grainfill, and mature, as seen in Figures 8a, 8b, and 8c. Planting progressed rapidly during this season, and comparing silking NSE values across all NN methods, it is clear that GSTD outputs from DSSAT were detrimental to silking accuracy during this year, when $LSTM_{GSTD}$ accuracy was worse than the unguided method. Given that the 2004 Illinois season had optimal growing conditions, this suggests that the value of GSTD guidance may be reduced during years without abnormal weather impact on crop phenological progress.

A similar issue arose in Illinois 2009, where delayed planting because of heavy rains appears to have reduced the value of simulation GSTD guidance as seen in Table 11. Simulations were planted during the historical planting window, so

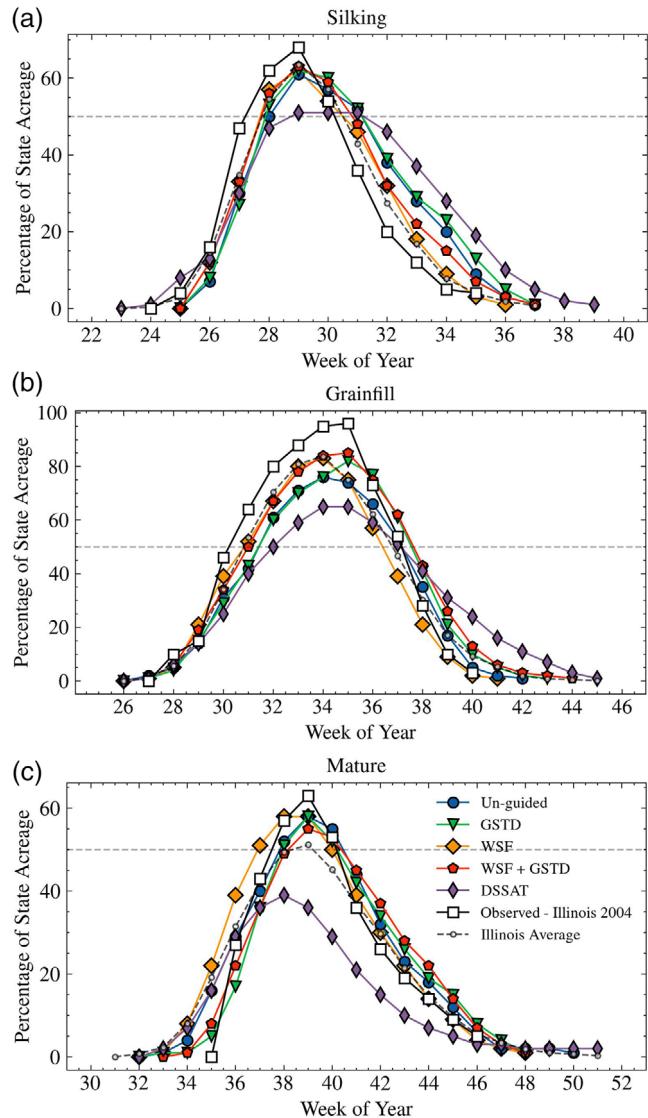


FIGURE 8 Estimated vs. observed phenological progress of corn through the (a) silking, (b) grainfill, and (c) mature stages in Illinois during 2004. DSSAT, Decision Support System for Agrotechnology Transfer; GSTD, growth stage; WSF, water stress factor.

DSSAT GSTD output are tethered to that window. While variation in the rate of GSTD progression mid-season is a useful indicator for model-guided methods when weather is slowing or hastening phenological progression, tethered GSTD information seemingly hinders NN methods for early season estimation in years with planting windows that differ significantly from the study period average. The $LSTM_{GSTD}$ had lower NSE than the unguided method for the pre-emergence and emerged stages in Illinois 2009 as seen in Table 11. The $LSTM_{GSTD}$ method performance for this season in Illinois only increased during the grainfill stage, when cooler temperatures became a significant component of phenological progress in Illinois.

TABLE 11 Median (Med), mean, and standard deviation (Std) of Nash-Sutcliffe efficiency (NSE) values for all methods for Illinois districts.

Model	Pre-emergence				Emerged				Siling				Grainfill				Mature				Harvested							
	Med		Mean		Std		Med		Mean		Std		Med		Mean		Std		Med		Mean		Std					
	2004	DSSAT-RR	0.937	0.928	0.039	0.913	0.907	0.038	0.64	0.661	0.193	0.787	0.785	0.1	0.715	0.67	0.186	0.992	0.989	0.008	0.995	0.995	0.008	0.995	0.995	0.008		
Un-guided	0.968	0.963	0.034	0.957	0.948	0.03	0.83	0.735	0.229	0.917	0.889	0.062	0.872	0.74	0.411	0.916	0.891	0.103	0.988	0.982	0.017	0.966	0.969	0.017	0.995	0.995	0.017	
GSTD	0.966	0.967	0.018	0.951	0.952	0.022	0.774	0.717	0.213	0.898	0.887	0.07	0.916	0.891	0.103	0.988	0.982	0.017	0.995	0.995	0.017	0.966	0.969	0.017	0.995	0.995	0.017	
WSF	0.969	0.961	0.032	0.967	0.958	0.027	0.922	0.867	0.101	0.932	0.919	0.045	0.852	0.836	0.112	0.992	0.989	0.007	0.995	0.995	0.007	0.995	0.995	0.007	0.995	0.995	0.007	
WSF+GSTD	0.985	0.977	0.018	0.968	0.971	0.015	0.887	0.829	0.139	0.914	0.913	0.047	0.892	0.859	0.121	0.981	0.972	0.023	0.995	0.995	0.023	0.995	0.995	0.023	0.995	0.995	0.023	
2009	DSSAT-RR	0.819	0.815	0.065	0.614	0.625	0.186	0.654	0.587	0.297	0.677	0.668	0.204	-0.038	-0.094	0.235	-0.071	0.047	0.414	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011
Un-guided	0.924	0.894	0.077	0.875	0.832	0.129	0.805	0.825	0.101	0.821	0.776	0.148	0.723	0.659	0.198	0.902	0.833	0.133	0.995	0.995	0.011	0.966	0.969	0.011	0.995	0.995	0.011	
GSTD	0.843	0.836	0.039	0.746	0.722	0.096	0.8	0.789	0.116	0.897	0.88	0.055	0.645	0.636	0.113	0.713	0.771	0.131	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011	
WSF	0.959	0.928	0.061	0.936	0.885	0.104	0.925	0.912	0.059	0.863	0.831	0.089	0.774	0.707	0.122	0.862	0.827	0.119	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011	
WSF+GSTD	0.894	0.88	0.044	0.86	0.802	0.098	0.884	0.84	0.105	0.892	0.887	0.044	0.699	0.671	0.131	0.74	0.79	0.128	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011	
2012	DSSAT-RR	0.851	0.834	0.162	0.92	0.917	0.05	0.912	0.903	0.039	0.851	0.828	0.101	0.49	0.457	0.296	0.956	0.962	0.02	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011
Un-guided	0.896	0.884	0.126	0.932	0.922	0.049	0.771	0.708	0.145	0.664	0.647	0.196	0.392	-0.02	1.404	0.934	0.87	0.196	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011	
GSTD	0.95	0.906	0.082	0.954	0.946	0.032	0.899	0.895	0.046	0.958	0.932	0.046	0.798	0.577	0.381	0.968	0.959	0.026	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011	
WSF	0.939	0.939	0.032	0.916	0.919	0.034	0.827	0.796	0.138	0.908	0.864	0.093	0.813	0.631	0.446	0.988	0.965	0.04	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011	
WSF+GSTD	0.938	0.944	0.036	0.957	0.959	0.021	0.932	0.927	0.024	0.951	0.925	0.08	0.785	0.754	0.15	0.97	0.969	0.011	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011	
2014	DSSAT-RR	0.979	0.975	0.011	0.969	0.961	0.017	0.86	0.848	0.089	0.868	0.864	0.057	0.6	0.548	0.149	0.944	0.927	0.046	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011
Un-guided	0.968	0.972	0.013	0.955	0.957	0.018	0.918	0.87	0.069	0.943	0.931	0.029	0.906	0.769	0.382	0.982	0.964	0.053	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011	
GSTD	0.951	0.956	0.032	0.949	0.951	0.037	0.92	0.902	0.037	0.954	0.952	0.007	0.943	0.937	0.059	0.994	0.992	0.006	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011	
WSF	0.978	0.966	0.032	0.974	0.96	0.028	0.911	0.915	0.051	0.969	0.955	0.027	0.951	0.878	0.163	0.987	0.978	0.021	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011	
WSF+GSTD	0.982	0.975	0.017	0.977	0.969	0.023	0.933	0.925	0.047	0.952	0.954	0.014	0.917	0.893	0.103	0.988	0.979	0.024	0.995	0.995	0.011	0.995	0.995	0.011	0.995	0.995	0.011	

Note. DSSAT-RR, Decision Support System for Agrotechnology Transfer -only ridge regression; GSTD, crop growth stage; WSF, water stress factor. Bold values indicate the method with the highest median NSE for each stage.

Mature and harvested NSEs for this season in Illinois were also affected by GSTD information, when the second wettest October led to the second slowest harvest on record. When faced with a very delayed season, all methods struggled to accurately describe crop phenology through maturity and harvest as seen in Figures 9c and 9d. Decreased accuracy in estimating crop phenology during seasons with late planting has also been observed in crop modeling studies (Hussain et al., 2018). While the $\text{LSTM}_{\text{GSTD}}$ could accurately estimate progression into the mature stage, estimation of progression into harvest was worse than for the unguided method as seen in Figure 9c. This is because, similar to planting, sampled simulated harvest dates are tied to historical harvest distribution windows. Although the $\text{LSTM}_{\text{GSTD}}$ has access to FPAR and rainfall data, the additional simulation GSTD information here clearly hinders the $\text{LSTM}_{\text{GSTD}}$ method's ability to estimate the rate of harvest. Simulated harvest is tied to its historical window, and any relationship learned between simulated harvest timing and in situ harvest during training does not hold for years of early or delayed harvest. The LSTM_{UG} method, which must use only FPAR and weather data to estimate harvest, produced the highest NSE during 2009 in both Iowa and Illinois as shown in Tables 10 and 11.

The GSTD guidance was most impactful during 2012, when high mid-summer temperatures meant that the corn crop in both states progressed much more quickly than the study period average. The GSTD method outperformed the unguided method across all stages as seen in Table 12. The unguided method, lacking the additional information provided by GSTD inputs from DSSAT simulations, was the latest of all methods to estimate progress through grainfill and maturity as seen in Figures 10a, 10b, 11a, and 11b.

3.3 | WSF guidance

The WSF outputs from DSSAT were found to improve in-season CPE when compared with the LSTM_{UG} method particularly during the Illinois 2009 season of delayed planting. The LSTM_{WSF} method outperformed all other methods in accurately estimating the onset of emergence and silking during this season as seen in Table 11 and Figures 9a and 9b. The WSF outputs from DSSAT remain at zero until the DSSAT simulations for the district are started during the historical planting window. During this delayed planting season, simulated WSFs were providing information to the LSTM_{WSF} method before in situ planting had begun. These WSFs, signaling excess water stress through EWSD (Table 5), may then act as a proxy for soil condition during the planting window, which does not have to be derived directly from learned relationships between soil properties and rainfall. Because this simulation–pre-planting overlap happens during all delayed planting seasons, early season simulation EWSD is correlated

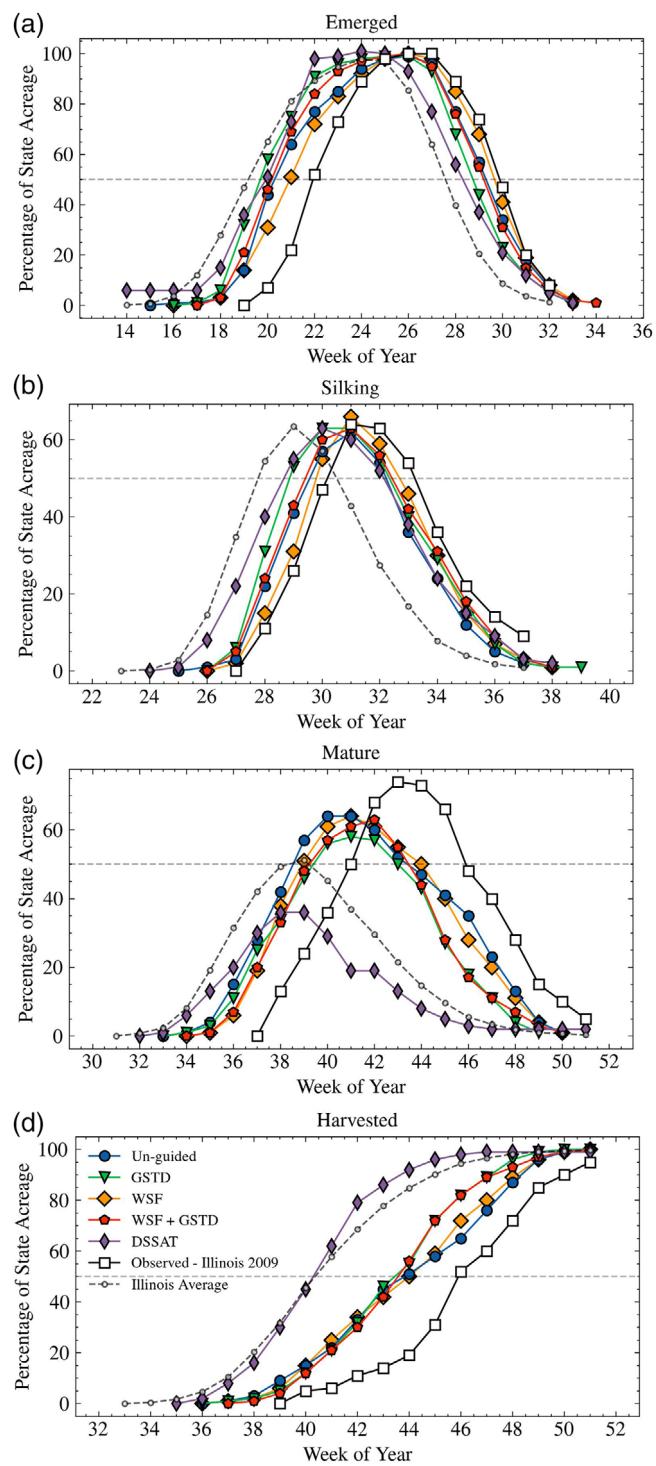


FIGURE 9 Estimated vs. observed phenological progress of corn through the (a) emerged, (b) silking, (c) mature, and (d) harvested stages in Illinois during 2009. DSSAT, Decision Support System for Agrotechnology Transfer; GSTD, growth stage; WSF, water stress factor.

TABLE 12 Median (Med), mean, and standard deviation (Std) of NSE values for all methods across all districts.

Model	Pre-emergence			Emerged			Silking			Grainfill			Mature			Harvested			
	Med		Mean	Med		Mean	Med		Mean	Med		Mean	Med		Mean	Med		Mean	Std
	2004	2009	2012	2014	2004	2009	2012	2014	2004	2009	2012	2014	2004	2009	2012	2014	2004	2009	2012
DSSAT-RR	0.961	0.947	0.034	0.947	0.933	0.038	0.867	0.772	0.18	0.749	0.746	0.111	0.596	0.585	0.232	0.989	0.986	0.01	
Un-guided	0.977	0.969	0.027	0.968	0.96	0.027	0.895	0.826	0.187	0.919	0.905	0.055	0.889	0.783	0.32	0.979	0.939	0.127	
GSTD	0.984	0.976	0.017	0.96	0.957	0.025	0.885	0.798	0.185	0.928	0.897	0.086	0.934	0.907	0.083	0.987	0.982	0.016	
WSF	0.928	0.939	0.035	0.928	0.931	0.042	0.794	0.804	0.128	0.859	0.831	0.12	0.815	0.808	0.098	0.982	0.966	0.036	
WSF+GSTD	0.988	0.985	0.015	0.982	0.975	0.021	0.918	0.871	0.115	0.944	0.92	0.067	0.935	0.898	0.105	0.983	0.974	0.033	
DSSAT-RR	0.947	0.896	0.093	0.911	0.789	0.21	0.8	0.719	0.256	0.833	0.761	0.176	0.252	0.156	0.31	0.517	0.308	0.4	
Un-guided	0.961	0.939	0.071	0.948	0.902	0.117	0.827	0.832	0.081	0.818	0.79	0.116	0.766	0.728	0.169	0.908	0.855	0.117	
GSTD	0.914	0.908	0.078	0.888	0.845	0.141	0.871	0.838	0.101	0.881	0.876	0.058	0.673	0.692	0.124	0.769	0.791	0.114	
WSF	0.974	0.95	0.062	0.972	0.936	0.089	0.915	0.909	0.052	0.877	0.855	0.078	0.797	0.771	0.118	0.866	0.841	0.12	
WSF+GSTD	0.952	0.936	0.064	0.941	0.894	0.115	0.881	0.856	0.092	0.88	0.859	0.078	0.765	0.737	0.121	0.843	0.821	0.105	
DSSAT-RR	0.969	0.905	0.135	0.96	0.945	0.046	0.923	0.92	0.036	0.903	0.881	0.092	0.559	0.499	0.334	0.949	0.935	0.046	
Un-guided	0.964	0.928	0.099	0.944	0.939	0.04	0.757	0.738	0.134	0.689	0.665	0.17	0.592	0.301	1.046	0.947	0.913	0.146	
GSTD	0.972	0.945	0.071	0.971	0.961	0.029	0.896	0.89	0.05	0.904	0.897	0.065	0.709	0.632	0.29	0.967	0.96	0.021	
WSF	0.94	0.933	0.032	0.923	0.919	0.026	0.854	0.825	0.121	0.906	0.863	0.087	0.818	0.719	0.335	0.979	0.963	0.034	
WSF+GSTD	0.935	0.938	0.037	0.951	0.95	0.027	0.934	0.937	0.032	0.93	0.903	0.069	0.717	0.665	0.206	0.953	0.945	0.034	
DSSAT-RR	0.986	0.981	0.01	0.975	0.97	0.016	0.84	0.813	0.117	0.86	0.85	0.072	0.631	0.567	0.157	0.946	0.939	0.036	
Un-guided	0.983	0.979	0.013	0.953	0.949	0.027	0.835	0.816	0.125	0.92	0.921	0.031	0.927	0.849	0.283	0.987	0.971	0.041	
GSTD	0.973	0.964	0.029	0.956	0.952	0.031	0.857	0.836	0.114	0.943	0.923	0.05	0.941	0.94	0.044	0.994	0.991	0.007	
WSF	0.989	0.978	0.026	0.969	0.958	0.031	0.891	0.844	0.144	0.952	0.939	0.038	0.954	0.911	0.126	0.991	0.981	0.025	
WSF+GSTD	0.991	0.984	0.016	0.981	0.975	0.02	0.935	0.924	0.061	0.963	0.916	0.047	0.919	0.08	0.988	0.981	0.019		

Note. DSSAT-RR, Decision Support System for Agrotechnology Transfer -only ridge regression; GSTD, crop growth stage; WSF, water stress factor. Bold values indicate the method with the highest median NSE for each stage.

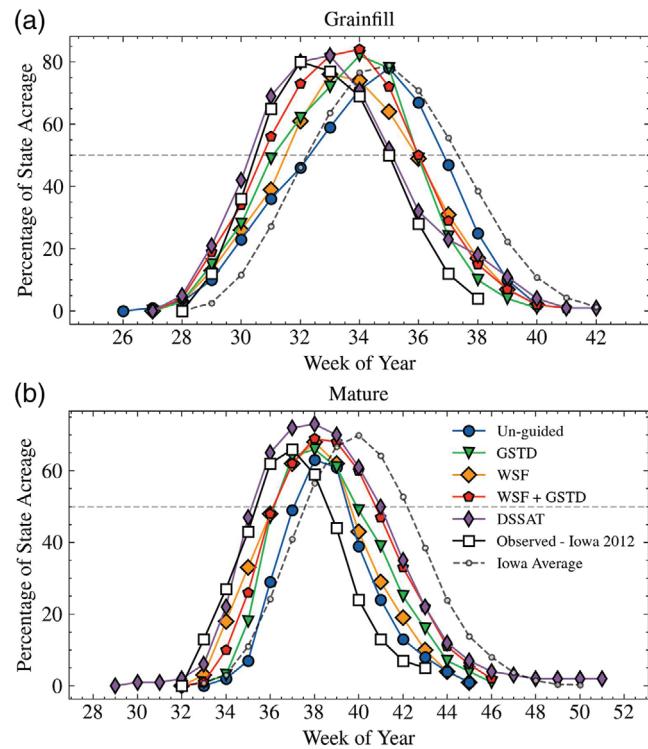


FIGURE 10 Estimated vs. observed phenological progress of corn through the (a) grainfill and (b) mature stages in Iowa during 2012. DSSAT, Decision Support System for Agrotechnology Transfer; GSTD, growth stage; WSF, water stress factor.

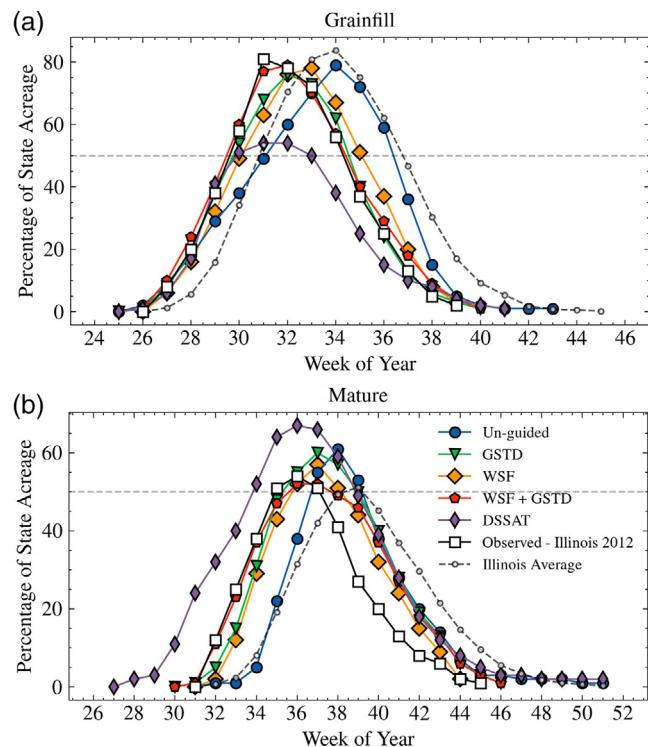


FIGURE 11 Estimated vs. observed phenological progress of corn through the (a) grainfill and (b) mature stages in Illinois during 2012. DSSAT, Decision Support System for Agrotechnology Transfer; GSTD, growth stage; WSF, water stress factor.

with adverse in situ planting conditions. The 2009 Illinois season indicates that the WSF method is able to use this information on soil conditions to better estimate delays in planting because of low field trafficability caused by heavy rainfall.

While EWSD may provide additional information on planting in the early season, LSTM_{WSF} method accuracy falls when excess water stress occurs further into the season. In 2004, the LSTM_{WSF} method performed poorly during excess water stress conditions in June in Iowa, with low NSE during silking shown in Table 10 caused by an early estimate of phenological progress into grainfill as seen in Figure 7a. The unguided LSTM_{UG} method did not experience as significant a decline in performance. While excess water stress does not affect crop growth stage progression within DSSAT, it has been shown to delay actual corn phenological progress (Zaidi et al., 2004). The WSF method, however, does not appear to have correctly learned this relationship.

The WSF-guided methods performed well for later phenological stages as seen in Table 12, such as for the mature stage in 2012. Compounding drought effects for that year in both Iowa and Illinois caused the corn crop to mature significantly earlier than the study period average, and the WSF method was closest to estimating the drought-induced early maturity as seen in Figures 10b and 11b. The WSF method also outperformed all other methods on the mature stage in Illinois during 2014 as seen in Table 11. Excess rainfall delayed harvest during that season, and the high NSE of the WSF method for the mature stage is another indicator that DSSAT WSF outputs are able to provide a proxy for field trafficability conditions.

3.4 | Effects of combined WSF and GSTD guidance

Analyses of each evaluation year show that GSTD and WSF factors are useful at different points in the season. GSTD guidance can provide information on a variable rate of mid-season crop phenology because of changing temperatures but can also negatively affect CPE performance early and late in the season because the simulations from which they are derived are tied to historical planting and harvest windows. The WSFs are able to provide information on soil condition, which NN methods may use to better estimate delayed planting seasons such as 2009. They also provide information useful for estimates of early maturity induced by drought as in 2012. The LSTM_{WSF} method, however, degrades when faced with in-season excess water stress such as for Iowa in 2004.

That the LSTM_{WSF+GSTD} method produced the best NSE across all districts and years for stages pre-emergence through grainfill, as shown in Table 7, is indicative of the complementary nature of GSTD and WSF guidance. Most striking is the performance boost for the silking stage, with an increase in median NSE of 0.044 vs. the next best method, LSTM_{WSF}.

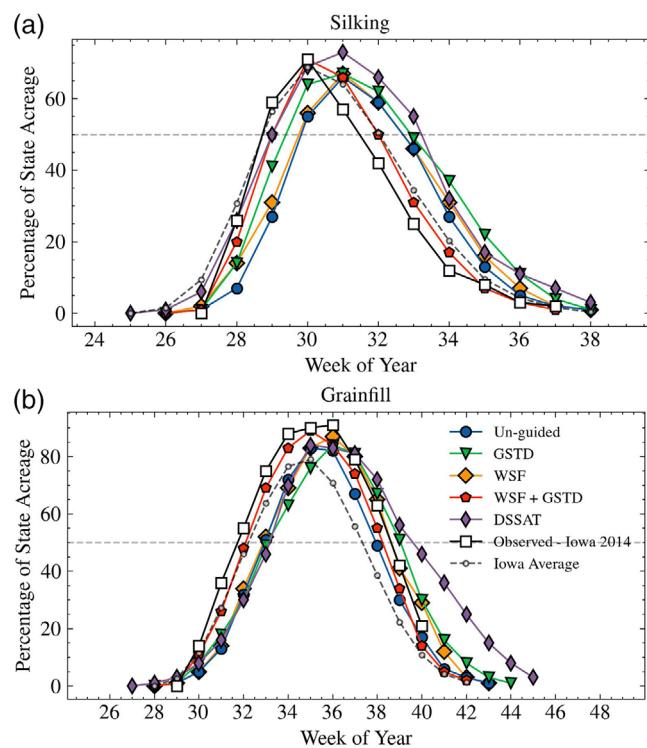


FIGURE 12 Estimated vs. observed phenological progress of corn through the (a) silking and (b) grainfill stages in Iowa during 2014. DSSAT, Decision Support System for Agrotechnology Transfer; GSTD, growth stage; WSF, water stress factor.

This is important given that silking is the phenological stage when corn is most vulnerable to yield loss caused by environmental stressors. This increased accuracy for silking held for both Iowa and Illinois during 2012, as shown in Tables 10 and 11, a year in which drought stress was a major factor in yield loss across the U.S. Midwest.

The Iowa 2004 and Iowa 2014 seasons, when the $\text{LSTM}_{\text{WSF+GSTD}}$ method outperformed both of the other guided methods and demonstrate how guiding DSSAT outputs are complimentary. During Iowa 2004, the $\text{LSTM}_{\text{WSF+GSTD}}$ method outperformed the $\text{LSTM}_{\text{GSTD}}$ method across all stages even when the LSTM_{WSF} method, affected by mid-season excess water stress, performed worse than the unguided method. Estimates by the LSTM_{WSF} and $\text{LSTM}_{\text{GSTD}}$ methods were late and early respectively, with the $\text{LSTM}_{\text{WSF+GSTD}}$ method able to leverage both guiding inputs to produce the most accurate CPE as seen in Figures 7a and 7b. During Iowa 2014, shown in Table 10 and Figures 12a and 12b, while all other NN methods were late in estimation of the onset of both silking and grainfill, the $\text{LSTM}_{\text{WSF+GSTD}}$ method was able to accurately track crop phenology through these stages, producing a median silking NSE of 0.943 vs. 0.824 for the next best method.

The $\text{LSTM}_{\text{WSF+GSTD}}$ performance during Illinois 2009, shown in Table 11, also suggests that the presence of WSF

guidance in the $\text{LSTM}_{\text{WSF+GSTD}}$ method was able to mitigate some NSE performance degradation caused by GSTD during a season of delayed planting. This mitigating ability of using both guiding inputs is important for accurate in-season CPE because the value of GSTD and WSF guiding inputs varies depending on seasonal conditions and upcoming conditions cannot be known *a priori*.

Issue with performance degradation caused by guiding inputs arose from simulation misalignment with in situ crop phenology. This could be rectified by masking in-season simulation values during periods when they have been shown to be detrimental for CPE such as GSTD guidance during the planted and harvested stages. Alternatively, a separate method could be developed to estimate field trafficability and planting dates from RS weather data, building upon recent studies from Paul et al. (2020) and Hoffmann et al. (2022). Dates produced by this method for estimating planting progress could then initiate crop model simulations so that GSTD and WSF outputs are more closely tied to the current season's planting window.

3.5 | Comparison with existing methods

Table 13 shows a comparison of the RMSEs of the four NN-based methods from this study with those reported for methods in the three reference studies (Seo et al., 2019; Diao, 2020; Diao et al., 2021). Patterns in RMSE of transition dates between methods are similar to those of NSE discussed above, with the exception that RMSE for the LSTM_{UG} method is the lowest on both the pre-emergence and emerged stages. The low RMSE of the model-guided methods for the emerged, silking, and grainfill stages in comparison to the reference studies is particularly notable. This is because all of the reference studies cited here use full-season data, meaning their methods use both past and future information from the growing season to produce a transition date estimate. In contrast, all the LSTM-based methods within this study received only in-season information, with no knowledge of the growing season past the target CPR release date. That the $\text{LSTM}_{\text{WSF+GSTD}}$ method produced lower RMSE than the reference methods for all stages except the harvested stage demonstrates the benefits of combining LSTM and crop model outputs for in-season CPE.

4 | CONCLUSION

In this study, three different methods were investigated for using uncalibrated DSSAT CSM-IXIM maize crop model outputs to guide NN-based nRT CPE methods. The crop model-guided methods, which use DSSAT simulation GSTD and WSF outputs, were compared with an unguided, LSTM_{UG} ,

TABLE 13 Root mean square error (RMSE) in days for median transitions dates estimated by all methods across all districts and evaluation years versus RMSE reported in other studies.

Method	Pre-emergence	Emerged	Silking	Grainfill	Mature	Harvested
DSSAT-RR	8.06	10.15	5.75	6.20	12.30	19.88
Unguided	6.63	6.96	4.45	6.06	11.38	15.28
GSTD	7.87	8.52	4.42	5.89	7.58	10.64
WSF	6.65	8.23	4.99	6.77	8.24	9.52
WSF+GSTD	7.05	7.38	3.37	5.41	6.37	11.61
NDVI double logistic fit (Seo et al., 2019)	—	8.00	9.80	—	10.20	—
Combined NDVI matching (Diao, 2020)	—	7.85	10.31	9.88	7.50	8.25
Hybrid phenology NDVI matching (Diao et al. (2021))	—	8.17	3.93	5.73	6.42	9.72

Note. DSSAT-RR, Decision Support System for Agrotechnology Transfer-only ridge regression; GSTD, crop growth stage; WSF, water stress factor; NDVI, normalized difference vegetation index. Bold values indicated the lowest RMSE for each stage.

NN method, and DSSAT-RR. Crop-model guided NNs outperformed the LSTM_{UG} and DSSAT-RR methods, and results showed that crop model outputs are particularly effective during abnormal growing season scenarios such as delayed planting, delayed harvest, and extended drought. Simulated GSTD outputs provided guidance during seasons when abnormal weather increased or slowed crop phenological progress, while simulated water stress factors aided the estimation of late planting, drought-induced early maturity, and delayed harvest. Additionally, guidance from both of these DSSAT outputs produced the best average performance over all districts during the evaluation years. The LSTM_{WSF+GSTD} method showed particularly strong performance on the silking stage, the most critical stage for corn in terms of yield loss, with a median NSE across all evaluation years and districts of 0.930 vs. 0.826 for the LSTM_{UG}. Finally, when compared with three full-season reference studies, the LSTM_{WSF+GSTD} method produced lower RMSE for all stages except harvested. This study demonstrates that uncalibrated crop model outputs can provide additional information to NN-based methods for in-season CPE at district scale in the U.S. states of Iowa and Illinois. Extending this hybrid approach to CPE to other regions could provide valuable nRT information of crop phenology useful for producers, agronomists, and policy makers to form assessments of crop stress risk and yield impact that are directly linked to crop phenology.

AUTHOR CONTRIBUTIONS

George Worrall: Conceptualization; data curation; formal analysis; investigation; methodology; software; validation; visualization; writing—original draft. **Jasmeet Judge:** Conceptualization; funding acquisition; methodology; resources; supervision; visualization; writing—review and editing. **Kenneth Boote:** Conceptualization; methodology; writing—review and editing. **Anand Rangarajan:** Conceptualization; methodology.

ACKNOWLEDGMENTS

The authors would like to thank anonymous reviewers and editors for their time and constructive comments. We gratefully acknowledge the USDA-NASS Upper Midwest and Heartland Regional Offices for providing the tabulated data from USDA Crop Progress and Condition reports. This study was funded in part by the University of Florida Informatics Institute and the Center for Remote Sensing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

George Worrall  <https://orcid.org/0000-0002-3885-4795>
Kenneth Boote  <https://orcid.org/0000-0002-1358-5496>

REFERENCES

- Akhavizadegan, F., Ansarifar, J., Wang, L., Huber, I., & Archontoulis, S. V. (2021). A time-dependent parameter estimation framework for crop modeling. *Scientific Reports*, 11, 11437. <https://doi.org/10.1038/s41598-021-90835-x>
- Barbosa dos Santos, V., Santos, A. M. F. d., & Rolim, G. d. S. (2021). Estimation and forecasting of soybean yield using artificial neural networks. *Agronomy Journal*, 113, 3193–3209. <https://doi.org/10.1002/agj2.20729>
- Claassen, M. M., & Shaw, R. H. (1970). Water deficit effects on corn. II. Grain components. *Agronomy Journal*, 62, 652–655. <https://doi.org/10.2134/agronj1970.00021962006200050032x>
- Diao, C. (2020). Remote sensing phenological monitoring framework to characterize corn and soybean physiological growing stages. *Remote Sensing of Environment*, 248, 111960. <https://doi.org/10.1016/j.rse.2020.111960>
- Diao, C., Yang, Z., Gao, F., Zhang, X., & Yang, Z. (2021). Hybrid phenology matching model for robust crop phenological retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 308–326. <https://doi.org/10.1016/j.isprsjprs.2021.09.011>
- Doss, B. D., Pearson, R. W., & Rogers, H. T. (1974). Effect of soil water stress at various growth stages on soybean yield.

- Agronomy Journal*, 66, 297–299. <https://doi.org/10.2134/agronj1974.00021962006600020032x>
- Gao, F., Anderson, M., Daughtry, C., Karnieli, A., Hively, D., & Kustas, W. (2020a). A within-season approach for detecting early growth stages in corn and soybean using high temporal and spatial resolution imagery. *Remote Sensing of Environment*, 242, 111752. <https://doi.org/10.1016/j.rse.2020.111752>
- Gao, F., Anderson, M. C., & Hively, W. D. (2020b). Detecting cover crop end-of-season using VENμS and Sentinel-2 satellite imagery. *Remote Sensing*, 12, 3524. <https://doi.org/10.3390/rs12123524>
- Gao, F., & Zhang, X. (2021). Mapping crop phenology in near real-time using satellite remote sensing: Challenges and opportunities. *Journal of Remote Sensing*, 2021, 8379391. <https://doi.org/10.34133/2021/8379391>
- Graves, A. (2012). Long short-term memory. In *Supervised sequence labelling with recurrent neural networks* (pp. 37–45). Springer. https://doi.org/10.1007/978-3-642-24797-2_4
- Han, E., Ines, A. V., & Koo, J. (2019). Development of a 10-km resolution global soil profile dataset for crop modeling applications. *Environmental Modelling & Software*, 119, 70–83. <https://doi.org/10.1016/j.envsoft.2019.05.012>
- Hoffmann, S., Schönauer, M., Heppelmann, J., Asikainen, A., Cacot, E., Eberhard, B., Hasenauer, H., Ivanovs, J., Jaeger, D., Lazzins, A., Mohtashami, S., Moskalik, T., Nordfjell, T., Stereñczak, K., Talbot, B., Uusitalo, J., Vuillermoz, M., & Astrup, R. (2022). Trafficability prediction using Depth-to-Water maps: The status of application in northern and central European forestry. *Current Forestry Reports*, 8, 55–71. <https://doi.org/10.1007/s40725-021-00153-8>
- Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., ... Keating, B. A. (2014). APSIM – Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, 62, 327–350. <https://doi.org/10.1016/j.envsoft.2014.07.009>
- Hussain, J., Khaliq, T., Ahmad, A., & Akhtar, J. (2018). Performance of four crop model for simulations of wheat phenology, leaf growth, biomass and yield across planting dates. *PLOS ONE*, 13, e0197546. <https://doi.org/10.1371/journal.pone.0197546>
- Jia, X., Willard, J., Karpatne, A., Read, J. S., Zwart, J. A., Steinbach, M., & Kumar, V. (2020). Physics-Guided machine learning for scientific discovery: An application in simulating lake temperature profiles. *ACM/IMS Transactions on Data Science*, 2, 1–26. <https://doi.org/10.1145/3447814>
- Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., Wilkens, P., Singh, U., Gijsman, A., & Ritchie, J. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18, 235–265. [https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7)
- Kamali, B., Abbaspour, K. C., Lehmann, A., Wehrli, B., & Yang, H. (2018). Spatial assessment of maize physical drought vulnerability in sub-Saharan Africa: Linking drought exposure with crop failure. *Environmental Research Letters*, 13, 074010. <https://doi.org/10.1088/1748-9326/aacb37>
- Kedem, B., & Chiu, L. S. (1987). On the lognormality of rain rate. *Proceedings of the National Academy of Sciences of the United States of America*, 84, 901–905. <https://doi.org/10.1073/pnas.84.4.901>
- Kerner, H. R., Sahajpal, R., Pai, D. B., Skakun, S., Puricelli, E., Hosseini, M., Meyer, S., & Becker-Reshef, I. (2022). Phenological normalization can improve in-season classification of maize and soybean: A case study in the central US corn belt. *Science of Remote Sensing*, 6, 100059. <https://doi.org/10.1016/j.srs.2022.100059>
- Liu, L., Zhang, X., Yu, Y., Gao, F., & Yang, Z. (2018). Real-time monitoring of crop phenology in the midwestern united states using VIIRS observations. *Remote Sensing*, 10, 1540. <https://doi.org/10.3390/rs10101540>
- Lizaso, J. I., Boote, K. J., Jones, J. W., Porter, C. H., Echarte, L., Westgate, M. E., & Sonohat, G. (2011). CSM-IXIM: A new maize simulation model for DSSAT version 4.5. *Agronomy Journal*, 103, 766–779. <https://doi.org/10.2134/agronj2010.0423>
- Mehrabi, Z., & Ramankutty, N. (2019). Synchronized failure of global crop production. *Nature Ecology & Evolution*, 3, 780–786. <https://doi.org/10.1038/s41559-019-0862-x>
- Myinen, R., Knyazikhin, Y., & Park, T. (2015). *MCD15A3H MODIS/Terra+Aqua leaf area Index/FPAR 4-day L4 global 500m SIN grid V006*. NASA EOSDIS Land Processes DAAC. <https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MCD15A3H>
- Paul, S., Coops, N., Johnson, M., Krzic, M., Chandna, A., & Smukler, S. (2020). Mapping soil organic carbon and clay using remote sensing to predict soil workability for enhanced climate change adaptation. *Geoderma*, 363, 114177. <https://doi.org/10.1016/j.geoderma.2020.114177>
- Peng, Y., Kira, O., Nguy-Robertson, A., Suyker, A., Arkebauer, T., Sun, Y., & Gitelson, A. A. (2019). Gross primary production estimation in crops using solely remotely sensed data. *Agronomy Journal*, 111, 2981–2990. <https://doi.org/10.2134/agronj2019.05.0332>
- Sakamoto, T., Wardlow, B. D., Gitelson, A. A., Verma, S. B., Suyker, A. E., & Arkebauer, T. J. (2010). A two-step filtering approach for detecting maize and soybean phenology with time-series MODIS data. *Remote Sensing of Environment*, 114, 2146–2159. <https://doi.org/10.1016/j.rse.2010.04.019>
- Seo, B., Lee, J., Lee, K.-D., Hong, S., & Kang, S. (2019). Improving remotely-sensed crop monitoring by NDVI-based crop phenology estimators for corn and soybeans in Iowa and Illinois, USA. *Field Crops Research*, 238, 113–128. <https://doi.org/10.1016/j.fcr.2019.03.015>
- Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US corn belt. *Scientific Reports*, 11, 1606. <https://doi.org/10.1038/s41598-020-80820-1>
- Shen, Y., Wu, L., Di, L., Yu, G., Tang, H., Yu, G., & Shao, Y. (2013). Hidden Markov models for real-time estimation of corn progress stages using MODIS and meteorological data. *Remote Sensing*, 5, 1734–1753. <https://doi.org/10.3390/rs5041734>
- Thornton, M., Shrestha, R., Wei, Y., Thornton, P., Kao, S., & Wilson, B. (2020). *Daymet: Daily surface weather data on a 1-km grid for North America, version 4*. ORNL Distributed Active Archive Center.
- U.S. Department of Commerce, NOAA, & USDA. (2021). Weekly weather and crop bulletin. <https://downloads.usda.library.cornell.edu/usda-esmis/files/cj82k728n/kp78hc468/9306tt80c/wwcb2021.pdf>
- USDA National Agricultural Statistics Service (USDA-NASS). (2005a). *2005 Illinois annual statistics bulletin*. USDA-NASS, Illinois Field Office. https://www.nass.usda.gov/Statistics_by_State/Illinois/Publications/Annual_Statistical_Bulletin/2005/index.php
- USDA National Agricultural Statistics Service (USDA-NASS). (2010a). *2010 Illinois annual statistics bulletin*. USDA-NASS Illinois Field Office. https://www.nass.usda.gov/Statistics_by_State/Illinois/Publications/Annual_Statistical_Bulletin/2010/index.php

- USDA National Agricultural Statistics Service (USDA-NASS). (2013a). *Illinois agricultural statistics 2013 annual bulletin*. USDA-NASS, Illinois Field Office.
- USDA National Agricultural Statistics Service (USDA-NASS). (2015a). *Illinois agricultural statistics 2015 annual bulletin*. USDA-NASS, Illinois Field Office. https://www.nass.usda.gov/Statistics_by_State/Illinois/Publications/Annual_Statistical_Bulletin/2015/2015-IL_Annual%20Bulletin_Entire_Report.pdf
- USDA National Agricultural Statistics Service (USDA-NASS). (2005b). *2005 Iowa agricultural statistics*. USDA-NASS, Iowa Field Office. https://www.nass.usda.gov/Statistics_by_State/Iowa/Publications/Annual_Statistical_Bulletin/2005_Iowa_Annual_Bulletin.pdf
- USDA National Agricultural Statistics Service (USDA-NASS). (2010b). *2010 Iowa agricultural statistics*. USDA-NASS, Iowa Field Office. https://www.nass.usda.gov/Statistics_by_State/Iowa/Publications/Annual_Statistical_Bulletin/2010_Iowa_Annual_Bulletin.pdf
- USDA National Agricultural Statistics Service (USDA-NASS). (2013b). *2013 Iowa agricultural statistics*. USDA-NASS, Iowa Field Office. https://www.nass.usda.gov/Statistics_by_State/Iowa/Publications/Annual_Statistical_Bulletin/2013_Iowa_Annual_Bulletin.pdf
- USDA National Agricultural Statistics Service (USDA-NASS). (2015b). *2015 Iowa agricultural statistics*. USDA-NASS, Iowa Field Office. https://www.nass.usda.gov/Statistics_by_State/Iowa/Publications/Annual_Statistical_Bulletin/2015_Iowa_Annual_Bulletin.pdf
- USDA National Agricultural Statistics Service (USDA-NASS). (2021a). *National crop progress*. USDA-NASS. https://www.nass.usda.gov/Charts_and_Maps/Crop_Progress_&_Condition/2021/index.php
- USDA National Agricultural Statistics Service (USDA-NASS). (2021b). *CropScape—Cropland Data layer: Published crop-specific data layer*. USDA-NASS. <https://nassgeodata.gmu.edu/CropScape/>
- USDA National Agricultural Statistics Service (USDA-NASS). (2021c). *Quick stats*. USDA-NASS. <https://quickstats.nass.usda.gov>
- Wang, S., Di Tommaso, S., Deines, J. M., & Lobell, D. B. (2020). Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive. *Scientific Data*, 7, 307. <https://doi.org/10.1038/s41597-020-00646-4>
- Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2020). Integrating physics-based modeling with machine learning: A survey. *arXiv*. <https://doi.org/10.48550/arXiv.2003.04919>
- Worrall, G., Rangarajan, A., & Judge, J. (2021). Domain-guided machine learning for remotely sensed in-season crop growth estimation. *Remote Sensing*, 13, 4605. <https://doi.org/10.3390/rs13224605>
- Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., Huang, J., Li, H., & Lin, T. (2020). DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sensing of Environment*, 247, 111946. <https://doi.org/10.1016/j.rse.2020.111946>
- Zaidi, P. H., Rafique, S., Rai, P., Singh, N., & Srinivasan, G. (2004). Tolerance to excess moisture in maize (*Zea mays* L.): Susceptible crop stages and identification of tolerant genotypes. *Field Crops Research*, 90, 189–202. <https://doi.org/10.1016/j.fcr.2004.03.02>
- Zeng, L., Wardlow, B. D., Wang, R., Shan, J., Tadesse, T., Hayes, M. J., & Li, D. (2016). A hybrid approach for detecting corn and soybean phenology with time-series MODIS data. *Remote Sensing of Environment*, 181, 237–250. <https://doi.org/10.1016/j.rse.2016.03.039>
- Zhang, C., Marzougui, A., & Sankaran, S. (2020). High-resolution satellite imagery applications in crop phenotyping: An overview. *Computers and Electronics in Agriculture*, 175, 105584. <https://doi.org/10.1016/j.compag.2020.105584>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Worrall, G., Judge, J., Boote, K., & Rangarajan, A. (2023). In-season crop phenology using remote sensing and model-guided machine learning. *Agronomy Journal*, 115, 1214–1236. <https://doi.org/10.1002/agj2.21230>