ISPRS Journal of Photogrammetry and Remote Sensing 231 (2026) 101–118

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Corn yield estimation under extreme climate stress with knowledge-encoded deep learning

Xingguo Xiong [a], Renhai Zhong [b], Hao Jiang [c], Ioannis Athanasiadis [d], Yi Yang [e], Linchao Zhu [e,*], Tao Lin [a,*]

[a] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, Zhejiang 310058, China
[b] Institute of Applied Remote Sensing and Information Technology, Zhejiang University, Hangzhou 310058, China
[c] China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou, Zhejiang 310006, China
[d] Artificial Intelligence Group, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands
[e] College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China

ARTICLE INFO

Keywords:
Crop yield estimation
Knowledge-encoded machine learning
Extreme climate stress
Remote sensing
Time lag
Cross attention

ABSTRACT

Accurately estimating crop yield under climate stress is vital for global food security, particularly as extreme weather events become more frequent. Data-driven models are increasingly adopted for enhancing yield estimation. They benefit from the effective learning of crop response to the environment from vast amounts of remote sensing and meteorological data. Extreme climate stress conditions that have few yield labels available failed to train these models for modeling crop-stress interactions. Crop responses to such extreme climate stress could exhibit significant delays and sensitivities that can be captured through remote sensing observations. However, this knowledge is not sufficiently utilized in data-driven yield estimation models to address the lack of labels under extreme stress. This study employs attention mechanisms to explicitly encode the time lag effect and phenology sensitivity within a deep learning framework for predicting crop yield under extreme climate stress. The framework consists of two-stream structure separately receiving climate and remote sensing data, with each aggregating input time series into multitemporal feature embeddings. A time-lag-encoded cross attention fuses feature embeddings between climate and remote sensing streams, while phenology-sensitivity-guided linear attention is applied atop the framework for processing these time-lag-encoded features. The proposed model is evaluated across nine Midwestern states within the US Corn Belt at the county level from 2006 to 2023, simulating extreme climate stress scenarios with limited samples. The time lag analysis indicates an average lag of approximately 45 days between the extreme stress event and the maximum vegetation decay event, revealing that such extreme events lead to delayed reductions in crop greenness. General model evaluation results demonstrate that the knowledge-encoded two-stream model (RMSE = 1.17 Mg ha^{-1}) outperforms both the feature-stacking-based two-stream model (RMSE = 1.43 Mg ha^{-1}) and the random forest model (RMSE = 1.68 Mg ha^{-1}) under conditions of extreme climate stress. Model ablation results show that cross attention and time-lag knowledge significantly improve model accuracy compared to direct sum of features, suggesting knowledge-encoded data fusion is more effective than simply summing multi-source input data. The incorporation of time lag-encoded cross-attention mechanisms also facilitates the identification of distinct time lag patterns since extreme climate stress happened, thereby enhancing the model's interpretability and providing insights into the interactions between environmental factors and crop responses. In-season analysis reveals that the time lag-encoded model captures extreme stress events once they occur, enabling accurate yield predictions up to 8 weeks in advance. The spatial–temporal transferability experiment shows that the knowledge-encoded two-stream model outperforms baseline models across counties from 2013 to 2023. Notably, the proposed model achieves substantial accuracy gains in regions experiencing extreme heat stress, and it also maintains robust performance across most years. Overall, the time lag knowledge could be extended to other forms of environmental stress as long as it's captured by multi-source data. The cross-attention mechanism as a basic unit enables integration with more knowledge to improve the modeling of complex biomass accumulation and yield formation.

* Corresponding authors.
E-mail addresses: zhulinchao@zju.edu.cn (L. Zhu), lintao1@zju.edu.cn (T. Lin).

https://doi.org/10.1016/j.isprsjprs.2025.10.020
Received 5 March 2025; Received in revised form 4 September 2025; Accepted 18 October 2025
Available online 28 October 2025

1. Introduction

Accurate crop yield estimation under climate stress is essential for global food security. The frequency of extreme events has increased across many regions of the world in recent years and is expected to rise further (Gabaldón-Leal et al., 2016; Lesk et al., 2022; Vogel et al., 2019). Major production areas have faced an increasing risk of yield failures due to the emergence of extreme stress, particularly during the reproductive stages (Rezaei et al., 2023; Vogel et al., 2019). Robust yield estimation models that highlight intensified stress and worsen canopy conditions due to various stressors are necessary to improve yield estimates and enhance understanding of crop responses to environmental stresses (Anderson et al., 2016; Peng et al., 2018; Prasad et al., 2015). Effectively representing the interactions between the environment and crop growth is a prerequisite for developing these models to assess how crops will ultimately respond to climate change (Schlenker and Roberts, 2009).

Crops interact dynamically with environmental changes, and the impacts of extreme climate stress can be captured through multi-source observations during the growing season. Crop responses to climate stress are often delayed, with canopy greenness showing lagged declines following the extreme stress (Jiang et al., 2025; Liu et al., 2018; Yang et al., 2024). These time-lag effects have been studied in both ecological and agricultural contexts to explore lagged impacts on vegetation and crops (Gao et al., 2025; Kong et al., 2020; Liu et al., 2016, 2018; Yang et al., 2024). The time-lag effect can occur throughout the crop growth cycle, but its impact on yield is unevenly distributed across stages. Crop responses to climate stress exhibit varying sensitivity across phenological stages, with reproductive periods being more vulnerable than others. The flowering stage is generally considered one of the most sensitive phases for yield determination in many crops, as it governs pollination success and final grain formation (Barnabás et al., 2008; Deryng et al., 2014; Kang et al., 2009; Lesk et al., 2022; Rezaei et al., 2023; Vogel et al., 2019). Therefore, jointly modeling time-lag effects and phenology sensitivity may enhance the modeling of crop response to climate stress in yield estimation.

Meteorological and remote sensing observations are the widely used dynamic data for characterizing the interaction of crop growth and the environment (Feng et al., 2021; Peng et al., 2018; Sakamoto, 2020). The time-lag effects are usually quantified by aligning vegetation dynamics with preceding climate anomalies using remote sensing and climate data (Gao et al., 2025; Liu et al., 2016; Yang et al., 2024). The common approach to quantify time lag is to calculate the lagged durations between the onset of climate stress and the subsequent decline in canopy greenness. However, it remains challenging to accurately link each vegetation decline event to a specific stress event. Besides the delays between events, the vegetation canopy may also exhibit prolonged damage following the extreme climate stress. Daily climate indicators could reveal the burst points of extreme stress, and vegetation time series reflect the cumulative physiological response until the largest decay. The interval between these two extremes may represent a meaningful window of the long-lasting crop response of extreme climate stress. Incorporating the features within this time-lag duration into yield estimation models offers a way to extract extreme stress-related signals, thereby improving the robustness of crop yield estimation under extreme climate conditions.

Data assimilation methods based on crop simulation models are one of the representative approaches to integrate extra observations into yield estimation frameworks. Process-based crop models simulate crop growth and development under different weather, soil, and management conditions (Challinor et al., 2014; Shuai and Basso, 2022). Data assimilation techniques incorporate remote sensing variables into crop models by updating internal state variables to improve simulation accuracy (Han et al., 2022; Huang et al., 2016; Ines et al., 2013, 2013; Jin et al., 2017; Kang and Özdoğan, 2019) (Table 1). However, these approaches often struggle under extreme climate conditions due to model structural limitations and uncertainty propagation (Shuai and Basso, 2022). They lack explicit mechanisms to represent observed physiological phenomena such as lagged crop responses and phenological sensitivity. These limitations highlight the need for more flexible frameworks that can integrate such domain-informed patterns.

Data-driven approaches, known for their strong learning capacity, have been increasingly applied to model crop yield using multi-source data (Table 1). Purely data-driven deep neural networks facilitate recognizing spatial–temporal-spectral pattern of crops from vast amounts of data. Many studies utilize long short-term memory (LSTM) networks to capture crop growth dynamics from remote sensing and meteorological data (Feng et al., 2021; Jiang et al., 2020; Zhong et al., 2023). Vision Transformers have been introduced to enhance temporal feature learning by operating on image patch sequences (Lin et al., 2024). To make full use of high temporal resolution data, pyramid architectures combining LSTM and Transformer modules have been proposed to extract fine-grained features from daily data (Xiong et al., 2024). Researchers have also attempted to use graph neural networks to learn spatial dependencies among agricultural regions, improving predictive performance (Cheng et al., 2024; Fan et al., 2022; Peng et al., 2024). These studies typically stack all input variables per time step, ignoring the potentially causal dependencies between variables over the temporal domain. For example, there is often a time lag between climate stress and canopy greenness responses, which can be observed through climate variables and vegetation indices. Recently developed models employ multiple streams of neural network to separately extract features from different data modalities, and often fuse the learned representations through a simple concatenation layer (Huang et al., 2022; Lin et al., 2024; Peng et al., 2024). A more recent extension uses gated fusion mechanisms to adaptively integrate heterogeneous features (Mena et al., 2025). Although these fully data-driven yield estimation models offer improve multi-modal learning, they typically align all inputs by calendar day, and fail to capture the temporal misalignment and lagged dependencies across variables. On the other hand, they often require extensive labeled data from real-world scenarios to achieve high accuracy. This reliance on large datasets poses challenges in climate stress scenarios with limited samples.

Encoding knowledge into data-driven models presents a potential strategy to effectively utilize the stress-related interactive signals like time lag and reduce the data requirements. Many studies have attempted to build hybrid models that integrate crop simulation knowledge with data-driven approaches (Table 1). The primary idea is to use simulation data and labels from crop simulation models to train a proxy machine learning model (Feng et al., 2020, 2019; Hsiao et al., 2019; Jeong et al., 2022; Liu et al., 2024; Shahhosseini et al., 2021; Xiao et al., 2024; Yang et al., 2023). However, traditional crop simulation models often struggle to accurately simulate crop responses to environmental stress (Nóia Júnior et al., 2025), making it challenging to generate simulation data representing stress-induced interactive signals for training proxy deep learning models. What is needed is a new approach that explicitly incorporates observed signals into the neural network structure. In deep learning, cross-attention mechanisms are frequently used as a flexible strategy for modeling interactive relationships between features from multiple sources (Chen et al., 2021; Wei et al., 2020). Cross-attention is a variant of the typical self-attention mechanism, realizing the “cross” function by switching the query and key to weight either side of the features. To broaden the applicability of attention mechanisms, many studies have employed predefined rules to enforce learned attention maps focusing on specific ranges of locations (Kim et al., 2023; Phung et al., 2024). Researchers have adapted this concept to manually adjust attention weights using phenological sensitivity in crop yield modeling, showing positive effects in improving crop yield estimation (Qiao et al., 2023). However, this approach primarily focuses on normal climate conditions. The incorporation of stress-related physiological knowledge, such as time-lag effects observed after heat or drought events and phenological sensitivity of crop response to extreme climate stress,

remains underexplored.

In this study, we develop a two-stream deep learning framework to allow for encoding time-lag effects and phenological sensitivity using attention mechanisms, facilitating crop yield estimation under extreme climate stress. The framework contains the remote sensing and climate streams, each stream employing a pyramid structure that progressively aggregates input time series into feature embeddings. We explicitly encode the time lag effect between the two streams based on cross attention. Phenology-sensitivity-guided linear attention processes the time-lag-encoded features on top of the two-stream framework. The proposed model is evaluated across nine Midwestern states within the US Corn Belt at the county level from 2006 to 2023, simulating climate stress situations with fewer samples to address the following questions:

(1) What are the patterns of the time lag and phenology sensitivity between remote sensing and meteorological data during extreme stress in crop growth?
(2) What are the contributions of the time-lag and phenological sensitivity to the yield estimation?
(3) How do these two knowledge improve the in-season estimation performance?
(4) How does the knowledge-encoded model perform across years and regions?

2. Method

2.1. Knowledge-encoded two-stream framework

The two-stream framework is proposed to embed the time-lag effect of climate stress on remote sensing data for estimating corn yield under climate stress (Fig. 1). The two streams, designed as the remote sensing and climate stream, receive time series data of remote sensing and climate data as input, respectively. Each stream implements a pyramid structure to hierarchically aggregate the input time series into multi-time-scale feature embeddings. These features of the two streams are then fused using the time-lag encoded cross attention to enhance the interaction between the two data sources. This fusion strategy differs from previous input-level fusion methods that stack multi-source data as input without any underlaying interaction knowledge. The resulting multi-temporal features are further constrained by a phenology sensitivity vector to estimate the final corn yield.

2.1.1. Remote sensing stream

The remote sensing stream employs a pyramid structure to extract the temporal features of the crop growth captured by weekly remote sensing observations. Weekly time steps are generated using temporal aggregation to alleviate the missing data problems (Khaki and Wang, 2019; You et al., 2017; Zhong et al., 2023). Two Vegetation Indices (VIs), Wide dynamic range vegetation index (WDRVI) and Normalized difference water index (NDWI), are utilized as they are effective in

Table 1
The representative yield estimation studies incorporating remote sensing and environmental data.

Fusion method	Base model	Technical points	Crop	Year	Spatial resolution	Author
Data assimilation	WOFOST	Assimilating observed soil moisture using EnKF	Corn	2007	50 km	(De Wit and Van Diepen, 2007)
	WOFOST	Assimilating RS retrieved LAI by recalibrating the crop model	Winter wheat	2011	Field	(Curnel et al., 2011)
	CERES-Maize	Assimilating RS retrieved multiple variables using MCMC EnKF	Corn	2013	County	(Ines et al., 2013)
	WOFOST	Assimilating RS retrieved LAI using EnKF	Winter wheat	2016	30 *m*	(Huang et al., 2016)
	AquaCrop	Assimilating RS retrieved biomass and canopy cover using PSO	Winter wheat	2017	Field	(Jin et al., 2017)
	SAFYsw	Assimilating RS retrieved LAI using MCMC EnKF	Corn	2019	30 *m*	(Kang and Özdoğan, 2019)
	AquaCrop	Assimilating DL estimated multiple variables using MCMC EnKF	Corn	2019	Field	(Han et al., 2019)
Data driven	Typical ML	RS estimates and CV as input	Soybean	2020	County	(Sakamoto, 2020)
	Typical ML	Time lag metric from multi-modal data as input	Corn, Wheat, Soy	2019	County	(Mateo-Sanchis et al., 2019)
	ANN	Concat raw data as input	Winter wheat	2021	County	(Feng et al., 2021)
	LSTM	Concat raw data as input	Corn	2019	County	(Jiang et al., 2020)
	Dual-stream DL	Concat features for final prediction	Winter wheat	2022	County	(Huang et al., 2022)
	LSTM–GNN	Concat features for GNN learning	Winter wheat	2024	County	(Peng et al., 2024)
	ViT	Multi-modal relationships learned by self-supervised learning	Corn etc.	2023	County	(Lin et al., 2024)
	Gated fusion	Features aggregated by learnable gated fusion weights	Wheat etc.	2025	Field	(Mena et al., 2025)
Hybrid modeling	Multiple ML models	Training using simulation data from APSIM	Corn	2021	County	(Shahhosseini et al., 2021)
	RF	Training using simulation data from APSIM	Wheat	2020	Site	(Feng et al., 2020)
	CNN and LSTM	Training using simulation data from RSCM	Rice	2022	500 *m*	(Jeong et al., 2022)
	GRU	Training using simulation data from ecosys	Corn and soybean	2024	Field/County	(Liu et al., 2024)
	GRU	Training using simulation data from ecosys	Corn and soybean	2023	Field/County	(Yang et al., 2023)
	XGBoost	Training using simulation data from APSIM	Wheat and corn	2024	Site/grid	(Xiao et al., 2024)
	Object optimization	Building a trainable crop process model	Corn	2023	County	(Chang et al., 2023)
	Attention	Adjust the attention weights for knowledge embedding	Corn and soybean	2023	County	(Qiao et al., 2023)

Abbreviations: EnKF means Ensemble Kalman Filter, RS represents Remote Sensing, LAI stands for Leaf Area Index, MCMC refers to Markov Chain Monte Carlo, PSO indicates Particle Swarm Optimization, ML denotes Machine Learning, CV means Climate Variables, ANN is short for Artificial Neural Network.
LSTM represents Long Short-Term Memory, GNN stands for Graph Neural Network, ViT means Vision Transformer, GRU refers to Gated Recurrent Unit, XGBoost indicates Extreme Gradient Boosting, APSIM stands for Agricultural Production Systems sIMulator, RSCM means Rice Simulation Crop Model.
ecosys refers to Ecosystem Model, RF denotes Random Forest, Attention indicates Attention Mechanism.

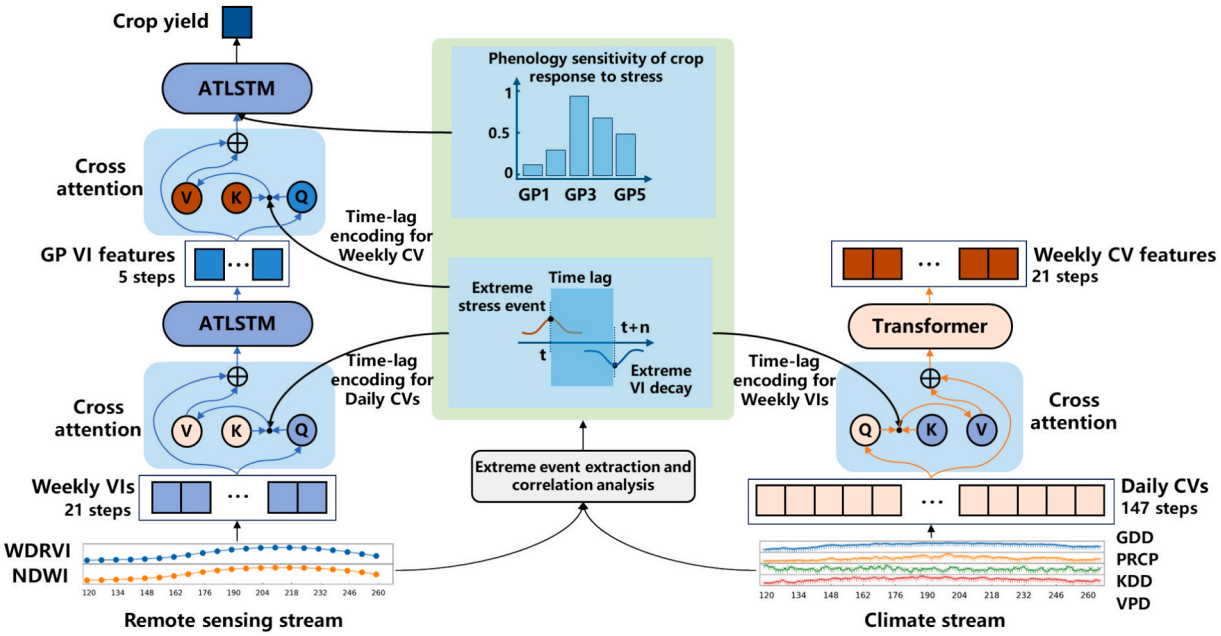

Fig. 1. The two stream framework to embed the time lag and phenology sensitivity in fusing remote sensing and climate data for corn yield estimation. The blue color represents the feature flow in remote sensing stream and the shallow yellow color represents the climate one. Shallow gray is the process of quantifying the time lag and phenology sensitivity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

representing the canopy cover and water status of the crop, thereby facilitating crop growth characterization (Gao, 1995; Gitelson, 2004). These two indicators also rank the highest important feature in previous deep learning attribution analysis (Xiong et al., 2024).

The pyramid structure comprises three layers and follows the hierarchical encoder structure from the previous study for hierarchical temporal feature learning (Xiong et al., 2024). The pyramid structure consists of multiple layers to aggregate the long input series into short sequences layer by layer. In the first layer, a linear transformation is employed on feature dimension to project the WDRVI and NDWI into high-dimensional feature embeddings. The second layer integrates these weekly feature embeddings with climate features through time-lag encoded cross-attention, enhancing the features relevant to durations affected by time-lag effects. The knowledge-encoded weekly input features are then segmented into 5 sequences corresponding to growth phases for further aggregation. During the whole growing season, the corn had 5 growing phases (GP) with varied length. GP1 was duration from corn planted to emerged, GP2 was the period from emerged to three weeks before silking. GP3 was from three weeks before silking to milk. GP4 was the period from milk to dented. GP5 was the duration from dented to mature. Therefore, the sequences hold temporal length of 4,5,4,4,4 to align with the growing phases. These sequences are aggregated using an attention-based long-short term memory network (ATLSTM). The ATLSTM is a one-layer LSTM appended by an attention layer. The attention layer is composed of a linear layer with SoftMax activation that learned the attention weights of each hidden feature from previous LSTM. In the third layer, the growth phase (GP) VI features are combined with climate features via an additional time-encoded cross attention for feature enhancement. Another ATLSTM then aggregates these enriched GP sequences into a single high-dimensional vector for estimating the final crop yield. The phenology sensitivity constrains the attention calculation in this ATLSTM, emphasizing the importance of time lags that cover key phenological stages.

2.1.2. Climate stream

The climate stream uses a similar pyramid structure but incorporates a daily aggregation layer at the input end to extract dynamic climate stress from daily climate variables (CVs). Daily CVs provide much more temporal information compared to weekly time series. These temporal details enhance the ability to exploit crop features across both short-term and long-term temporal ranges. Four climate variables, which represent diverse climate conditions, constitute the daily CVs. Growing degree days (GDD) quantify effective heat accumulation, while precipitation (PRCP) indicates water supply. Additionally, Killing Degree Days (KDD) and Vapor Pressure Deficit (VPD) are employed to reflect extreme heat accumulation and the degree of dryness in the air, respectively (Butler et al., 2018; Hsiao et al., 2019).

The pyramid structure of the climate stream consists of two layers utilizing Transformer encoder temporal aggregations. In the first layer, a linear transformation converts the 147-time-step daily climate variables (CVs) of 4 features into feature embeddings that match the dimensionality of the weekly VI feature embeddings. In the second layer, the daily CV embeddings are enhanced by incorporating weekly VI feature embeddings through time-lag encoded cross-attention. The enhanced daily CV features are segmented into 21 sequences corresponding to seven-day windows. Each sequence is aggregated separately using the Transformer encoder into one feature, thus forming weekly CV embeddings. The Transformer encoder consists of the sinusoidal positional encoding, multi-head self-attention and feedforward network. The self-attention and feedforward network adopt the one-layer architecture. Subsequently, these features are fused with growth phase (GP) VI features from the remote sensing stream using another time-lag encoded cross-attention. The output of the cross attention is forwarded to the remote sensing stream for ultimate yield estimation.

2.1.3. Knowledge-encoded fusion

2.1.3.1. Time lag quantification. In this study, time lag is defined as the temporal offset between the strongest climate stress events and the largest extreme vegetation decline. This definition differs from the conventional notion of lag, which typically considers the delay between the onset of stress and the initial vegetative response. Instead, it aims to capture the lagged and long-lasting damage within intervals of extreme stress events and the extreme canopy decay. This strategy enables the model to focus on the burst of stress conditions and the subsequent prolonged damage period to vegetation structure, which are an effective

strategy for identifying crop responses under extreme climate stress.

The duration of the time lag is determined by the temporal position of extreme events as characterized by Killing Degree Days (KDD) and the Weighted Difference Vegetation Index (WDRVI). Drought and heat stress are among the two most critical environmental factors influencing crop growth, development, and yield processes. Climatic variables like VPD and KDD are often utilized as the drought and heat indicator respectively. Recently, heat stress has been observed with the steadily growing relative importance trend (Barnabás et al., 2008; Lobell et al., 2015; Prasad et al., 2015). KDD is a temperature-based index used to quantify damaging heat units that can negatively impact crop yields. It accumulates when daily maximum temperatures exceed a defined upper threshold T_{max} = 29 ℃, reflecting the cumulative effect of high-temperature stress (Butler and Huybers, 2015, 2013). WDRVI is linearly correlation with vegetation fraction and has high sensitivity to moderate-to-high biomass density, representing canopy greenness dynamics (Gitelson, 2004).

The normal and stress conditions are defined based on the yield deviation to historical means. 200 county-year samples are randomly selected from 2012 to represent stress condition, with yield deviation to historical means smaller than −30 %. The same 200 counties are randomly selected from 2006 to 2011 to indicate normal conditions, which have yield deviation to historical means bigger than 0. The difference time series of KDD and WDRVI, generated by subtracting normal conditions from stressed ones, reveal clear signals of extreme events. High-temperature stress is defined as KDD exceeding 3 °C·day according to the finds of harmful impact of high temperature on yield (Deryng et al., 2014; Lobell et al., 2015; Schlenker and Roberts, 2009). Significant WDRVI decay is considered to occur when surpassing 0.3 through trial and error. The durations from the start of maximum extreme climate event to the end of maximum extreme vegetation decay are regarded as influenced most by time lag periods.

The time lag quantification process involves comparing two time series representing normal and stress conditions, respectively. By calculating their difference over time, periods where the deviation exceeds a predefined threshold are identified as potential extreme events. A binary mask is generated to mark these periods, and contiguous sequences of ones in the mask, bounded by zeros, are recognized as distinct extreme events. The duration of each event is computed, and the one with the longest duration is selected as the maximum extreme event. The time lag is then defined as a tuple representing the durations of the maximum extreme events in the respective climate and vegetation time series. The detailed quantification process is as follows:

Let $T_n(t)$ and $T_s(t)$ represent the time series under normal and stress conditions, respectively. The difference between the two time series is calculated as:

$$T_d(t) = T_n(t) - T_s(t) \tag{1}$$

To identify periods of extreme events, a binary mask $M(t)$ is constructed as:

$$M(t) = \begin{cases} 1, T_d(t) > threshod \\ 0, otherwise \end{cases} \tag{2}$$

Here the *threshold* defines the criterion for identifying extreme deviations. An extreme event $EE_{i,j}$ is defined as a continuous sequence from time i to j where all values in $M(t)$ are 1, and it is bounded by *0 s* at both ends (i.e., $M(i-1) = 0$ and $M(j+1) = 0$:

$$EE_{i,j} = j - i, if M(i-1) = 0 and M(j+1) = 0 and \sum_{t=i}^{j} M(t) = j - i + 1 \tag{3}$$

Among all identified extreme events, the one with the longest duration is selected as the maximum extreme event:

$$ME = \underset{(i,j)}{argmax} EE_{i,j} \tag{4}$$

Finally, the time lag duration TL is defined as a pair (ME_C, ME_R), where ME_C and ME_R represent the maximum extreme event durations associated with climate and vegetation signals, respectively. TL represents the start and end of the time lag, which is used to mask the time region in the knowledge encoding process when modeling.

$$TL = (ME_C, ME_R) \tag{5}$$

2.1.3.2. Knowledge encoding using cross attention. The cross attention is the core components of the fusion between the two streams, with their attention weights constrained by the time lag knowledge to construct interpretable interactions (Fig. 2). Cross attention built on the self-attention mechanism is inherently capable of modeling interdependence between any two sets of embeddings across different temporal scales. This property enables the fusion of features from both weekly and daily temporal resolutions, as well as between growth phase (GP) and weekly scales. The cross attention involves attending query embeddings from one source to key and value embeddings from another, and switching the query, key and values embeddings from both sources for cross computing, thus allowing the model to align significant information across both. However, the purely data-driven nature of cross attention may overlook potential cascading relationships that hold significant biological meaning in crop growth. To address this, we extract observed time-lag signals of climate stress on crop canopy greenness dynamics from the climate stress indicators and VIs. These time-lag signals are organized into a matrix that masks the attention weights in the cross-attention process, thereby incorporating biological mechanisms into the attention calculations. This masking strategy allows the model to focus more on time-lag regions, facilitating the learning of stress-related features for improved crop yield estimation under climate stress.

Taking the time-lag encoding for weighting daily CVs as an example, the entire encoding process consists of three stages: cross-attention weight learning, weight constraining, and feature enhancement. The weighting process of weekly RS and weekly CV follows the same time-lag encoding strategy as weighting daily CVs. The only differences are the variables for query, key, values in computing cross attention and temporal length of time lag in adjusting the time lag matrix. When implementing the weighting of weekly RS, the query is daily CVs and key and values are weekly RS. The time-lag matrix is transposed from the one used in weighting daily CVs. For weighting weekly CV, the query is changed into GP RS and key and values take weekly CV for computing. The time-lag matrix reserves the same meaning of rows and columns as the weighting process of daily CVs, but shortens the temporal length using statistical aggregation to align with GP RS and weekly CV.

Cross-attention weight learning. Initially, the cross-attention weights are learned following a standard self-attention structure. It uses the features from weekly VIs as query embeddings and attends it on the key embeddings generated from daily CVs. The resulting intermediate attention map illustrates the temporal response of VIs on CVs anomaly, with each element reflecting the response at specific time steps. The detailed computing process is as follows:

$$\boldsymbol{Q} = \boldsymbol{H}_R \boldsymbol{W}_q \tag{6}$$

$$\boldsymbol{K} = \boldsymbol{H}_C \boldsymbol{W}_k \tag{7}$$

$$\boldsymbol{\alpha}_{CA} = softmax\left(\boldsymbol{Q}\boldsymbol{K}^T / \sqrt{C/n}\right) \tag{8}$$

where $\boldsymbol{H}_R$ and $\boldsymbol{H}_C$ are the hidden features of remote sensing and climate streams, $\boldsymbol{W}_q$ and $\boldsymbol{W}_k$ are learnable weight matrices, C and n are the embedding dimension and number of heads, *softmax* is the activation function, $\boldsymbol{\alpha}_{CA}$ is the cross-attention map reflecting the impact of CVs on VIs where rows represent the temporal dimension of CVs while the columns representing the VIs.

Weight constraining. Notably, elements in the lower triangle of the

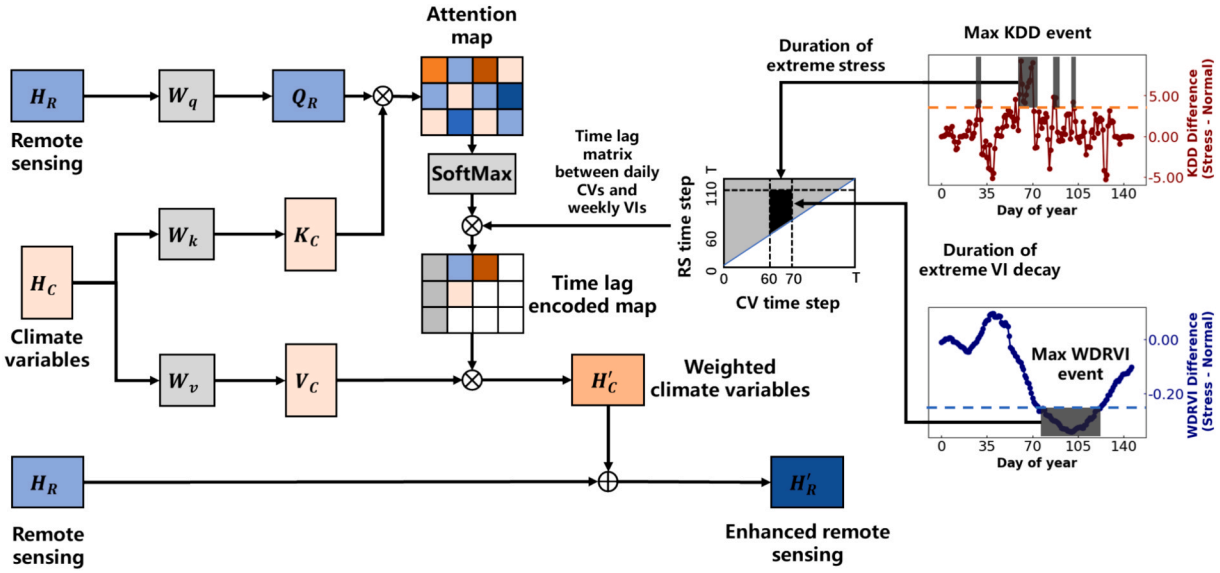

Fig. 2. Time-lag encoded fusion using cross-attention between remote sensing and climate variables. The learned attention map between remote sensing and climate variables is weighted using knowledge matrix showing their time lag effect. The knowledge matrix has values 0 for the lower triangle to represent temporal causality. The difference time series of KDD and WDRVI, generated by subtracting normal conditions from stressed ones, are used to extract extreme events for quantifying the time lag duration. High-temperature events are defined as KDD exceeding 3 °C·day, while significant greenness decay is considered to occur when WDRVI difference surpassing 0.3. The durations of extreme high-temperature event and extreme greenness decay event are projected to a two-dimensional matrix to express the time lag, with the overlay region set as 1.

map indicate the impact of late-step CVs on early-step VIs, which is biologically impossible in crop growth. To constrain the cross-attention weights, we develop a time-lag knowledge matrix with zeros in the lower triangle and non-negative values in the upper triangle. This matrix effectively constrains the attention map through multiplication, inactivating attention weights in the lower and placing varied impact of time lag on the weights in the upper. The detailed constraining process is as follows:

$$\boldsymbol{\alpha}_{TL} = \boldsymbol{\alpha}_{CA} \bullet \boldsymbol{P} \tag{9}$$

where $\boldsymbol{P}$ is the time-lag knowledge matrix, $\boldsymbol{\alpha}_{TL}$ is the time-lag-encoded attention map reflecting the delayed impact of CVs on VIs. The upper triangle of $\boldsymbol{P}$ contains values of 1 or 0 to provide varied response of VIs to CVs (Fig. 2). The time-lag regions are reshaped as the overlaps between the extreme climate event and the extreme vegetation decay in the map per sample. Samples exhibit varied length of time lags due to the varying intensities of stress, and the union of these overlaps is compiled to encompasses all time lags. This union region in the knowledge matrix assigns a value of 1 in the upper triangle, while the remaining areas in the upper triangle are assigned a much lower value of 0. A sensitivity analysis was conducted in the training set to determine this lower threshold, testing values from 0 to 0.5 with an interval of 0.1. As the RMSE differences were small across this range (Fig. S2), 0 was selected to effectively suppress non-time-lag regions and emphasize learning from time-lag-relevant features.

Feature enhancement. The knowledge matrix works as a gate on attention map to allow time-lag influenced features to pass, while partially blocking those outside the time-lag durations. The daily CVs are multiplied by the encoded attention map to generate weighted CVs with time-lag implications. These weighted CVs are summed with the VIs to produce enhanced remote sensing features, which are passed to the upper layer of the pyramid structure. The detailed enhancement process is as follows:

$$\boldsymbol{V} = \boldsymbol{H}_C \boldsymbol{W}_v \tag{10}$$

$$\boldsymbol{H}'_c = \boldsymbol{\alpha}_{TL} \boldsymbol{V} \tag{11}$$

$$\boldsymbol{H}'_R = \boldsymbol{H}'_c + \boldsymbol{H}_R \tag{12}$$

where $\boldsymbol{W}_v$ is the learnable weight matrices, $\boldsymbol{H}'_c$ is the time-lag-encoded daily CVs, $\boldsymbol{H}'_R$ is the enhanced remote sensing features using time-lag-encoded CVs.

2.1.4. Knowledge-encoded prediction

2.1.4.1. Phenology sensitivity quantification. Phenology sensitivity is measured by the correlation of principal features of drought and heat stress indicators with corn yield at each growing phase. KDD, VPD, and WDRVI are linearly projected using principal component analysis (PCA) to generate the principal features, reflecting heat stress, drought stress, and canopy chlorophyll over the growing season. The first principal component is used as the principal features as it captures the most variance in the three stress indicator. Pearson's correlation coefficients for these principal features and yield are calculated at each growth period to construct a phenology sensitivity vector.

2.1.4.2. Phenology encoding using linear attention. The growing phase features generated from time-lag encoding are constrained by phenological sensitivity, allowing the model to focus on the reproductive stages that hold the highest importance throughout the growing season. The features, aggregated by a pyramid structure within a single stream and fused through time-lag encoded cross attention between the two streams, represent the complex dynamics of crop growth (Fig. 1). Time lag effects can occur at any stage, thereby strengthening the features at the corresponding stages. However, not all phenological stages are equally important for ultimate yield. Corn yield responses to climate stress exhibit higher sensitivity during the flowering stage, with lower sensitivity observed in other stages. Therefore, embedding this phenological sensitivity in the time-lag encoded growing phase features may enhance the modeling of critical time lag impacts on ultimate yield.

The attention mechanism is employed to integrate phenological sensitivity into the growing phase features by adjusting the initially learned attention weights (Fig. 3). Linear attention mechanisms are widely used to emphasize important features based on learnable

attention weights (Mnih et al., 2014; Xu et al., 2021; Zhong et al., 2023). Many current studies adjust the distribution of these initially learned attention weights using quantified prior knowledge vectors for knowledge constraining in deep neural networks (Kim et al., 2023; Phung et al., 2024). These adjustments typically treat prior knowledge as coefficients for the attention weights. In this context, the linear attention receives growing phase features h1, …, h5 as input. A fully connected layer transforms these features into initially learned attention weights α1, …, α5. These weights are then multiplied by a predefined phenology sensitivity vector s1, …, s5 to generate sensitivity-constrained attention weights β1, …, β5, which encode the phenological characteristics under climate stress. The original growing phase features are weighed by these encoded weights to produce phenology sensitivity-constrained feature embeddings. Finally, another fully connected layer aggregates these features to obtain the ultimate yield estimate.

The proposed model utilized the mean squared error function combined with L2 norm regularization as the loss function. The Adam optimizer is used to optimize the network. The epoch is set as 300 for training. The hyperparameters consist of structural hyperparameters including the embedding dimension and training hyperparameters like the learning rate. The range of hyperparameter values is pre-defined and optimal ones are determined through grid search. The search space and optimal hyperparameter configuration are provided in Table S1. The computation and analysis are performed on a Linux workstation (Ubuntu 16.04 LTS) with two Intel Xeon E5-2683 v4 Processors (2.1G/16 Core/40 M), 128 GB of RAM, and one NVIDIA Tesla V100 graphics card (32 GB of memory).

2.2. Datasets and data preprocessing

To validate the knowledge-encoded yield prediction framework, this study selects the nine Midwestern states within the U.S. Corn Belt from 2006 to 2023 for case studies. The region in the year 2012 experienced an unprecedented and widespread extreme drought and heat event, resulting in significant yield reductions and a scarcity of samples for studying potential future extreme events. Validating the effectiveness of the proposed knowledge-encoded model becomes particularly valuable in such a scenario with limited data. These states primarily cultivate rain-fed corn, with yield records collected from 769 counties via the USDA's National Agricultural Statistics Service (NASS) database (USDA-NASS, 2020a). Only counties with yield records spanning more than five years were retained in the dataset to ensure the representativeness of yield samples over long temporal durations. The Crop Data Layer (CDL) provides corn maps for focusing on the corn area and generating corn-relevant weekly VIs and daily CVs for yield estimation (USDA-NASS, 2020b). The masking process follows previous work (Johnson, 2014; Lin et al., 2020). Each county intersects with each year to establish the county-year entries, which were paired with daily CVs and weekly VIs to construct a ready-to-use dataset. The temporal length of input series covering from the 18th to the 38th week of each calendar year to encompass the entire growing season across various years and regions.

Weekly VIs are generated from the daily bands of the MCD43A3 product, temporally aggregated into weekly time steps to reduce missing data (Schaaf and Wang, 2015). The MODIS MCD43A3 product provides optical bands at a spatial resolution of 500 m. Weekly VIs are further spatially aggregated at the county level using a median operation to match yield records. Linear interpolation and the Savitzky-Golay filtering approach were employed to fill narrow data gaps caused by clouds and cloud shadows in the weekly time series. Daily CVs are calculated using meteorological data from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) dataset (Daly et al., 2008). The meteorological data are aggregated to county levels by averaging. Cumulative VPD and PRCP were calculated from daily precipitation and daily VPD, respectively. The calculation of GDD and KDD followed the methodologies of previous studies (Butler and Huybers, 2015). Ultimately, 5,383 county-year records coupled with weekly VIs and daily CVs were obtained for model building and evaluation.

2.3. Baselines and performance evaluation

In this study, three widely used models are selected as baselines: Random Forest (RF), Long Short-Term Memory (LSTM), and

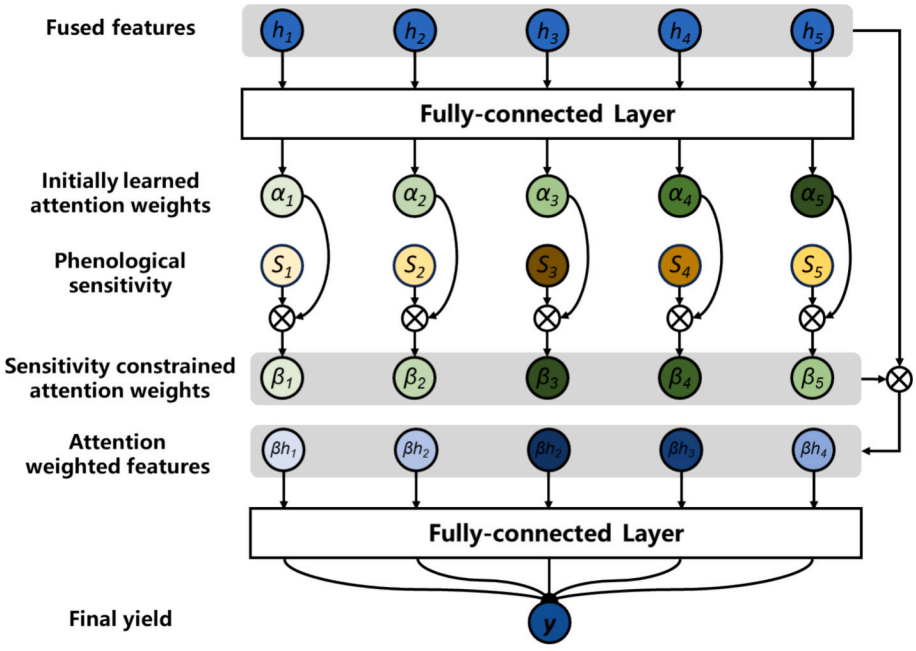

Fig. 3. Encoding phenology sensitivity to enhance the time-lag features for predicting final yield. The process operates on top of the two-stream framework, where the fused features are encoded by time lag and generated from the bottom up within the framework. Phenology sensitivity is generated by the correlation of corn yield with principal features of KDD, VPD, and WDRVI. KDD, VPD, and WDRVI respectively reflect heat stress, drought stress, and canopy greenness over the growing season. The first principal component of the three indicators is calculated using PCA for measuring the correlation. Pearson's correlation coefficients are calculated at each growth period to construct a phenology sensitivity vector.

Transformer. RF is effective at handling high-dimensional data by treating each time step as a variable in time series tasks. The weekly VIs and daily CVs are concatenated to form a high-dimensional vector for RF. LSTM and Transformer are two representative sequential models inherently designed for temporal feature extraction. The standard LSTM network serves as the backbone to build the LSTM baseline. The Transformer backbone consists of a linear layer that projects the input variables to high dimensional features and an encoder of the original Transformer that processes output of linear layer. The encoder is composed of sinusoidal positional encoding, multi-head self-attention and feedforward network. These backbones employ a two-stream structure to utilize VIs and CVs separately, named TS-Transformer and TS-LSTM. Feature fusion occurs in the last layer of the model by concatenating the high-level features from the two data sources, following the same fusion strategy as previous studies (Huang et al., 2022; Peng et al., 2024). The evaluation criteria used are the Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2) between the estimated and actual yields. The training strategy and computation platform were the same as the proposed model. The search space and optimal hyperparameter configuration of the three baselines are provided in Tables S2–S4.

Five experiments are conducted to demonstrate the effectiveness of the proposed model and address the research questions summarized in Section 1 (Fig. 4). We simulate the actual situation where training samples only acquired from historical years to examine model performance under extreme stress conditions. Counties from 2012 are selected for testing, while samples from 2006 to 2011 were used for model training. Furthermore, counties in 2013–2023 are also used for assessing the scalability of the model over spatial and temporal domain.

In the first experiment, we analyzed the time lag durations and phenological sensitivity for samples experiencing climate stress. For time lag analysis, we visualized the time lag patterns of the data from 2006 to 2012 using the KDD and WDRVI time series. KDD indicates high temperature stress, while WDRVI represents the dynamics of canopy greenness in corn plants. The time lag durations quantified by VPD and NDWI are also analyzed, as VPD represents the drought severity and NDWI indicates the water status of corn plants. The time lag durations of samples in 2006–2011 are extracted for model training while the ones extracted from 2012 are used for testing. To assess the phenological sensitivity of these durations on corn yield, we analyzed the correlation between yield and a linear combination of KDD, VPD, and WDRVI on sample suffering from stress in 2006–2011.

The second experiment involves accuracy comparisons for evaluating the performance of the proposed model. The general performance is assessed by comparison between the proposed knowledge-encoded model and baseline models. To highlight the importance of each component, an ablation analysis is employed to assess the contributions of time lag knowledge, cross attention, and the two-stream backbone within the knowledge-encoded model. The two-stream backbone that combines the features using a direct sum of features of two streams serves as the baseline in the ablation analysis. In the third experiment, we validate the interpretability of the model by visualizing the attention map on the sample with time lag signals in 2012, confirming whether detailed time lag signals are effectively learned from the time series data. Additionally, the phenological sensitivity encoded through linear attention on top of the two-stream structure is visualized to analyze the effects of this sensitivity knowledge.

In the fourth experiment, we simulate in-season corn yield prediction based on the established dataset to further validate the proposed knowledge-encoded model. Performing real-world in-season predictions is challenging due to limitations in the in-season prediction algorithms and difficulties in obtaining accurate in-season crop maps and

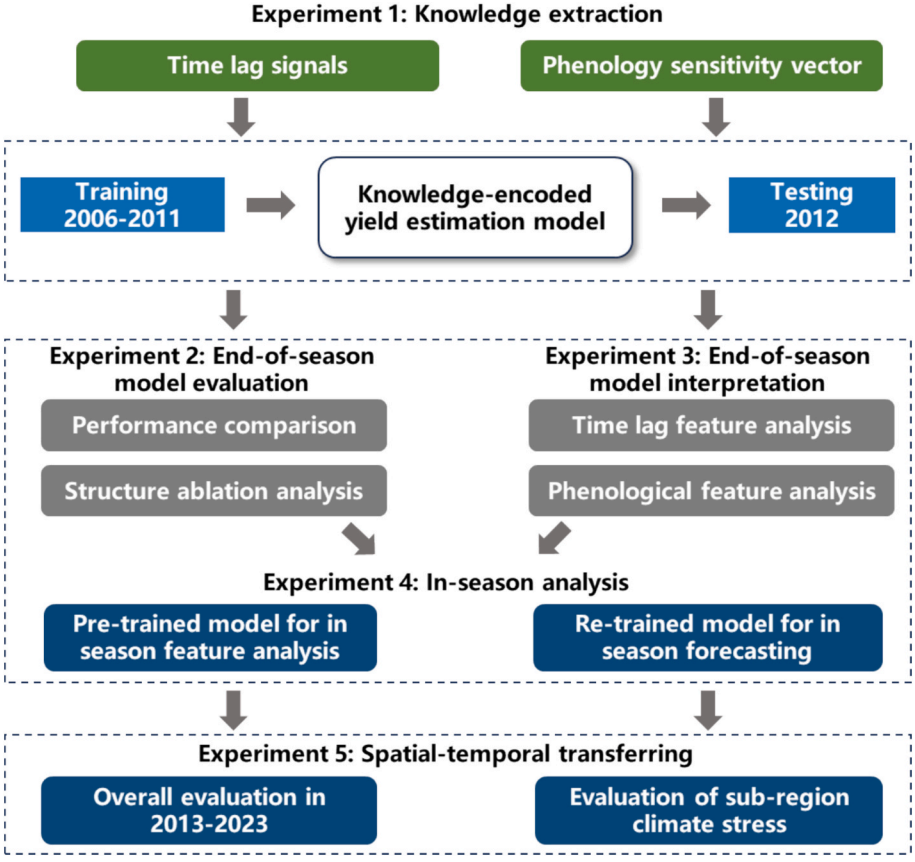

Fig. 4. Experimental design to validate the effectiveness of knowledge encoded strategy in end-of-season, in-season, and spatial–temporal transferring performance of the proposed model.

constructing high-quality input time series. To gain insight into the in-season potential of proposed knowledge-encoding strategy, we set aside the preprocessing challenges inherent in real-world situations and regard this dataset as an optimally prepared resource. This allows us to focus on testing the capabilities of the proposed knowledge-encoding models for in-season yield estimation.

The experiment is conducted at various yield anomaly levels to demonstrate the capability of proposed model across varied degrees of climate stress. The yield anomaly is calculated as the percentage change of yield in 2012 relative to the historical average yield from 2006 to 2011. The yield anomaly is then divided into four quartiles, with each group containing an equal number of samples. The first quartile represents the lowest yield loss that is closer to yield distribution in normal years, while the fourth quartile corresponds to the highest yield loss. The pre-trained end-of-season model, combined with the knowledge matrix and historical mean-masked in-season time series, is first used to validate the in-season feature extraction behavior. The later time steps of the full-season input time series are masked using historical mean values to capture the yield response of earlier time steps. We then compare these predictions across different yield anomaly levels to reveal how the knowledge matrix enhances model prediction behavior. In the second part of this experiment, the proposed model structure is re-trained using historical in-season time series to predict ultimate yield. The dynamics of prediction performance across different yield anomaly levels could indicate the occurring time of the largest accuracy improvements and optimal lead time for various degrees of climate stress.

In the final experiment, we deeply conduct the spatial–temporal transferring experiment within the US Corn Belt in 2013–2023, especially geographically diverse regions with sub-regional extreme climate stress. The proposed model, pre-trained using data from 2006 to 2011, is evaluated on all counties in 2013–2023. Furthermore, we validate the model performance in counties with severe yield loss related to extreme sub-region climate stress. As corn yield is significantly influenced by climate stress during flowering stages, the target counties are those having large KDD anomalies within 55–80 days from 18th week, which is also the main duration that caused the severe yield loss in 2012. The KDD anomalies are quantified using the percent deviation to historical averages of total KDD. The spatial distribution of the KDD anomalies during this period shows similar patterns of yield loss, which is represented by detrended yield calculated from linearly fitted trend from 2006 to 2023 (**Fig. S5**). The counties with top 10 % yield loss and KDD anomalies exceeding 30 % compared to the historical average are selected for evaluating the proposed model in sub-region extreme climate stress.

3. Results

3.1. Prior knowledge under climate stress

Large gaps are observed in both KDD and WDRVI under stress and normal conditions, though the temporal positions of the largest gaps differ between these two indicators (Fig. 5). The mean time series of daily KDD, averaged over the study area for the historical year, is visualized alongside the daily KDD for 2012, using the same averaging method. The daily mean KDD in normal conditions shows significant differences compared to stress conditions, particularly during the period ranging from the Day 60 to Day 70. The difference values in KDD between stress and normal conditions reaches nearly 10.00 ℃ days during this period, indicating that an extreme high-temperature event occurred. Similarly, the mean time series of WDRVI for both climate conditions is generated using the same aggregation method as KDD. A substantial reduction of WDRVI between normal conditions and stress conditions begins around the 60th day, coinciding with the onset of high-temperature stress. This reduction continues from the Day 60 to approximately Day 105 to reach its peak. The similar large gaps and their temporal positions are also observed in both VPD and NDWI (Fig. S1).

Time lags exist between the extreme high-temperature event and the maximum reduction in crop greenness as shown in the difference time series (Fig. 5c). The difference time series of KDD and WDRVI represents the contrast between stress and normal conditions. The lowest trough of the WDRVI difference lags the highest peak of the KDD difference by around 45 days, with the WDRVI trough occurring around Day 105 and the temperature peak starting from the Day 60. The similar time lag phenomena caused by extreme climate stress is also observed in summer corn in North China Plain (Liu et al., 2018) and in the global cropland (Wu et al., 2015). Day 60 aligns with the beginning of the silking stages when stress has the most significant impact on ultimate yield throughout the growing season. It may explain the severe yield loss experienced in the U.S. Corn Belt in 2012 as extensively studied in previous research (Butler and Huybers, 2015; Challinor et al., 2014). The prolonged reduction of WDRVI from the Day 60 to the Day 105 indicates a lasting and time lagged decline in canopy greenness. This decline reflects reduced chlorophyll content and fewer leaves on corn plants, limiting photosynthesis for generating the organic matter that determines ultimate yield. Thus, the extended duration and time lags might present the inevitable negative impacts of climate stress on ultimate yield, which crops cannot overcome through recovery. The similar time lags between the extreme high-VPD event and the maximum reduction in NDWI are also captured in the VPD and NDWI difference time series (Fig. S1(c)).

The climate stress indicator and vegetation index exhibit the highest

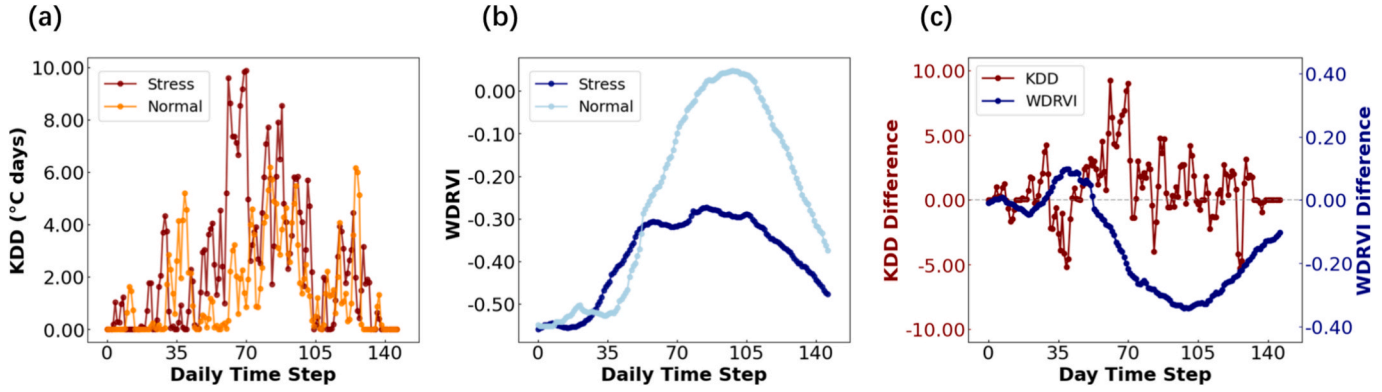

Fig. 5. Temporal profiles of KDD and WDRVI and their difference under stress and normal condition. (a) Mean time series of KDD in stress and normal conditions. 200 samples of the stress condition are randomly selected from 2012 with percent yield deviation to historical means smaller than −30%. 200 samples of the normal condition, the same counties as in 2012, are randomly selected from 2006 to 2011 with yield deviation to historical means bigger than 0. (b) Mean time series of WDRVI in stress and normal conditions. (c) Differences between indicators under stress condition and under normal condition.

correlation with yield during GP3 and GP4, respectively, throughout the growing season (Fig. 6). The crop progress report from the USDA outlines the following growth periods: GP1 represents the duration from corn planting to emergence; GP2 covers the period from emergence to three weeks before silking; GP3 spans three weeks before silking to the milk stage; GP4 extends from milk to dent; and GP5 lasts from dent to maturity. KDD and VPD show the strongest negative correlation with corn yield in GP3. This indicates that heat and drought stress occurring during the silking stage have the worst impact on yield. The milk stage (GP4) ranks second in terms of negative effects on yield, as it displays the second-largest negative correlation. WDRVI demonstrates the highest positive correlation with corn yield in GP4, with a slightly lower correlation in GP3. This underscores the importance of crop canopy greenness in the grain-filling process for yield formation.

Phenology sensitivity measured by the correlation of principal features with corn yield is highest in GP3, followed by GP4 (Fig. 6d). KDD, VPD, and WDRVI are linearly projected using principal component analysis (PCA) to generate the principal features, reflecting heat stress, drought stress, and canopy chlorophyll over the growing season. Pearson's correlation coefficients for these principal features and yield are calculated at each growth period to construct a phenology sensitivity vector. The profile of phenology sensitivity resembles a hill with GP3 as the peak and other GPs descending, aligning with previous field experimental results on stress impacts on reproductive processes in crops (Barnabás et al., 2008). GP4 is identified as the second most sensitive stage after GP3, emphasizing the critical role of chlorophyll in photosynthesis for producing grain organic matter.

3.2. Yield estimation performance

The accuracy of the knowledge-encoded two-stream framework significantly surpasses that of the baselines (Fig. 7). The proposed model achieves the best end-of-season yield estimation for 2012, with a RMSE of 1.17 Mg ha^{-1} and a R^2 of 0.78. The other two two-stream neural networks follow the proposed model, with RMSE values of 1.43 Mg ha^{-1} and 1.54 Mg ha^{-1}, and R^2 values of 0.67 and 0.62, respectively. The superior performance of the proposed model compared to the other two-stream models highlights the advantages of knowledge-encoded fusion over previous simple concatenation of separate data sources in extracting features. In contrast, the random forest model (RF) underperforms the three two-stream networks, achieving an RMSE of 1.68 Mg ha^{-1} and an R^2 of 0.54. This indicates that treating each time step as an independent variable is insufficient for processing multivariate time series data characterized by complex temporal interactions.

KETS generally exhibits lower absolute errors across the entire region, with different spatial patterns of error distribution in northern and southern (Fig. 8). Baseline models show notable overestimation in the southern regions, where extreme heat and drought were most pronounced. By contrast, KETS significantly reduces this overestimation, indicating better performance under extreme climate conditions. In the northern regions where the model is prone to underestimation, the KETS achieve comparable performance with the baselines. In several western counties where TS-LSTM tends to significantly underestimate yield, KETS produces more accurate estimates with smaller underestimation errors. These results highlight the advantages of KETS in capturing spatial heterogeneity of crop responses under climate stress and reducing regional biases in yield estimation.

Incorporating cross attention into the two-stream framework significantly enhances model accuracy, while embedding prior knowledge also contributes to improvements in accuracy (Fig. 9). The Two-Stream backbone fusing features using a direct feature sum, yield poor accuracy with an RMSE of 1.97 Mg ha^{-1} and an R^2 of 0.37. It suggests that a simple sum does not effectively align features from different sources. In contrast, the incorporation of cross attention into the Two-Stream backbone learns the attention weights of both VIs and CVs for a weighted sum. This approach significantly enhances performance, achieving a lower RMSE of 1.36 Mg ha^{-1} and a higher R^2 of 0.70. The substantial improvement in performance demonstrates the effectiveness of the cross-attention mechanism in aligning features from diverse sources and modeling their interactions. Furthermore, the knowledge-encoded cross attention outperforms the pure cross attention implementation, with RMSE improved from 1.36 Mg ha^{-1} to 1.17 Mg ha^{-1} and R^2 increased from 0.70 to 0.78. The interaction modeling capability of vanilla cross attention derived from natural language processing,

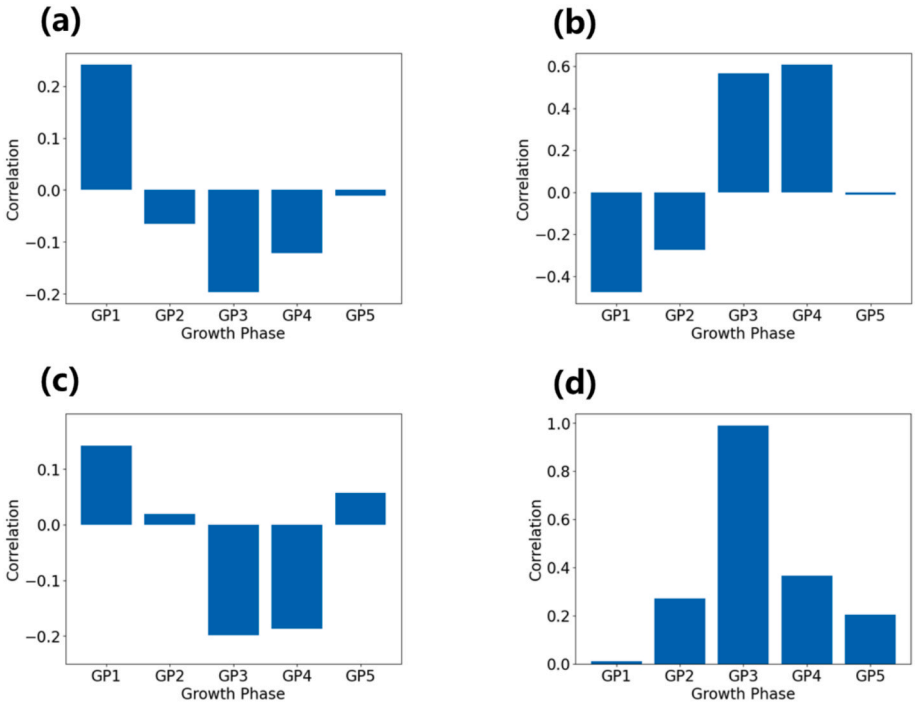

Fig. 6. Phenology sensitivity quantified from historical stress samples in 2006–2011. (a) KDD; (b) WDRVI; (c) VPD; (d) Phenology sensitivity of yield on principal features over all GPs. Principal features are the first principal component of PCA reduced features from KDD, VPD and WDRVI.

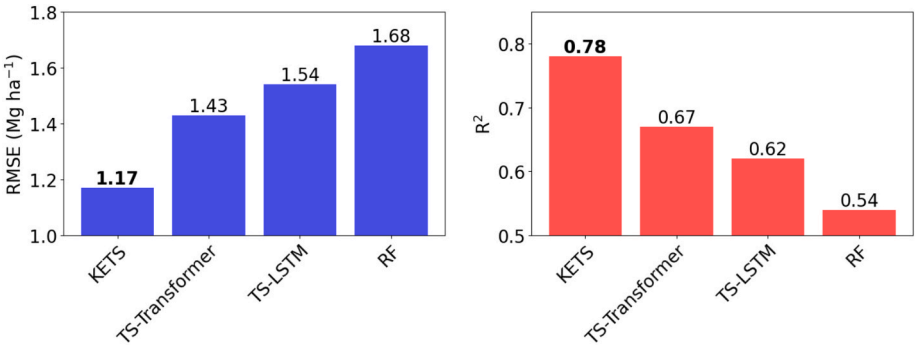

Fig. 7. The end-of-season performance of the proposed model compared to the baselines. KETS" means the knowledge-encoded two-stream framework.

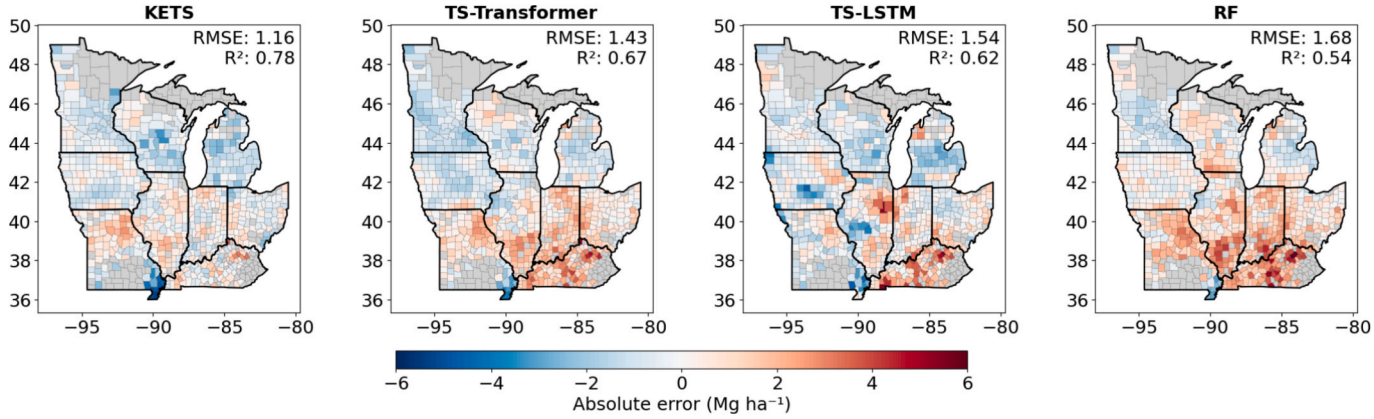

Fig. 8. The spatial distribution of the absolute yield error of the proposed model compared to the baselines in 2012.

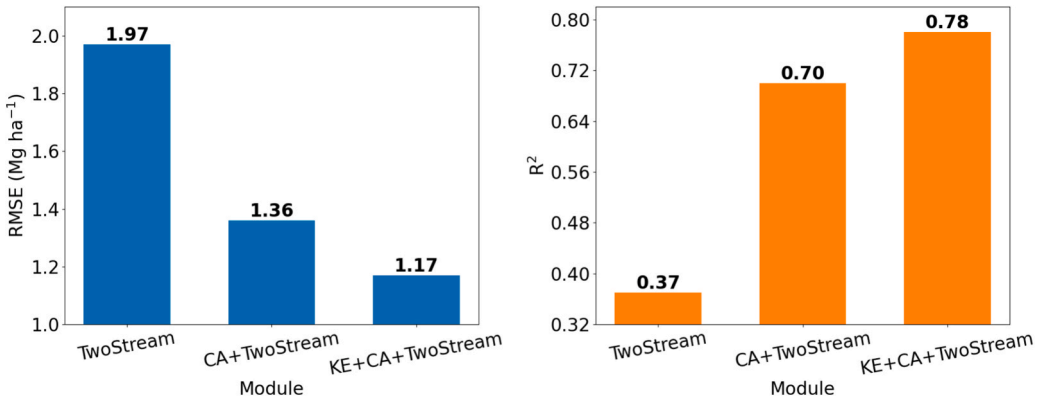

Fig. 9. The ablation results of different main parts of proposed knowledge encoded two-stream framework. "CA" means cross attention. "KE" means knowledge encoded.

typically overlook the temporal causal relationships between time steps. The results indicate the temporal causation brought by the knowledge encoding could improve performance compared to pure cross attention modeling.

3.3. Model interpretation through cross-attention map and linear attention weights

The knowledge-encoded cross attention map shows time lag signals after the flowering stage, while the vanilla cross attention distribution exhibits some randomness (Fig. 10). The encoded weights for weighting daily CVs show hot regions after the silking and dent stages, representing the time lag response of VIs to CVs. The weights for weighting weekly CVs also display the highest time lag response after the silking stage, aligning with the distribution observed in the input time series shown in Fig. 5. The more compact time lag patterns for weekly CVs compared to daily CVs may stem from the pyramid structure for high level feature aggregation, resulting in less noise. The cross-attention weights for weighting VIs differ from those on CVs, with CVs having a higher influence on VIs from early emergence to the middle silking stages. These results indicate that the knowledge encoded cross attention captures the temporal dynamics during this period, where VIs vary significantly along the growing season and are sensitive to environmental perturbations. In contrast, the vanilla cross attention weights show a random distribution on daily CVs and weekly VIs, retaining some time lag patterns on GP CVs during the planting stages, which are less

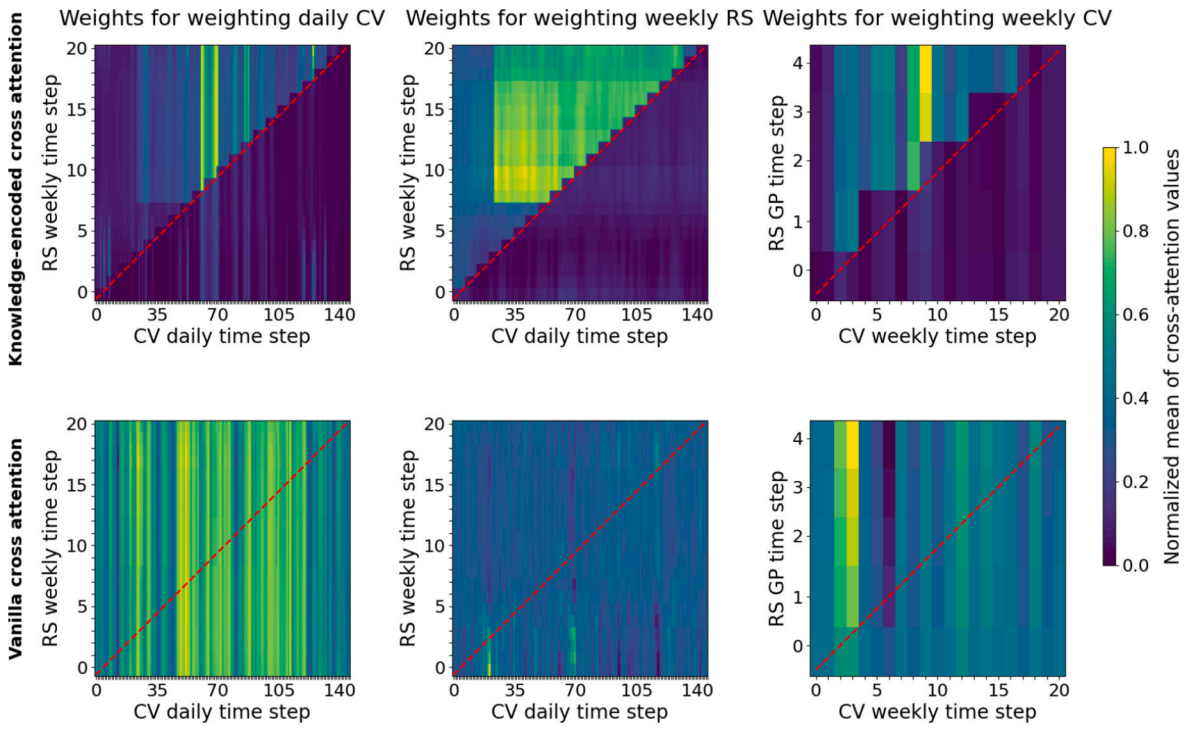

Fig. 10. The knowledge encoded and pure data-driven cross attention map of cross attention on daily CVs, weekly VIs, and GP CVs with sample having time lag signals in 2012.

critical for yield compared to the silking stage. The results demonstrate that cross attention enables the two-stream model to learn patterns from separate data sources, while the knowledge-encoded attention weights can direct these patterns, emphasizing the importance of incorporating knowledge to guide the model's focus on critical time lag patterns during key stages.

The phenological sensitivity encoded in the attention maintains its highest value during the silking to dent stages, compared to the highest value in maturity stage observed without attention constraints (Fig. 11). The knowledge-encoded attention weights increase from GP1, peaking at GP4, followed by a decline at GP5. This distribution indicates that phenological sensitivity effectively guides the linear attention to focus on the silking to dent stages, thereby strengthening the time lag-encoded features. In contrast, the other two data-driven linear attention mechanisms exhibit a monotonically increasing trend with varying magnitudes across each GP. The two-stream model shows gradually increasing attention weights along the GPs, while the cross attention two-stream model exhibits a more pronounced upward trend, with GP5 significantly higher than the other four GPs. The increasing trend of the data-driven linear attention weights has also been documented in previous data driven studies (Lin et al., 2020), which indicates that without any prior knowledge constraints, the model would struggle to capture the biophysical meaningful features during the silking to dent stages.

3.4. *In-season prediction based on time lag*

The pre-trained KETS continues to predict lower yields for samples suffering from stress, while predictions for normal samples remain unchanged (Fig. 12). Prior to the 60th day, predictions on both stressed and normal samples show an increasing value to their peaks. This phenomenon may be attributed to the model's ability to extract more effective features as real data accumulates, resulting in higher yield forecasts. Following the 60th day when stress occurred, the model predicts increasingly lower yields for stressed samples, while the predicted yield

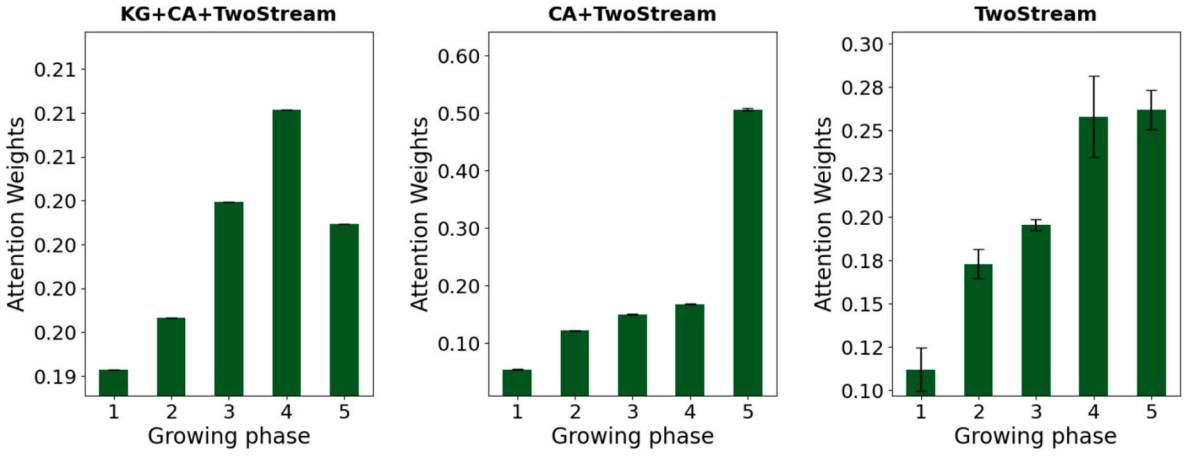

Fig. 11. The linear attention weights of knowledge encoded two stream model, cross attention based two stream model, and the two stream model.

for normal samples remains at its peak. The decline of predicted yield in stressed samples aligns well with the occurrence of extreme high-temperature drought stress, indicating that the time-lagged model effectively captures stress signals to enhance yield estimation. Furthermore, across various yield anomaly levels, greater yield anomalies correlate with lower predicted yield values, demonstrating the model's capacity to detect differences in yield loss due to stress.

In scenarios of extreme yield reduction, the model employing the lag effect matrix achieves lower prediction errors compared to the model without this feature (Fig. 12). Before the 60th day, both models experienced increasing prediction errors as extreme stress has not yet occurred, and predicted yields are closer to normal conditions. After the 60th day, the model with the lag effect matrix begins to reduce estimation errors immediately, whereas the model without this effect only starts to lower errors around the 80th day. This finding suggests that incorporating lag effects enables the model to respond more swiftly to stress, which is advantageous for mid-season yield estimation.

The lag effect significantly improves the accuracy starting from the flowering period at yield anomaly levels of 2nd-4th quartile, allowing for yield predictions up to 8 weeks in advance (Fig. 13). From the end of GP2 to the end of GP3, the In-season yield estimation accuracy improved the most, and this period coincided with the beginning of the extreme event lag. From the end of the flowering period, the mid-season yield estimation accuracy of the yield estimation model reached a high level, and the end-of-season yield could be predicted 8 weeks in advance. The lowest yield anomaly level has low in-season yield prediction performance of all the GPs, but showing a slight decreasing trend in the performance, with RMSE rising from 1.40 to 1.60 Mg ha^{-1}. This might be related to closer feature distribution to the normal samples in 0–25 % yield anomaly level while the knowledge encoded model is prone to handling the sample suffering from extreme stress.

3.5. Spatial-temporal transferring

The KETS model outperforms all baseline models across counties over the years of 2013–2023, including the recently developed DDCN (Fig. 14). KETS achieves the lowest overall RMSE of 1.34 Mg ha^{-1}, followed by closely by TS-Transformer (1.35 Mg ha^{-1}) and TS-LSTM (1.37 Mg ha^{-1}). These three two-stream models exhibit comparable performance in counties during normal years, with KETS performing slightly better across most scenarios. In contrast, DDCN and RF show the lowest accuracies, with RMSEs of 1.41 Mg ha^{-1} and 1.47 Ma ha^{-1}, respectively. The higher accuracy of the two-stream models over DDCN and RF highlights the effectiveness of modeling remote sensing and climate data separately and fusing them at deeper feature levels in normal years. The relatively poor performance of DDCN, despite its advanced hierarchical structure, suggests that unified feature encoding without explicit variable separation may limit the model's ability to disentangle modality-specific signals.

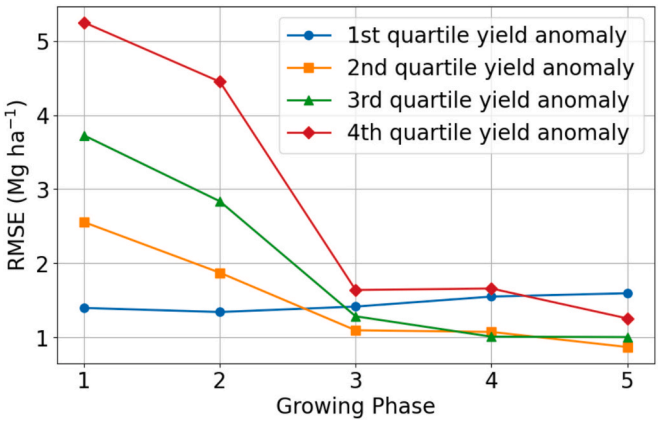

Fig. 13. The in-season prediction performance of the re-trained model among the four yield anomaly levels. The yield anomaly is the percentage change of yield in 2012 to the historical average yield in 2006–2011. The four quartiles of yield anomaly contain an equal number of samples, with the first quartile indicating the lowest yield loss and the fourth quartile indicating the highest yield loss.

KETS outperformed all baseline models in counties affected by extreme heat stress and maintained strong performance across most years from 2013 to 2023 (Fig. 15). These counties were geographically diverse and distributed throughout the US Corn Belt, covering a wide temporal span. KETS achieved the lowest overall RMSE of 1.05 Mg ha^{-1}, followed by TS-Transformer (1.15 Mg ha^{-1}) and RF (1.20 Mg ha^{-1}). The DDCN performed comparably to RF with RMSE of 1.23 Mg ha^{-1}. The TS-LSTM exhibited the highest RMSE of 1.78 Mg ha^{-1}. The results demonstrate the effectiveness of explicitly encoding time lag and phenology sensitivity for yield estimation under extreme stress conditions. In the yearly evaluation, KETS outperformed the baselines in most years, except for 2020 and 2021. In contrast, the TS-LSTM performed very well in these two years, while it underperformed significantly in 2013, 2018, and 2023. The stressed counties in 2020 and 2021 were primarily located in the northern Corn Belt, while those in 2018 and 2023 were concentrated in the southern regions. This contrast may be attributed to latitudinal differences in stress conditions.

4. Discussion

4.1. Performance across varying extreme levels

The model accuracy exhibits distinct patterns across varying levels of

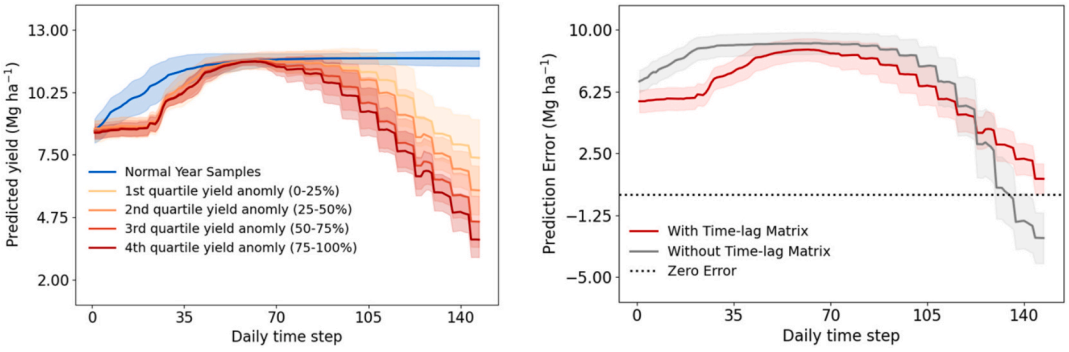

Fig. 12. The model predictions based on in-season time series generated using the historical mean mask. (a) Prediction using the time lag matrix on normal samples compared to various yield anomaly samples. The yield anomaly is calculated as the percentage change of yield in 2012 relative to the historical average yield from 2006 to 2011. The four quartiles of yield anomaly contain an equal number of samples. The first quartile indicates the lowest yield loss, while the fourth quartile indicates the highest yield loss. (b) Prediction errors with and without the time lag matrix for yield anomaly samples in the fourth quartile.

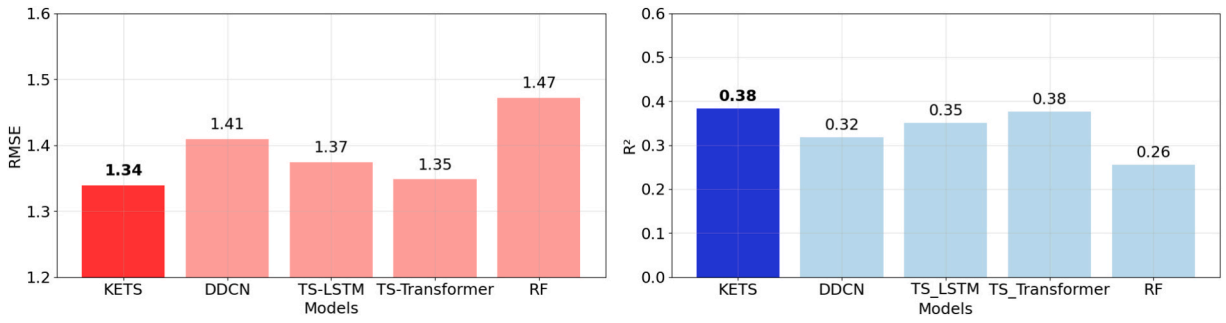

Fig. 14. The overall performance of KETS over all counties in 2013–2023 compared to baselines.

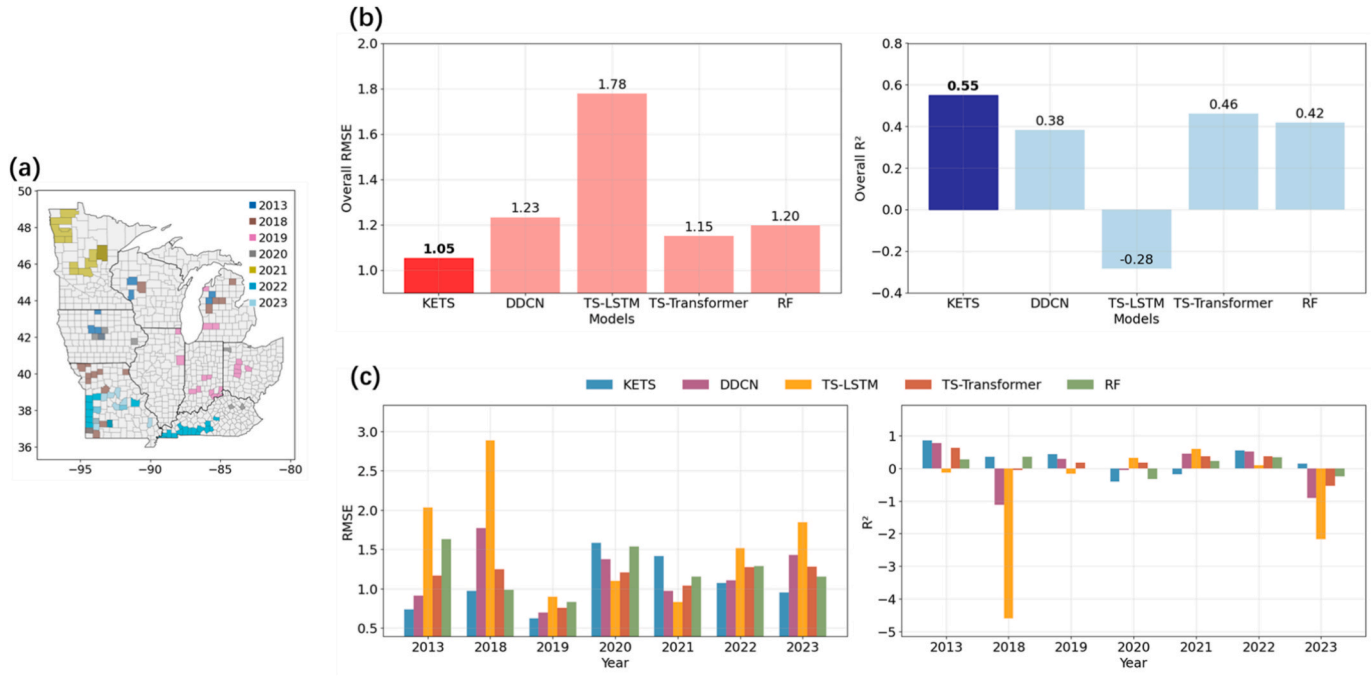

Fig. 15. Model performance in counties experienced severe yield loss due to extreme sub-regional heat stress. (a) Spatial distribution of counties with significant yield loss attributed to extreme heat stress. (b) Model accuracy in these heat-stressed counties. (c) Year-by-year performance of different models in heat-stressed counties.

extreme climate stress in 2012 which experienced extreme climate event (**Fig. S3**). All counties in 2012 were included in the analysis, and the samples were divided into four bins based on the quartiles of the daily average KDD to represent different stress levels. The daily average KDD is quantified during the flowering period (55–80 days from the 18th week). The results show that the model maintained relatively high accuracy under KDD values of 3 ~ 5 ℃·day, which represent the moderate levels of extreme heat stress. This demonstrates the robustness of the model under extreme conditions. However, performance declined under the most severe stress levels, likely due to crop failure that disrupted the extraction of time-lag signals. Future improvements might require other advanced measurement of stress-induced signals and develop dedicated modules to more effectively capture them.

4.2. Spatial and temporal transferability

The performance of the proposed KETS model exhibits distinct spatiotemporal patterns across regions and years. The spatial distribution of the training and testing sets (**Figs. S4 and S5**) shows that counties experiencing extreme stress and yield loss were mainly located in the southern to central regions of the U.S. Corn Belt in 2010 and 2011, with a few counties in the northern Corn Belt in 2006 and 2007. KETS achieved superior performance in the central Corn Belt but showed relatively lower accuracy in the northernmost and southernmost counties (Fig. 8). Similar latitudinal patterns were observed in Fig. 15, where KETS attained the highest accuracy in southern and central counties but lower accuracy compared with baselines in northern counties in 2021. This suggests that KETS may be more prone to capturing patterns in major southern and central regions, which differ phenologically from minor northern regions. Since time lags are extracted from raw time series, shifts in phenological timing between northern and southern areas may result in different temporal patterns.

Temporally, the model performs well in most years but fails in a few years influenced by latitudinal differences (Fig. 15). Years characterized by a higher proportion of southern counties, such as 2018, 2022, and 2023, generally exhibit better KETS performance. In contrast, years dominated by northwestern counties, particularly 2020 and 2021, show reduced accuracy. These declines are likely associated with mismatches in crop phenology, where later planting dates, delayed vegetative development, and shorter growing seasons in northern regions introduce temporal shifts that challenge the model's ability to generalize (Abendroth et al., 2021; Yang et al., 2025; Zhou et al., 2024).

Both spatial and temporal analyses reveal limitations of KETS along the latitudinal dimension, with reduced performance in geographically

diverse regions remaining an open question and a critical area for future research. Future extensions of the KETS framework could incorporate spatial prior knowledge into the cross-attention mechanism to better account for regional phenological variability. Introducing spatially aware modules may help mitigate biases in time-lag extraction and phenology sensitivity modeling arising from ecological differences across latitude zones.

4.3. Application for other stress-induced signals

The proposed method of fusing prior knowledge can be extended to various climate stresses on different vegetations. Existing studies have revealed the time lag effects of other climate stresses on different crops (Jiang et al., 2025; Liu et al., 2018). The cross-attention mechanism in KETS is modular and flexible, allowing it to integrate time-lag signals between climate variables and vegetation responses regardless of the stressor types or the target crop. By converting these time lag effects into numerical matrices representing their interactive effects, the proposed framework could be used to encode the time lag effects of different climate stresses (such as floods and frost) on different crops. In a larger ecosystem range, natural vegetation such as forests have also been remotely sensed to have significant time lag response to climate stress (Dou et al., 2024; Gao et al., 2025; Liu et al., 2016; Pastor-Guzman et al., 2018; Wu et al., 2015). The method proposed in this paper could also be applied to the encoding of the delayed response of different vegetation in the ecosystem to climate dynamics, which could be used for different modeling and estimation tasks in the ecosystem.

However, the extraction of time-lag knowledge from raw time series still requires crop-specific domain expertise. Different crops exhibit distinct phenological stages and physiological responses to environmental stress, which may manifest as different lag structures in the remote sensing and climate signals. For instance, the 45-day time lag used in this study is tailored to heat stress in maize (Fig. 5), and may not directly apply to crops like wheat, where sensitivity to heat often peaks during grain filling (Lobell et al., 2015; Zampieri et al., 2017). Extending KETS to other crops would require careful definition of stress indicators and time-lag intervals based on the physiological characteristics of each crop. This may involve domain knowledge or automated extraction techniques. In future work, the model could involve adaptive modules that can learn crop-specific or stress-specific time-lag patterns directly from data, enabling more generalizable applications across a broader range of agricultural systems.

In addition to the time lag effect, which is the temporal response of vegetation to the environmental drivers, this knowledge-encoded multi-source data fusion method can also adapt to other interactive modeling of vegetation and the environment. As long as the interactive effects of different data sources can be quantified as numerical vectors, the cross-attention mechanism can match their mutual effects to promote their embedding. This flexibility of the proposed knowledge encoding method paves the way for future studies exploring the interactions between different environmental variables and their collective impact on agricultural systems.

4.4. Cross attention for embedding knowledge

As a method that empowers learnable weighting between variables, cross attention can be used as a model-agnostic basic component to combine crop mechanism knowledge with multi-source data. At the county level, data-driven models have gradually become the main method for crop yield estimation in recent years (Butler and Huybers, 2015; Ma et al., 2021; You et al., 2017). The lack of yield labels for random events such as extreme stress makes it difficult for data-driven models to make full use of their feature learning advantages. Cross-attention can learn patterns directly from data while allowing attention weights to be adjusted through prior knowledge. This ability is useful in extreme stress scenarios because it can finish partial feature expression by embedding knowledge, thereby reducing the reliance on learning features using data. At the same time, when the number of yield labels increases, cross-attention can use its pure data-driven ability to learn complex patterns from vast amounts of remote sensing observations and climate records, ensuring its adaptability to varied data volume.

At the field scale, the typical method characterizing the impact of stress on crop growth is to multiply the stress impact coefficient to a specific process equation in the crop model. This characterization is simplified and requires several years of field experiments for parameter improvement and correction, which is labor-intensive and time-consuming. As a learnable weighting method, cross attention is similar to the weighting strategy in the crop model. It can be used as an alternative to automatically learning weights from measured data for weighting the existing process in crop model, reducing the huge workload of manual parameter adjustment. Cross attention can also fuse data on different time scales. This feature enables us to fuse weekly and monthly remote sensing observations and daily state variables of crop model, which offer a new method for making full use of remote sensing data.

Cross attention not only supports simple pairwise fusion, but it can also accommodate the integration of three or more data sources, thereby promoting a more comprehensive modeling of complex interactions in crop growth. The crop interacts with soil content, atmospheric conditions, and human practices from seedling to maturity. Understanding the complex interactions between impacts of various environmental factors facilitate effective regional-specific decision making. Remote sensing technology provides effective tools for retrieving the environmental variables to monitor the interactions in crop production systems (Shuai and Basso, 2022). Cross attention could be used as a basic unit to model the multivariate relationship among these remotely sensed soil, climate, and management variables through the multiple cascading in a hierarchical manner. That is, cross attention weighted features from two variables are fused with the third variable using another cross attention. This strategy has been demonstrated in biological area where researchers employ the cascading attention calculations to model three or more multi-source signals for protein folding (Abramson et al., 2024; Jumper et al., 2021).

4.5. Limitations and future work

This study validates the feasibility of cross attention to encoding stress-related knowledge using several publicly accessible datasets. The public datasets are often of low spatial resolution in remote sensing, climate variable, and the yield records, which limit in precisely characterizing the complex interactions between crop growth and environmental factors. The lack of reliable and widespread sub-county yield records also presents a significant challenge to fully capture the sub-county heterogeneity and generate high-resolution maps. Addressing this limitation is a crucial direction for future work. More field-level datasets are expected to be released to further evaluate the effectiveness of the proposed methods, as an increasing number of dataset journals and platforms are being established to support data sharing. Advances in satellite technology and climate monitoring are also expected to produce more higher-resolution crop surface indicators, providing richer and more diverse agricultural data for better application at finer spatial scales (Prasad et al., 2015). Although collecting high-resolution field-level yield labels is resource-intensive, many efforts have been taken to develop the weakly learning or transfer learning on the county level yield to generate field or subfield yield predictions (Ma et al., 2024; Paudel et al., 2022). By integrating our framework with these advanced learning strategies, future research could extend our model to produce high-resolution yield prediction maps, offering more precise and actionable insights for agricultural practice.

The spatial heterogeneity of time lag effects on corn yield estimates has not been fully explored in this study. This study focuses on using

cross-attention to effectively embed prior knowledge into the temporal dimension. Crop production exhibits spatial heterogeneity that could be partially captured by temporal feature extraction. Specially designed spatial modules may help adequately accommodate feature differences across larger regions. Future studies can extend the application of cross-attention to effectively integrate spatial prior knowledge from different regions. In addition, the use of large language models to incorporate text information from large spatial scales might be a potential method for enhancing the spatial features.

Additionally, this study utilizes a crop type map and commonly used preprocessing techniques to ensure a high-quality and ready-to-use dataset for validating the capabilities of proposed model. However, the Cropland Data Layer (CDL) dataset used to extract the corn area is only available at the end of the season, making it challenging to use for in-season corn feature extraction in current real-world applications. One direct strategy is to use crop type data from the previous year, although this approach may introduce some errors. A better option is to use potential alternatives to the CDL, as an increasing number of studies have proposed advanced methods for mapping in-season crop types (Cai et al., 2018; Gallo et al., 2023; Lin et al., 2022; Rußwurm et al., 2023). Moreover, the Savitzky-Golay filtering employed for optimally smoothing the weekly time series requires information from several future observations. Recently launched satellites, such as Sentinel-2 and Landsat 9, offer high temporal resolution images, which can reduce errors when compositing images every week. Furthermore, newly developed algorithms for cloud removal and missing value reconstruction present promising tools for generating high-quality time series data, reducing the need for smoothing techniques during the preprocessing step (Gonzalez-Calabuig et al., 2025; Xu et al., 2022).

5. Conclusion

In this study, we explicitly encode the time lag effect and phenology sensitivity into a framework by attention mechanisms for estimating crop yield under extreme climate stress. The framework employs two streams of pyramid structure to respectively aggregate remote sensing and climate time series into feature embeddings. A cross attention is adopted to encode time lag effect in fusing these two streams of embeddings and a phenology-sensitivity-guided linear attention is applied atop the framework for encoding sensitivities. Evaluation results demonstrate significant temporal delays in the extreme stress event and the maximum vegetation decay event with an average of 45 days, indicating that extreme temperature events lead to delayed reductions in crop greenness and negative impact yield. The knowledge-encoded two-stream model achieved an RMSE of 1.17 Mg ha^{-1}, significantly outperforming baseline models, which underscores the superiority of knowledge-guided data fusion over traditional methods. Furthermore, model structure ablation analysis revealed that the cross-attention mechanism and time-lag encoding significantly enhance performance. The model interpretation analysis showed that cross-attention mechanisms effectively identify time lag patterns and elucidate the interactions between environmental factors and crop responses, providing valuable insights into yield formation processes. The in-season analysis demonstrates the proposed model could capture extreme stress occurrence, allowing for accurate yield predictions up to 8 weeks in advance, thereby supporting proactive agricultural management and decision-making. The spatial–temporal transferring results highlight the superior generalization capability of the proposed model. The model generally outperformed baseline models across counties from 2013 to 2023. The superior performance was pronounced in regions significantly impacted by extreme heat stress. Moreover, the model demonstrated stable performance across most years. Overall, this research highlights the effectiveness of cross attention to integrating crop-environment interaction signals induced by extreme stress. This encoding strategy could improve early warning systems for crop failures under extreme climate conditions, thereby contributing to enhanced agricultural resilience and food security.

CRediT authorship contribution statement

Xingguo Xiong: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Renhai Zhong:** Writing – review & editing, Methodology, Conceptualization. **Hao Jiang:** Writing – review & editing, Validation. **Ioannis Athanasiadis:** Writing – review & editing, Validation. **Yi Yang:** Writing – review & editing. **Linchao Zhu:** Writing – review & editing, Validation, Supervision, Methodology. **Tao Lin:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Zhejiang Provincial Natural Science Foundation of China (Grant Number LR24C130002), National Natural Science Foundation of China (Grant number 32471987), China National Key Research and Development Plan (Grant Number 2022YFD2001001), and China Scholarship Council (Grant Number 202406320434). We would also like to thank Qiyu Tian for improving the figures and Professor Jingfeng Huang for his valuable comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2025.10.020.

References

Abendroth, L.J., Miguez, F.E., Castellano, M.J., Carter, P.R., Messina, C.D., Dixon, P.M., Hatfield, J.L., 2021. Lengthening of maize maturity time is not a widespread climate change adaptation strategy in the US Midwest. Glob. Chang. Biol. 27, 2426–2440. https://doi.org/10.1111/gcb.15565.

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., Jumper, J.M., 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

Anderson, M.C., Zolin, C.A., Sentelhas, P.C., Hain, C.R., Semmens, K., Tugrul Yilmaz, M., Gao, F., Otkin, J.A., Tetrault, R., 2016. The Evaporative stress Index as an indicator of agricultural drought in Brazil: an assessment based on crop yield impacts. Remote Sens. Environ. 174, 82–99. https://doi.org/10.1016/j.rse.2015.11.034.

Barnabás, B., Jäger, K., Fehér, A., 2008. The effect of drought and heat stress on reproductive processes in cereals. Plant Cell Environ. 31, 11–38. https://doi.org/10.1111/j.1365-3040.2007.01727.x.

Butler, E.E., Huybers, P., 2015. Variations in the sensitivity of US maize yield to extreme temperatures by region and growth phase. Environ. Res. Lett. 10, 034009. https://doi.org/10.1088/1748-9326/10/3/034009.

Butler, E.E., Huybers, P., 2013. Adaptation of US maize to temperature variations. Nature Clim Change 3, 68–72. https://doi.org/10.1038/nclimate1585.

Butler, E.E., Mueller, N.D., Huybers, P., 2018. Peculiarly pleasant weather for US maize. PNAS 115, 11935–11940. https://doi.org/10.1073/pnas.1808035115.

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. Remote Sens. Environ. 210, 35–47. https://doi.org/10.1016/j.rse.2018.02.045.

Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R., Chhetri, N., 2014. A meta-analysis of crop yield under climate change and adaptation. Nature Clim Change 4, 287–291. https://doi.org/10.1038/nclimate2153.

Chang, Y., Latham, J., Licht, M., Wang, L., 2023. A data-driven crop model for maize yield prediction. Commun. Biol. 6, 439. https://doi.org/10.1038/s42003-023-04833-y.

Chen, C.-F.R., Fan, Q., Panda, R., 2021. CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Montreal, QC, Canada, pp. 347–356. 10.1109/ICCV48922.2021.00041.

Cheng, E., Wang, F., Peng, D., Zhang, B., Zhao, B., Zhang, W., Hu, J., Lou, Z., Yang, S., Zhang, H., Lv, Y., 2024. A GT-LSTM Spatio-Temporal Approach for Winter Wheat Yield Prediction: from the Field Scale to County Scale. IEEE Trans. Geosci. Remote Sensing 62, 1–18. https://doi.org/10.1109/TGRS.2024.3418046.

Curnel, Y., De Wit, A.J.W., Duveiller, G., Defourny, P., 2011. Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. Agric. For. Meteorol. 151, 1843–1855. https://doi.org/10.1016/j.agrformet.2011.08.002.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. 28, 2031–2064. https://doi.org/10.1002/joc.1688.

De Wit, A.J.W., Van Diepen, C.A., 2007. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. Agric. For. Meteorol. 146, 38–56. https://doi.org/10.1016/j.agrformet.2007.05.004.

Deryng, D., Conway, D., Ramankutty, N., Price, J., Warren, R., 2014. Global crop yield response to extreme heat stress under multiple climate change futures. Environ. Res. Lett. 9, 034011. https://doi.org/10.1088/1748-9326/9/3/034011.

Dou, Y., Tian, F., Wigneron, J.-P., Li, X., Zhang, W., Chen, Y., Feng, L., Xie, Q., Fensholt, R., 2024. Satellite observations indicate slower recovery of woody components compared to upper-canopy and leaves in tropical rainforests after drought. Commun. Earth Environ. 5, 1–10. https://doi.org/10.1038/s43247-024-01892-9.

Fan, J., Bai, J., Li, Z., Ortiz-Bobea, A., Gomes, C.P., 2022. A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction. Presented at the Proceedings of the AAAI Conference on Artificial Intelligence, pp. 11873–11881. 10.1609/aaai.v36i11.21444.

Feng, L., Wang, Y., Zhang, Z., Du, Q., 2021. Geographically and temporally weighted neural network for winter wheat yield prediction. Remote Sens. Environ. 262, 112514. https://doi.org/10.1016/j.rse.2021.112514.

Feng, P., Wang, B., Liu, D.L., Waters, C., Xiao, D., Shi, L., Yu, Q., 2020. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. Agric. For. Meteorol. 285–286, 107922. https://doi.org/10.1016/j.agrformet.2020.107922.

Feng, P., Wang, B., Liu, D.L., Waters, C., Yu, Q., 2019. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia. Agric. For. Meteorol. 275, 100–113. https://doi.org/10.1016/j.agrformet.2019.05.018.

Gabaldón-Leal, C., Webber, H., Otegui, M.E., Slafer, G.A., Ordóñez, R.A., Gaiser, T., Lorite, I.J., Ruiz-Ramos, M., Ewert, F., 2016. Modelling the impact of heat stress on maize yield formation. Field Crop Res 198, 226–237. https://doi.org/10.1016/j.fcr.2016.08.013.

Gallo, I., Ranghetti, L., Landro, N., La Grassa, R., Boschetti, M., 2023. In-season and dynamic crop mapping using 3D convolution neural networks and sentinel-2 time series. ISPRS J. Photogramm. Remote Sens. 195, 335–352. https://doi.org/10.1016/j.isprsjprs.2022.12.005.

Gao, B.-C., 1995. Normalized difference water index for remote sensing of vegetation liquid water from space, in: Descour, M.R., Mooney, J.M., Perry, D.L., Illing, L.R. (Eds.), Presented at the SPIE's 1995 Symposium on OE/Aerospace Sensing and Dual Use Photonics, Orlando, FL, United States, p. 225. 10.1117/12.210877.

Gao, S., Huang, S., Singh, V.P., Deng, X., Duan, L., Leng, G., Guo, W., Li, Y., Zhang, L., Han, Z., Huang, Q., 2025. Dynamic response of vegetation to meteorological drought and driving mechanisms in Mongolian Plateau. J. Hydrol. 650, 132541. https://doi.org/10.1016/j.jhydrol.2024.132541.

Gitelson, A.A., 2004. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. J. Plant Physiol. 161, 165–173. https://doi.org/10.1078/0176-1617-01176.

Gonzalez-Calabuig, M., Fernández-Torres, M.-Á., Camps-Valls, G., 2025. Generative networks for spatio-temporal gap filling of Sentinel-2 reflectances. ISPRS J. Photogramm. Remote Sens. 220, 637–648. https://doi.org/10.1016/j.isprsjprs.2025.01.016.

Han, C., Zhang, B., Chen, H., Wei, Z., Liu, Y., 2019. Spatially distributed crop model based on remote sensing. Agric Water Manag 218, 165–173. https://doi.org/10.1016/j.agwat.2019.03.035.

Han, J., Shi, L., Yang, Q., Chen, Z., Yu, J., Zha, Y., 2022. Rice yield estimation using a CNN-based image-driven data assimilation framework. Field Crop Res 288, 108693. https://doi.org/10.1016/j.fcr.2022.108693.

Hsiao, J., Swann, A.L.S., Kim, S.-H., 2019. Maize yield under a changing climate: the hidden role of vapor pressure deficit. Agric. For. Meteorol. 279, 107692. https://doi.org/10.1016/j.agrformet.2019.107692.

Huang, H., Huang, J., Feng, Q., Liu, J., Li, X., Wang, X., Niu, Q., 2022. Developing a Dual-Stream Deep-Learning Neural Network Model for improving County-Level Winter Wheat Yield estimates in China. Remote Sens. (Basel) 14, 5280. https://doi.org/10.3390/rs14205280.

Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., Tian, L., Zhang, X., Fan, J., Wu, W., 2016. Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. Agric. For. Meteorol. 216, 188–202. https://doi.org/10.1016/j.agrformet.2015.10.013.

Ines, A.V.M., Das, N.N., Hansen, J.W., Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. Remote Sens. Environ. 138, 149–164. https://doi.org/10.1016/j.rse.2013.07.018.

Jeong, S., Ko, J., Yeom, J.-M., 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., Hu, H., Zhong, R., Xu, J., Xu, J., Huang, J., Wang, S., Ying, Y., Lin, T., 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. Glob. Chang. Biol. 26, 1754–1766. https://doi.org/10.1111/gcb.14885.

Jiang, Z., Zhou, Y., Gao, S., Dong, Z., Wang, Y., Duan, Z., He, W., Liu, Y., Ju, W., 2025. Stronger effects of accumulated soil moisture deficit on gross primary productivity and light use efficiency than lagged soil moisture deficit for cropland and forest. Agric. For. Meteorol. 361, 110317. https://doi.org/10.1016/j.agrformet.2024.110317.

Jin, X., Li, Z., Yang, G., Yang, H., Feng, H., Xu, X., Wang, J., Li, X., Luo, J., 2017. Winter wheat yield estimation based on multi-source medium resolution optical and radar imaging data and the AquaCrop model using the particle swarm optimization algorithm. ISPRS J. Photogramm. Remote Sens. 126, 24–37. https://doi.org/10.1016/j.isprsjprs.2017.02.001.

Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sens. Environ. 141, 116–128. https://doi.org/10.1016/j.rse.2013.10.027.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Kang, Y., Khan, S., Ma, X., 2009. Climate change impacts on crop yield, crop water productivity and food security – a review. Prog. Nat. Sci. 19, 1665–1674. https://doi.org/10.1016/j.pnsc.2009.08.001.

Kang, Y., Özdoğan, M., 2019. Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. Remote Sens. Environ. 228, 144–163. https://doi.org/10.1016/j.rse.2019.04.005.

Khaki, S., Wang, L., 2019. Crop Yield Prediction using Deep Neural Networks. Front. Plant Sci. 10, 621. https://doi.org/10.3389/fpls.2019.00621.

Kim, Y., Lee, J., Kim, J.-H., Ha, J.-W., Zhu, J.-Y., 2023. Dense Text-to-Image Generation with Attention Modulation. 10.48550/arXiv.2308.12964.

Kong, D., Miao, C., Wu, J., Zheng, H., Wu, S., 2020. Time lag of vegetation growth on the Loess Plateau in response to climate factors: Estimation, distribution, and influence. Sci. Total Environ. 744, 140726. https://doi.org/10.1016/j.scitotenv.2020.140726.

Lesk, C., Anderson, W., Rigden, A., Coast, O., Jägermeyr, J., McDermid, S., Davis, K.F., Konar, M., 2022. Compound heat and moisture extreme impacts on global crop yields under climate change. Nat. Rev. Earth Environ. 3, 872–889. https://doi.org/10.1038/s43017-022-00368-8.

Lin, C., Zhong, L., Song, X.-P., Dong, J., Lobell, D.B., Jin, Z., 2022. Early- and in-season crop type mapping without current-year ground truth: Generating labels from historical information via a topology-based approach. Remote Sens. Environ. 274, 112994. https://doi.org/10.1016/j.rse.2022.112994.

Lin, F., Guillot, K., Crawford, S., Zhang, Y., Yuan, X., Tzeng, N.-F., 2024. An Open and Large-Scale Dataset for Multi-Modal Climate Change-aware Crop Yield Predictions, in: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 5375–5386. 10.1145/3637528.3671536.

Lin, T., Zhong, R., Wang, Y., Xu, J., Jiang, H., Xu, J., Ying, Y., Rodriguez, L., Ting, K.C., Li, H., 2020. DeepCropNet: a deep spatial-temporal learning framework for county-level corn yield estimation. Environ. Res. Lett. 15, 034016. https://doi.org/10.1088/1748-9326/ab66cb.

Liu, L., Zhou, W., Guan, K., Peng, B., Xu, S., Tang, J., Zhu, Q., Till, J., Jia, X., Jiang, C., Wang, S., Qin, Z., Kong, H., Grant, R., Mezbahuddin, S., Kumar, V., Jin, Z., 2024. Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems. Nat. Commun. 15, 357. https://doi.org/10.1038/s41467-023-43860-5.

Liu, Q., Fu, Y.H., Zhu, Z., Liu, Y., Liu, Z., Huang, M., Janssens, I.A., Piao, S., 2016. Delayed autumn phenology in the Northern Hemisphere is related to change in both climate and spring phenology. Glob. Chang. Biol. 22, 3702–3711. https://doi.org/10.1111/gcb.13311.

Liu, X., Pan, Y., Zhu, X., Yang, T., Bai, J., Sun, Z., 2018. Drought evolution and its impact on the crop yield in the North China Plain. J. Hydrol. 564, 984–996. https://doi.org/10.1016/j.jhydrol.2018.07.077.

Lobell, D.B., Hammer, G.L., Chenu, K., Zheng, B., McLean, G., Chapman, S.C., 2015. The shifting influence of drought and heat stress for crops in northeast Australia. Glob. Chang. Biol. 21, 4115–4127. https://doi.org/10.1111/gcb.13022.

Ma, Y., Liang, S.-Z., Myers, D.B., Swatantran, A., Lobell, D.B., 2024. Subfield-level crop yield mapping without ground truth data: a scale transfer framework. Remote Sens. Environ. 315, 114427. https://doi.org/10.1016/j.rse.2024.114427.

Ma, Y., Zhang, Z., Kang, Y., Özdoğan, M., 2021. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. Remote Sens. Environ. 259, 112408.

Mateo-Sanchis, A., Piles, M., Muñoz-Marí, J., Adsuara, J.E., Pérez-Suay, A., Camps-Valls, G., 2019. Synergistic integration of optical and microwave satellite data for crop yield estimation. Remote Sens. Environ. 234, 111460. https://doi.org/10.1016/j.rse.2019.111460.

Mena, F., Pathak, D., Najjar, H., Sanchez, C., Helber, P., Bischke, B., Habelitz, P., Miranda, M., Siddamsetty, J., Nuske, M., Charfuelan, M., Arenas, D., Vollmer, M.,

Dengel, A., 2025. Adaptive fusion of multi-modal remote sensing data for optimal sub-field crop yield prediction. Remote Sens. Environ. 318, 114547. https://doi.org/10.1016/j.rse.2024.114547.
Mnih, V., Heess, N., Graves, A., 2014. Recurrent models of visual attention. Adv. Neural Inf. Process. Syst. 27.
Nóia Júnior, R.D.S., Asseng, S., Müller, C., Deswarte, J.-C., Cohan, J.-P., Martre, P., 2025. Negative impacts of climate change on crop yields are underestimated. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2025.05.002. S136013852500130X.
Pastor-Guzman, J., Dash, J., Atkinson, P.M., 2018. Remote sensing of mangrove forest phenology and its environmental drivers. Remote Sens. Environ. 205, 71–84. https://doi.org/10.1016/j.rse.2017.11.009.
Paudel, D.R., Marcos, D., de Wit, A., Boogaard, H., Athanasiadis, I.N., 2022. A weakly supervised framework for high-resolution crop yield forecasts.
Peng, B., Guan, K., Pan, M., Li, Y., 2018. Benefits of Seasonal climate Prediction and Satellite Data for forecasting U.S. Maize Yield. Geophys. Res. Lett. 45, 9662–9671. https://doi.org/10.1029/2018GL079291.
Peng, D., Cheng, E., Feng, X., Hu, J., Lou, Z., Zhang, H., Zhao, B., Lv, Y., Peng, H., Zhang, B., 2024. A Deep–Learning Network for Wheat Yield Prediction Combining Weather forecasts and Remote Sensing Data. Remote Sens. (Basel) 16, 3613. https://doi.org/10.3390/rs16193613.
Phung, Q., Ge, S., Huang, J.-B., 2024. Grounded Text-to-Image Synthesis with Attention Refocusing, in: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA, pp. 7932–7942. 10.1109/CVPR52733.2024.00758.
Prasad, P.V.V., Staggenborg, S.A., Ristic, Z., 2015. Impacts of Drought and/or Heat stress on Physiological, Developmental, growth, and Yield Processes of Crop Plants. In: Ahuja, L.R., Reddy, V.R., Saseendran, S.A., Yu, Q. (Eds.), Advances in Agricultural Systems Modeling. American Society of Agronomy and Soil Science Society of America, Madison, WI, USA, pp. 301–355, 10.2134/advagricsystmodel1.c11.
Qiao, M., He, X., Cheng, X., Li, P., Zhao, Q., Zhao, C., Tian, Z., 2023. KSTAGE: a knowledge-guided spatial-temporal attention graph learning network for crop yield prediction. Inf. Sci. 619, 19–37. https://doi.org/10.1016/j.ins.2022.10.112.
Rezaei, E.E., Webber, H., Asseng, S., Boote, K., Durand, J.L., Ewert, F., Martre, P., MacCarthy, D.S., 2023. Climate change impacts on crop yields. Nat. Rev. Earth Environ. 4, 831–846. https://doi.org/10.1038/s43017-023-00491-0.
Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D., Tavenard, R., 2023. End-to-end learned early classification of time series for in-season crop type mapping. ISPRS J. Photogramm. Remote Sens. 196, 445–456. https://doi.org/10.1016/j.isprsjprs.2022.12.016.
Sakamoto, T., 2020. Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. ISPRS J. Photogramm. Remote Sens. 160, 208–228. https://doi.org/10.1016/j.isprsjprs.2019.12.012.
Schaaf, C., Wang, Z., 2015. MCD43A4 MODIS/Terra+ aqua BRDF/albedo nadir BRDF adjusted RefDaily L3 global 500 m V006. NASA EOSDIS Land Processes DAAC.
Schlenker, W., Roberts, M.J., 2009. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. PNAS 106, 15594–15598. https://doi.org/10.1073/pnas.0906865106.
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. Sci. Rep. 11, 1606. https://doi.org/10.1038/s41598-020-80820-1.
Shuai, G., Basso, B., 2022. Subfield maize yield prediction improves when in-season crop water deficit is included in remote sensing imagery-based models. Remote Sens. Environ. 272, 112938. https://doi.org/10.1016/j.rse.2022.112938.
USDA-NASS, 2020a. Quick Stats 2.0. SDA-NASS, Washington, DC.
USDA-NASS, 2020b. USDA national agricultural statistics service cropland data layer. Published crop-specific data layer.
Vogel, E., Donat, M.G., Alexander, L.V., Meinshausen, M., Ray, D.K., Karoly, D., Meinshausen, N., Frieler, K., 2019. The effects of climate extremes on global agricultural yields. Environ. Res. Lett. 14, 054010. https://doi.org/10.1088/1748-9326/ab154b.
Wei, X., Zhang, T., Li, Y., Zhang, Y., Wu, F., 2020. Multi-Modality Cross Attention Network for Image and Sentence Matching, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA, pp. 10938–10947. 10.1109/CVPR42600.2020.01095.
Wu, D., Zhao, X., Liang, S., Zhou, T., Huang, K., Tang, B., Zhao, W., 2015. Time-lag effects of global vegetation responses to climate change. Glob. Chang. Biol. 21, 3520–3531. https://doi.org/10.1111/gcb.12945.
Xiao, L., Wang, G., Wang, E., Liu, S., Chang, J., Zhang, P., Zhou, H., Wei, Y., Zhang, H., Zhu, Y., Shi, Z., Luo, Z., 2024. Spatiotemporal co-optimization of agricultural management practices towards climate-smart crop production. Nat. Food 5, 59–71. https://doi.org/10.1038/s43016-023-00891-x.
Xiong, X., Zhong, R., Tian, Q., Huang, J., Zhu, L., Yang, Y., Lin, T., 2024. Daily DeepCropNet: a hierarchical deep learning approach with daily time series of vegetation indices and climatic variables for corn yield estimation. ISPRS J. Photogramm. Remote Sens. 209, 249–264. https://doi.org/10.1016/j.isprsjprs.2024.02.008.
Xu, J., Yang, J., Xiong, X., Li, H., Huang, J., Ting, K.C., Ying, Y., Lin, T., 2021. Towards interpreting multi-temporal deep learning models in crop mapping. Remote Sens. Environ. 264, 112599. https://doi.org/10.1016/j.rse.2021.112599.
Xu, M., Deng, F., Jia, S., Jia, X., Plaza, A.J., 2022. Attention mechanism-based generative adversarial networks for cloud removal in Landsat images. Remote Sens. Environ. 271, 112902. https://doi.org/10.1016/j.rse.2022.112902.
Yang, J., Wang, H., Wu, S., Wu, X., Wu, J., 2024. Responses of the Key Phenological Characteristics and Carbon Flux of Winter Wheat to Climatic Time-lag Effects and Crop Management on the North China Plain. Int. J. Plant Prod. 18, 301–312. https://doi.org/10.1007/s42106-024-00292-5.
Yang, Q., Liu, L., Zhou, J., Ghosh, R., Peng, B., Guan, K., Tang, J., Zhou, W., Kumar, V., Jin, Z., 2023. A flexible and efficient knowledge-guided machine learning data assimilation (KGML-DA) framework for agroecosystem prediction in the US Midwest. Remote Sens. Environ. 299, 113880. https://doi.org/10.1016/j.rse.2023.113880.
Yang, Y., Tao, B., Ruane, A.C., Shen, C., Matteson, D.S., Cousin, R., Ren, W., 2025. Widespread advances in Corn and soybean Phenology in Response to Future climate Change across the United States. J. Geophys. Res. Biogeo. 130, e2024JG008266. https://doi.org/10.1029/2024JG008266.
You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep Gaussian Process for Crop Yield Prediction based on Remote Sensing Data. AAAI 31. https://doi.org/10.1609/aaai.v31i1.11172.
Zampieri, M., Ceglar, A., Dentener, F., Toreti, A., 2017. Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales. Environ. Res. Lett. 12, 064008. https://doi.org/10.1088/1748-9326/aa723b.
Zhong, R., Zhu, Y., Wang, X., Li, H., Wang, B., You, F., Rodríguez, L.F., Huang, J., Ting, K.C., Ying, Y., Lin, T., 2023. Detect and attribute the extreme maize yield losses based on spatio-temporal deep learning. Fundam. Res. 3, 951–959. https://doi.org/10.1016/j.fmre.2022.05.006.
Zhou, Q., Guan, K., Wang, S., Hipple, J., Chen, Z., 2024. From satellite-based phenological metrics to crop planting dates: Deriving field-level planting dates for corn and soybean in the U.S. Midwest. ISPRS J. Photogramm. Remote Sens. 216, 259–273. https://doi.org/10.1016/j.isprsjprs.2024.07.031.