PhenoYieldNet: Learning Crop-Aware Phenological Responses for Multi-Crop Yield Prediction

Supplementary Material

This supplementary material is organized as follows:

- Sec. A provides additional descriptions of the datasets and crop growing seasons.
- Sec. B elaborates on the implementation details of both single-crop and multi-crop baselines.
- Sec. D summarizes the calculation of evaluation metrics.
- Sec. C presents further experimental results, including hyperparameter analysis and visualizations.

A. Dataset Details

We employ two publicly available datasets, CropNet [21] and [40]'s dataset (denoted as MODIS) for multi-crop yield prediction. CropNet contains the satellite imagery of surface reflectance at a 2-week interval, and 9 meteorological variables from HRRR at a daily interval, including averaged temperature, maximal temperature, minimal temperature, precipitation, relative humidity, wind gust, wind speed, downward shortwave radiation flux and vapor pressure deficit. MODIS includes 8-day satellite imagery and temperature variables.

Both datasets were masked along the temporal dimension to retain only the data within crop-specific growing seasons, ensuring the analysis targets critical developmental stages. These seasons are defined based on USDA calendars (Table A.1).

Dataset	Crop Type	Growing Season
CropNet	Corn	Apr. - Dec.
	Cotton	Apr. - Dec.
	Soybean	Apr. - Sep.
	Winter Wheat	Sep. - Aug. (next year)
MODIS	Soybean	49th - 281st day of year

Table A.1. The growing seasons of different crop types.

B. Implementation Details of Baselines

To ensure a fair comparison, all baseline models were re-implemented within our framework and adapted to the specific modalities available.

B.1. Single-Crop Methods

MMST-ViT [21] predicts crop yields based on Sentinel-2 imagery and meteorological data. It first employs a spatial Transformer to extract the image features, and an MLP to extract the meteorological data, and then utilizes a multimodal Transformer with a cross-attention mechanism to fuse the two modalities, and finally an MLP for prediction.

UNet-ConvLSTM [17] is originally developed on Sentinel-2 imagery and managing practice data. Here, we adapt it by replacing the management data branch with our meteorological data. An MLP is used to project the meteorological features, which are then concatenated to the bottleneck of the UNet feature extractor before being passed to the ConvLSTM to output predictions.

Transformer-based [12] is developed solely on satellite imagery, and we remove the meteorological data input, and follow the original Transformer architecture to extract features and finally employ an MLP to output the predicted yields.

MMVF [27] employs different encoders according to different modalities. Here, we remove other modalities like soil and DEM, which are not presented in our dataset, and preserve the LSTM-based encoder for satellite imagery and meteorological data, with a gated fusion mechanism to fuse each pixel from the two modalities. The fused representations are finally fed into a fully connected layer to predict yield maps.

B.2. Multi-Crop Method

YieldNet [20] following the paper, we remove the meteorological data and preserve the satellite imagery input, and feed it into a shared CNN, which employs a 2D convolution operation to capture the temporal dependency. The prediction head consists of two convolutional layers followed by two fully connected layers.

C. Additional Experimental Results

D. Evaluation Metrics

We adopt three metrics, including Root Mean Square Error (RMSE), R-squared (R^2), and Pearson Correlation Coefficient (Corr), for comparative evaluation. Given the ground-truth(GT) yield y and the corresponding predicted yield $\hat{y}$, these metrics can be calculated as follows:

$$RMSE = \sqrt{\frac{\sum(\mathbf{y} - \hat{\mathbf{y}})^2}{N}} \tag{D.1}$$

weight	Corn			Cotton			Soybean			Winter Wheat		
	RMSE(↓)	R^2(↑)	Corr(↑)	RMSE(↓)	R^2(↑)	Corr(↑)	RMSE(↓)	R^2(↑)	Corr(↑)	RMSE(↓)	R^2(↑)	Corr(↑)
$\lambda_\mu = 0.0$, $\lambda_\nu = 1.0$	16.94	0.445	0.668	78.88	0.513	0.716	6.36	0.547	0.739	8.25	0.078	0.277
$\lambda_\mu = 0.5$,$\lambda_\nu = 0.5$	16.52	0.516	0.718	54.88	0.638	0.799	5.91	0.616	0.785	8.32	0.136	0.368
$\lambda_\mu = 1.0$,$\lambda_\nu = 0.0$	20.43	0.453	0.673	87.33	0.504	0.711	5.96	0.608	0.780	8.13	0.095	0.308

Table C.1. Hyperparameter analysis on the weights of trend and variation components on CropNet.

$$R^2 = 1 - \frac{\sum(\mathbf{y} - \hat{\mathbf{y}})^2}{\sum(\mathbf{y} - \bar{\mathbf{y}})^2} \tag{D.2}$$

$$Corr = \frac{\sum(\mathbf{y} - \bar{\mathbf{y}})(\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}})}{\sqrt{\sum(\mathbf{y} - \bar{\mathbf{y}})^2}\sqrt{\sum(\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}})^2}} \tag{D.3}$$

where $\bar{\mathbf{y}}$ and $\bar{\hat{\mathbf{y}}}$ denote the average GT yield and predicted yield values, respectively. N is the total number of samples in the testing set.

Method	Corn	Cotton	Soybean	Winter Wheat
SMA	19.88	85.3	6.04	**8.25**
Fourier Feature	18.21	81.13	6.94	8.72
CPA (Ours)	**16.52**	**54.88**	**5.91**	8.32

Table D.1. RMSE comparison of time-series feature modeling methods on CropNet dataset.

D.1. Visualizations of Prediction Error Map

To provide a qualitative assessment of prediction performance across different regions, we further visually illustrate the RMSE distribution of the proposed method and YieldNet on four states, including Illinois (IL), Louisiana (LA), Iowa (IA) and Mississippi (MS). As shown in Fig.D.1, the maps cover four key agricultural states for *corn* and *soybean*. It is visually evident that the proposed method consistently produces lower prediction errors across most counties compared to the YieldNet with brighter colors, which suggests that our method not only improves prediction accuracy, but also the robustness across different regions.

D.2. Analysis of CPA/CPB modules

Hyperparameter Sensitivity. To evaluate the effect of trend and variation components within the CPA module, we conducted a hyperparameter analysis on their corresponding weights, λ_μ and λ_ν. As presented in Table C.1, setting the equal weights yields the best performance across all crops except *winter wheat*, which achieved the lowest RMSE by solely relying on the trend component ($\lambda_\mu = 1.0$).

Figure D.1. Yield prediction error distribution for (a) Soybean and (b) Corn.

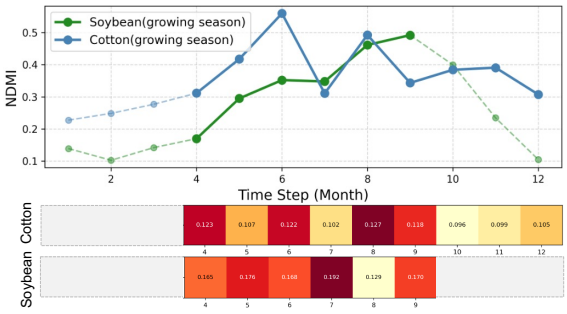

Figure D.2. The NDMI curves and temporal attention score heatmaps for *soybean* and *cotton*.

Comparison with Simple Time-Series Features. We conducted additional ablation studies comparing our approach with simple time-series modeling methods, including smoothing moving average (SMA) and Fourier features. The results in Table D.1 show that PhenoYieldNet consis-

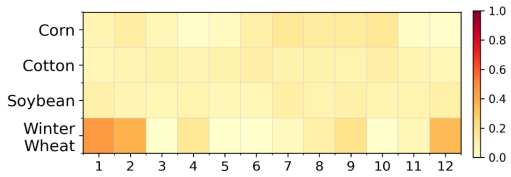

Figure D.3. The visualization of one-hot embedding for *soybean*.

tently outperforms these baselines. This confirms that fixed mathematical biases are insufficient for modeling complex agricultural data, whereas our learnable CPA can effectively capture more complex phenological dynamics.

Attention Visualization and Phenological Alignment. We further visualize the NDMI curves and learned temporal attention scores for two crops, *soybean* and *cotton*, with distinct phenological calendars. As shown in Fig. D.2, it can be observed that these learned attention peaks consistently coincide with the crop-relevant growing-season windows and key transition phases derived from the NDMI curves. This structural consistency demonstrates that our CPA module, guided by the crop-specific queries from the CPB, effectively learns phenology-aligned temporal relevance rather than mismatched patterns.

Necessity of Dynamic Crop Queries. To further justify the design of our crop-aware attention, we compared it against a static one-hot crop encoding baseline. As shown in Fig. D.3, while one-hot vectors provide a deterministic representation of crop identity, their signals remain almost constant across time steps. Under non-stationary growing cycles and climate variability, crop identity alone does not specify temporal dynamics. In contrast, crop-aware attention learns temporal relevance, yielding significantly better predictive performance than the static one-hot alternative (shown in Table D.2).

Method	Corn	Cotton	Soybean	Winter Wheat
One-hot	20.87	129.02	6.11	10.26
CPB (Ours)	**16.52**	**54.88**	**5.91**	**8.32**

Table D.2. RMSE comparison with static one-hot alternative.