# Interpreting the Impact of Weather on Crop Yield Using Attention

**Tryambak Gangopadhyay**
Department of Mechanical Engineering
Iowa State University
Ames, IA 50011
tryambak@iastate.edu

**Johnathon Shook**
Department of Agronomy
Iowa State University
Ames, IA 50011
jmshook@iastate.edu

**Asheesh K. Singh**
Department of Agronomy
Iowa State University
Ames, IA 50011
singhak@iastate.edu

**Soumik Sarkar**
Department of Mechanical Engineering
Iowa State University
Ames, IA 50011
soumiks@iastate.edu

## Abstract

Accurate prediction of crop yield supported by scientific and domain-relevant interpretations can improve agricultural breeding by providing monitoring across diverse climatic conditions. The use of this information in plant breeding can help provide protection against weather challenges to crop production, including erratic rainfall and temperature variations. In addition to isolating the important time-steps, researchers are interested to understand the effect of different weather variables on crop yield. In this paper, we propose a novel attention-based model that can learn the most significant variables across different weeks in the crop growing season and highlight the most important time-steps (weeks) to predict the annual crop yield. We demonstrate our model's performance on a dataset based on historical performance records from Uniform Soybean Tests (UST) in North America. The interpretations provided by our model can help in understanding the impact of weather variability on agricultural production in the presence of climate change and formulating breeding strategies to circumvent these climatic challenges.

## 1 Introduction

One important challenge in plant breeding and crop production is to predict performance (for example, seed yield) in different environments. In addition to generating predictions, interpretability is an important aspect to generate domain insights. It is important to understand how agricultural production is affected by the variability of weather parameters in the presence of global climate change, especially with a higher occurrence of extreme weather events. The ability to interpret prediction outcomes from a machine learning model (learning temporal dependencies from multivariate time series) can significantly benefit the domain experts.

Among different crops, soybean has a long history of cultivation in the US and Canada [1, 2, 3]. North American annual soybean yield trials (known as Uniform Soybean Tests (UST)) have been coordinated through the United States Department of Agriculture (USDA) since 1941 [4, 5]. These trials are used to evaluate current and experimental varieties in multiple environments within their range of adaptation. These trials are valuable sources of historical and current data.

AI for Earth Sciences Workshop at NeurIPS 2020.

Long Short Term Memory (LSTM) networks, which can effectively capture the long-term temporal dependencies in multivariate time series [6], have been utilized in different applications, including yield prediction [7, 8]. Attention-based model [9] was initially introduced for neural machine translation to outperform the Encoder-Decoder model [10, 11]. Attention based models have also been proposed for time series prediction [12, 13, 14, 15, 16, 17, 18]. Some models [19] have only spatial attention while some models are non-causal [12, 14] or non-scalable [13]. LSTM based model has been used for corn yield estimation [20], but the model lacks interpretability and temporal resolution in the absence of daily weather data. Attention-based LSTM has been used along with multi-task learning (MTL) output layers [21] for county-level corn yield anomaly prediction without field-scale farming data. Previous works using deep learning for yield prediction has utilized multi-spectral data [22] and applied deep neural networks [23] without considering model interpretability. Only temporal attention has been studied in [15] without considering the importance of different variables for yield prediction.

In this paper, we propose a model based on LSTM and a dual-attention mechanism that is accurate in predicting crop yield and can provide interpretations across 30 weeks of weather data in the growing season. The model is causal (i.e., only depends on past inputs and does not use future inputs while learning the temporal dependencies) and scales well with an increase in the number of variables. To the best of our knowledge, this is the first work on attention-based model for spatiotemporal interpretability in yield prediction. The model works on daily weather data, aggregating weekly information using spatial attention. It highlights which weather variables are significant in each week. After the spatial attention phase and encoding with LSTM layers, the temporal attention learns the most significant time-steps (weeks). We envision broad applicability of this approach for soybean and other crop species under different climatic conditions.

## 2 Dual Attention (Dual-Att) Model

We denote the daily multivariate time series input as $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^N]^\top \in \mathbb{R}^{N \times T}$, where $T$ denotes the entire crop growing season, which is 210 days. $N$ denotes the number of weather variables. We divide the entire dataset into weekly time windows with no overlap between two consecutive windows. Therefore, in total, there are 30 ($T_x$) time-windows or weeks, referred to as time-steps here. At time-step (week) $t \in \{1, 2, ..., T_x\}$, all the associated weather variables for that week are denoted as $\mathbf{x}_t = [\mathbf{x}_t^1, \mathbf{x}_t^2, ..., \mathbf{x}_t^N]^\top \in \mathbb{R}^{N \times 7}$. We can also express the time series as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{T_x}]^\top$. With $\mathbf{X}$ as input, we predict $\mathbf{y} \in \mathbb{R}$, the yearly value of crop yield.

We propose a novel attention-based model called the Dual Attention (Dual-Att) Model, which can provide insights regarding the relative significance of different weather variables on yield prediction throughout the crop growing season. The model comprises spatial attention before the encoder (LSTM layers) and temporal attention after the encoding phase. The model is illustrated in Fig 1.

The inputs to the spatial attention are spatial embeddings generated from the data of the corresponding week $t$. At time-step $t$, a feed forward neural network is used to compute the spatial embedding for each feature $\mathbf{x}_t^i \in \mathbb{R}^7, i \in \{1, 2, ..., N\}$. From $\mathbf{x}_t = [\mathbf{x}_t^1, \mathbf{x}_t^2, ..., \mathbf{x}_t^N]^\top$, the spatial embeddings are computed as $\mathbf{d}_t = [\mathbf{d}_t^1, \mathbf{d}_t^2, ..., \mathbf{d}_t^N]^\top$, where $\mathbf{d}_t^i \in \mathbb{R}^m$.

We use a feed-forward neural network as an alignment model for computing the associated energies of spatial attention. Given the spatial embedding $\mathbf{d}_t^i \in \mathbb{R}^m$, the spatial attention weight $\beta_t^i$ for the $i$-th variable is computed. ReLU activation function is used instead of tanh due to slightly better results observed in the empirical studies. The parameters to learn are $W_e \in \mathbb{R}^m$. Then, the *spatial context* vector $\mathbf{g}_t$ is computed using the spatial attention weights.

$$e_t^i = \mathrm{ReLU}(W_e^\top \mathbf{d}_t^i),\ \beta_t^i = \frac{\exp(e_t^i)}{\sum_{o=1}^N \exp(e_t^o)},\ \mathbf{g}_t = \sum_{i=1}^N \beta_t^i \mathbf{d}_t^i \tag{1}$$

After the spatial attention phase, the sequence of spatial context vectors can be denoted as $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_{T_x}]^\top$, where $\mathbf{g}_t \in \mathbb{R}^m$. $\mathbf{G}$ is the input to the encoder which consists of two stacked LSTM layers. After reading the input sequence in order from $\mathbf{g}_1$ to $\mathbf{g}_{T_x}$, the encoder learns the temporal dependencies with two LSTM layers and computes the sequence of hidden states (temporal embeddings) $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_{T_x}]^\top$, where $\mathbf{h}_t \in \mathbb{R}^m$.

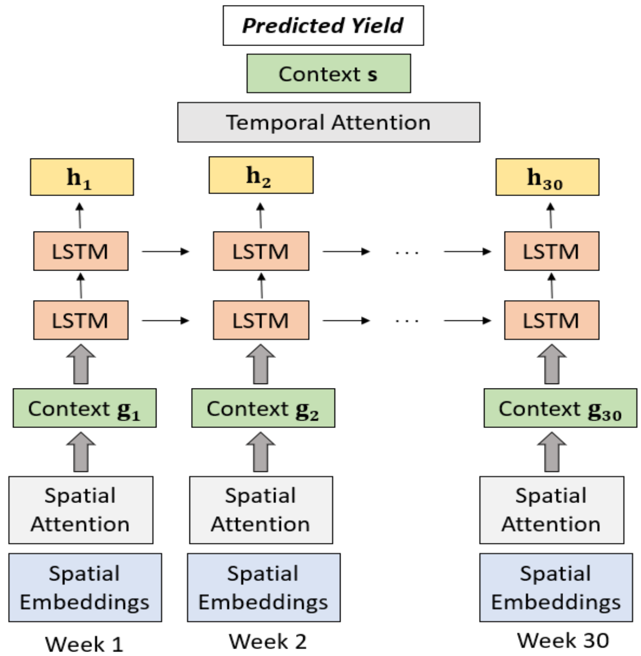


Figure 1: *Illustration of the proposed Dual Attention (Dual-Att) Model.*

The temporal attention mechanism takes the temporal embeddings as input. The associated energy $a^t$ is first computed using a feed-forward neural network to get the temporal attention weight $\alpha^t$ corresponding to the hidden state $\mathbf{h}_t$. The weight $\alpha^t$ shows the importance of $\mathbf{h}_t$ in predicting the yield $\mathbf{y}$. The *temporal context* vector $\mathbf{s}$ is then computed to aggregate the information across all the weeks in the crop growing season.

$$a^t = \mathrm{ReLU}(W_a^\top \mathbf{h}_t),\ \alpha^t = \frac{\exp(a^t)}{\sum_{l=1}^{T_x} \exp(a^l)},\ \mathbf{s} = \sum_{t=1}^{T_x} \alpha^t \mathbf{h}_t \tag{2}$$

The parameters to learn are $W_a \in \mathbb{R}^m$. The proposed Dual-Att Model aggregates the weekly information using spatial attention and learns the temporal dependencies followed by aggregating the temporal information across the growing season. It can highlight the important weather variables for each week and also identify the significant weeks across the year.

# 3 Experiments

## 3.1 Dataset

From 2003-2015 USTs, files were were downloaded as PDFs [4, 5]. The on-line utility Zamzar (zamzar.com) was used for data pre-processing. Manually, we curated the tables to align all performance records for corresponding genotype/location combinations. We do not consider records without yield data (due to a variety not being planted in a specific location or dying before production of seed). After data cleaning, the final dataset used for this paper's experiments comprise 103,365 performance records over 13 years, representing 5839 unique genotypes. The performance records were augmented with the weather data based on the nearest available weather station (25km grid) on Weather.com. We compiled the daily weather records throughout the growing season (defined April 1 through October 31). The dataset consists of 7 weather variables - Average Direct Normal Irradiance (ADNI), Average Precipitation (AP), Average Relative Humidity (ARH), Maximum Direct Normal Irradiance (MDNI), Maximum Surface Temperature (MaxSur, $^\circ F$), Minimum Surface Temperature (MinSur, $^\circ F$) and Average Surface Temperature (AvgSur, $^\circ F$). The training, validation and test sizes are approximately 82,692, 10,336, and 10,337, respectively.

Table 1: Comparison of empirical results with baseline models on the test set. Tr. Time / Iter: Training Time / 1000 iterations, Num. Params: Number of trainable parameters

| Model | RMSE | MAE | $R^2$ Score | Tr. Time / Iter | Num. Params |
|---|---|---|---|---|---|
| LSTM | 8.714 | 6.502 | 0.704 | 35.4 s | 202,369 |
| LSTM-Att | 8.759 | 6.599 | 0.701 | 36.2 s | 202,497 |
| Dual-Att | 9.268 | 7.008 | 0.665 | 43.1 s | 265,601 |

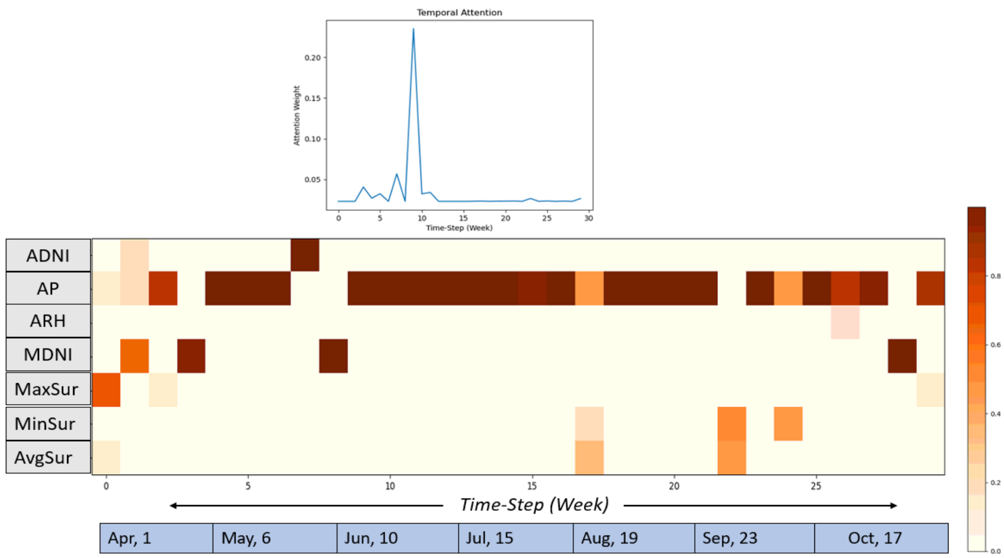


Figure 2: *Distribution of temporal attention weights and weights for different variables spanning the growing season for a sample performance record in the test set.*

### 3.2 Baseline Models and Results

**Baseline Models:** We use two baseline models for comparison of the empirical results: LSTM model and LSTM with temporal attention (LSTM-Att) model. In the LSTM model, there are two encoder LSTM layers with no use of an attention mechanism. The LSTM-Att model comprises of encoder LSTM layers and temporal attention without the use of spatial attention. The hidden state dimensions for the LSTM and LSTM-Att models are kept the same as that of the Dual-Att model.

**Results:** We select the optimal hyper-parameters for the Dual-Att model after performing experiments. We keep the hidden state dimensions of the LSTM layers and the spatial embedding dimension the same for simplicity. The dimension of 128 shows better performance in our experiments. We use Adam optimizer with a constant learning rate of 0.001 to train our model on NVIDIA Titan RTX GPU. Three evaluation metrics are used: root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination or R-squared score ($R^2$). From the empirical results in Table 1, Dual-Att maintains high accuracy, with the performance quite close to the baseline models on all the three evaluation metrics. Dual-Att is computationally tractable and provides the added benefit of spatial and temporal interpretability. From Fig. 2, for a test sample, we observe that average precipitation (AP) is found to be the most important variable for most of the weeks in the growing season. Irradiance (ADNI, MDNI) is also highlighted the most by Dual-Att for few weeks. The temporal attention weights highlight the increasing importance of the features in the time period of weeks 8 to 10.

## 4 Conclusion

In this paper, we propose a model based on a dual-attention mechanism, which can be a great resource for domain experts who seek trustworthy (not just a black-box) deep learning models for crop yield prediction, particularly in the context of weather variables and their role in crop yield. Valuable insights can be gained by understanding predictions from the perspectives of ‘what’ and ‘where’ highlighted by our proposed model. The insights obtained by using this model can help us understand the impact of weather on yield prediction and strategize plant breeding activities with data-driven decision-making. In the future, we plan to continue fine-tuning our model to achieve higher accuracy and perform detailed interpretation studies with different test-case scenarios.

## References


[1] Theodore Hymowitz and Jack Rodney Harlan. Introduction of soybean to north america by samuel bowen in 1765. *Economic Botany*, 37(4):371–379, 1983.

[2] Soybeans: Planted acreage by county. `https://www.nass.usda.gov/Charts_and_Maps/Crops_County/sb-pl.php`. Accessed: 2020-05-06.

[3] Canadian soybean seeded acres (1980 to current). `http://soycanada.ca/statistics/seeded-area-acres/`. Accessed: 2020-05-06.

[4] Uniform soybean tests, northern region. `https://www.ars.usda.gov/midwest-area/west-lafayette-in/crop-production-and-pest-control-research/docs/uniform-soybean-tests-northern-region/`. Accessed: 2020-05-06.

[5] Uniform soybean tests. `https://www.ars.usda.gov/southeast-area/stoneville-ms/crop-genetics-research/docs/uniform-soybean-tests/`. Accessed: 2018-01-10.

[6] Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. Long short term memory networks for anomaly detection in time series. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium)*, page 89. Presses universitaires de Louvain, 2015 April 22-24.

[7] Johnathon Shook, Linjiang Wu, Tryambak Gangopadhyay, Baskar Ganapathysubramanian, Soumik Sarkar, and Asheesh K Singh. Integrating genotype and weather variables for soybean yield prediction using deep learning. *bioRxiv*, 2018.

[8] Tryambak Gangopadhyay, Anthony Locurto, James B Michael, and Soumik Sarkar. Deep learning algorithms for detecting combustion instabilities. In *Dynamics and Control of Energy Systems*, pages 283–300. Springer, 2020.

[9] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[10] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[11] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

[12] Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*, 2017.

[13] Ritambhara Singh, Jack Lanchantin, Arshdeep Sekhon, and Yanjun Qi. Attend and predict: Understanding gene regulation by selective attention on chromatin. In *Advances in neural information processing systems*, pages 6785–6795, 2017.

[14] Huan Song, Deepta Rajan, Jayaraman J Thiagarajan, and Andreas Spanias. Attend and diagnose: Clinical time series analysis using attention models. In *Thirty-second AAAI conference on artificial intelligence*, 2018.

[15] Johnathon Shook, Tryambak Gangopadhyay, Linjiang Wu, Baskar Ganapathysubramanian, Soumik Sarkar, and Asheesh K Singh. Crop yield prediction integrating genotype and weather variables using deep learning. *arXiv preprint arXiv:2006.13847*, 2020.

[16] Tryambak Gangopadhyay, Sin Yong Tan, Genyi Huang, and Soumik Sarkar. Temporal attention and stacked lstms for multivariate time series prediction. In *NeurIPS Workshop on Modeling and Decision-Making in the Spatiotemporal Domain*. NeurIPS, 2018.

[17] Tryambak Gangopadhyay, Sin Yong Tan, Zhanhong Jiang, Rui Meng, and Soumik Sarkar. Spatiotemporal attention for multivariate time series prediction and interpretation. *arXiv preprint arXiv:2008.04882*, 2020.

[18] Tryambak Gangopadhyay, Sin Yong Tan, Zhanhong Jiang, and Soumik Sarkar. Interpretable deep attention model for multivariate time series prediction in building energy systems. In *International Conference on Dynamic Data Driven Application Systems*, pages 93–101. Springer, 2020.

[19] Xuan Zhang, Xun Liang, Aakas Zhiyuli, Shusen Zhang, Rui Xu, and Bo Wu. At-lstm: An attention-based lstm model for financial time series prediction. In *IOP Conference Series: Materials Science and Engineering*, volume 569, page 052037. IOP Publishing, 2019.

[20] Hao Jiang, Hao Hu, Renhai Zhong, Jinfan Xu, Jialu Xu, Jingfeng Huang, Shaowen Wang, Yibin Ying, and Tao Lin. A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the us corn belt at the county level. *Global change biology*, 2019.

[21] Tao Lin, Renhai Zhong, Yudi Wang, Jinfan Xu, Hao Jiang, Jialu Xu, Yibin Ying, Luis Rodriguez, KC Ting, and Haifeng Li. Deepcropnet: a deep spatial-temporal learning framework for county-level corn yield estimation. *Environmental Research Letters*, 15(3):034016, 2020.

[22] Jiaxuan You, Xiaocheng Li, Melvin Low, David Lobell, and Stefano Ermon. Deep gaussian process for crop yield prediction based on remote sensing data. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[23] Saeed Khaki and Lizhi Wang. Crop yield prediction using deep neural networks. In *INFORMS International Conference on Service Science*, pages 139–147. Springer, 2019.