CNN-LSTM models show non-linear relationships between weather and US corn yields

Isabella Smythe
Columbia University

Abstract

Understanding the impacts of annual weather and long-term climate change on crop yields is a key area of research, but has often faced tradeoffs between prediction accuracy and model interpretability. This research proposes an approach to bridging that gap using a hybrid CNN-LSTM yield model in conjunction with model interpretability methods. This model reduces out-of-sample RMSE by 10.7% versus a widely used linear baseline, and by 3.6% relative to a linear model using complex hand-engineered features. Key to this improvement is the model's recovery of a number of nuanced weather features that are highly consistent with agronomic theory. The estimated implications of these non-linearities for yields in a changing climate are significant, and represent an important avenue for future work.

1 Introduction

Crop yields have substantial economic and humanitarian significance, but are highly sensitive to weather. Accurate models of how weather affects yield are thus necessary for predicting the impacts of climate change and identifying adaptation strategies. Historically, models have balanced interpretability with predictive accuracy, but "top-down" approaches to feature selection can have downsides: if important features are excluded or relationships are misspecified, understanding of climate impacts will suffer. This work proposes an approach to bridging the gap between traditional parametric models and "black box" deep learning (DL) models. A DL model for yield prediction is presented, and shown to obtain high out-of-sample predictive accuracy. Using interpretable machine learning tools, I demonstrate that (1) the DL model recovers nuanced links between weather and yield that are remarkably consistent with expectations from agronomy; and (2) these relationships have substantial influence on predictions about yields under climate change.

This analysis contributes to unifying modeling approaches from two strands of literature: traditional statistical models focused on estimating the effect of a specific weather feature on yields (e.g. [1], [2], [3], [4], [5], [6], [7]), and machine learning models focused on improving predictive accuracy (e.g. [8], [9], [10], [11]). The models used in this work are consistent with this latter strand of literature; however, I adopt a primary focus on explanation rather than prediction, using tools for model interpretation to identify the structure and importance of the identified relationships.

2 Data

Analysis is conducted for corn in the eastern United States from 2000-2021, a sample of 32,969 county-year pairs in 1,921 distinct counties. Crop yields are available from the US Department of Agriculture [12], and weather is obtained from the Daily Weather Data for Contiguous United States dataset at 2.5 mile resolution [13]. Daily rainfall and high and low temperature are used as inputs to the deep learning model. For linear models (LMs), cumulative rainfall is used along with growing degree days (GDDs) and killing degree days (KDDs), all aggregated from March to August. GDDs are a metric of time spent between 10 and 29°C, calculated as the area beneath the daily temperature

Preprint. Under review.

curve bounded above and below by the temperature range; KDDs are analogous but only bounded from below at 29°C. Weather is aggregated to the county level weighted by fraction of corn cropland in each pixel, per the USDA Cropland Data Layer.

3 Methods

3.1 Model architectures and training

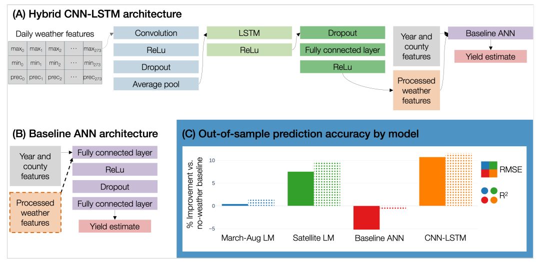

Figure 1: **Model architectures and performance.** Panels A and B show deep learning model architectures. Panel C shows percent improvement in RMSE and R^2 for each model versus a baseline linear model with only county and year controls.

The primary model uses a hybrid CNN-LSTM architecture, shown in Fig. 1 Panel A. Input data includes county dummies and year, along with a 3 by 274 array of daily high and low temperature and precipitation for March through November. All features and labels are scaled to a standard normal distribution, and training uses an Adam stochastic optimizer. The years 2000-2016 are used for training and 2017-2021 for evaluation. One hundred random hyperparameter[1] combinations are evaluated using out-of-sample RMSE (5-fold CV on training data, split by year). Test set predictions are generated using the best-performing hyperparameters, with predictions averaged over 100 models each trained on a random sample with replacement of the training data.

Three LMs and one additional DL model are fit as reference points. The baseline LM and artificial neural network (ANN, Fig. 1 Panel B) are trained using only county and year (i.e., no weather). The March-Aug. LM uses weather data aggregated March through August, and the Satellite LM uses a more complex specification accounting for satellite-estimated location-specific growing season and crop stage-specific heat sensitivity [14].

$$y_{it} = \delta z_{s_i t} + c_i + \beta W_{it} + \varepsilon_{it} \tag{1}$$

Equation 1 shows the general form of the LMs: log yield y_{it} is estimated for county i in year t as a function of $z_{s_i t}$, a quadratic state-level time trend, c_i, a county fixed effect, and optional matrix of weather variables W_{it}. LMs are trained using 100 bagging folds of 5-fold CV with results reported for test years only, as opposed to a train-test split. This avoids overestimating the relative improvement of the DL model, since extrapolating the quadratic time trend can be a substantial source of error.

3.2 Interpretation and climate simulations

Permutation feature importance [15] and partial dependence plots [16] are used to interpret features learned by the DL model. Models are trained in-sample on all data (2000-2021), with results averaged across ten model instances. The LM used for reference is the March-August specification. Yields are re-estimated for four warming scenarios (1.5, 2.7, 3.6, and 4.4°C), simulated as a fixed increase in daily low and high temperature. Two variations on the DL model are used to address possible sources of error in these simulations. First, flexible models may conflate time trends in climate and technological progress [17], [18]; to partially account for this, one model is trained enforcing the time trend learned by the LM. Second, models trained in-sample are likely to be overfit. While accuracy is not evaluated in this stage, similarity of input features between training data and climate simulations may bias estimates towards their corresponding historical reference point. To address this, both baseline and warming scenario yields are re-estimated with 5-fold cross-validation by year.

[1] Learning rate, number of epochs, dropout fraction, CNN kernel dimension and stride, number of LSTM layers, and dimensions of CNN, LSTM, and fully connected layers.

4 Results

4.1 Model performance

The DL model performs best among the models tested, reducing RMSE by 10.7% versus the baseline (LM with no weather data; Fig. 1 Panel C). The baseline ANN performs slightly worse than the linear baseline, though the LM is allowed to "look ahead" when fitting a time trend, while the DL models must extrapolate (if the linear baseline is fit using the same strategy as the ANN, its RMSE increases by nearly 40%). The ANN's performance demonstrates that the improvement of the primary DL model is due to better handling of weather data, not more flexible modeling of controls. The satellite LM is outperformed by the DL model, but does substantially better than the March-August LM.

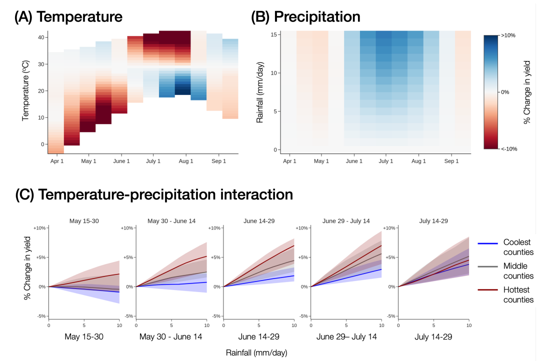

Figure 2: **Impact of weather on yield by time of season.** Panel A shows the percent change in yield as daily high varies from a baseline of 29°C; Panel B is analogous for precipitation versus a no-rain baseline. Panel C plots effects of increased precipitation separately for the coolest, hottest, and middle third of counties.

4.2 Feature importance and impacts

For both DL and linear models, county and high temperature are the most important features. However, the DL model relies more on other weather variables, and much less on time trends. For the LM, the most important features (increase in RMSE from permutation) are: KDDs (120%), county (112%), and year (34%); GDDs and precipitation increase RMSE by <0.5%. For the DL model, results are: county (126%), high temperature (65%), low temperature (27%), precipitation (16%); year has almost no impact. KDDs are an integrated measure taking into account both high and low temperature, potentially accounting for the greater reliance of the DL model on low temperature.

The DL model learns substantially different relationships between temperature and yields through the growing season (Fig. 2 Panel A). The U-shaped curve with optimal temperatures around 29°C found in e.g. [7] is evident in the mid-season, but warmer temperatures are favored during planting (April/May) and dry-down (late August/September). In late July and early August, cooler temperatures are strictly favored. Relationships between rainfall and yields (Fig. 2 Panel B) are more linear, but positive in the mid-season and negative during planting and dry-down. Fig 2 Panel C shows that benefits of rainfall during key periods of the season are much larger in the hottest regions than elsewhere.

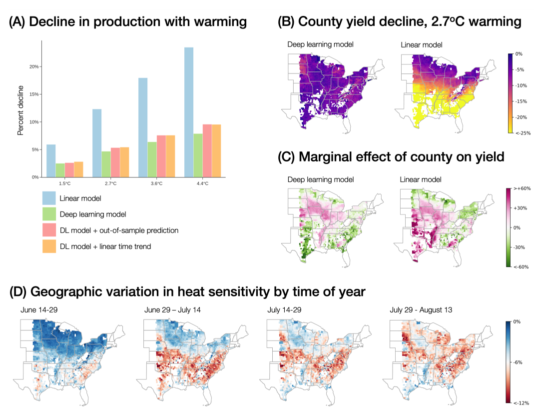

Figure 3: **Geographic variation and yield response to warming.** Panel A shows predicted yield declines for four warming scenarios, as estimated by the LM and three DL model variants; results for the LM and primary DL model are mapped in Panel B for +2.7°C. Panel C compares the impacts of county on yield for LM and DL models, and Panel D shows the DL model's estimated impact of increasing temperature from 29 to 35°C. All maps show percent change in yield.

Estimated county yield effects (Fig. 3 Panel C) are comparable between DL and linear models for much of the country, but the LM finds large positive effects in parts of the South. Fig. 3 Panel D shows substantial geographic variation in DL model estimates of the timing of maximum heat sensitivity, which occurs earlier in the South than further north. There is also variation in overall heat sensitivity; e.g., South Dakota shows high vulnerability, while Iowa and Minnesota appear more robust.

4.3 Climate simulations

The DL models estimate two to three times less yield loss than the LM (Fig. 3 Panel A). Two attempts to address potential sources of error in the DL model slightly reduce this gap, but yield loss remains at most half of LM predictions. Geographic trends diverge substantially between the two approaches (Fig. 3 Panel B): the strong north-south trend predicted by the LM is not evident in the DL estimates, which are driven mainly by location-specific heat sensitivity. However, results should be interpreted with caution, since the complex structure of DL models can pose barriers to causal inference, and the simulated warming is highly simplified.

5 Discussion

While the DL model improves prediction accuracy, this does not necessarily mean it is a better causal model than the LM [19]. Instead, the DL model should be thought of as a tool for hypothesis generation [20], [19], where important features are identified for further causal inference analysis.

The strong alignment between relationships learned by the DL model and agronomic theory is, however, a positive indicator for their usefulness. During planting (April/May), warm and dry conditions are favorable [21], while temperatures of approximately 29°C are optimal in June and early July [7]. Later in July is the silking stage, when crops are most sensitive to heat [1], [22], though these results are novel in their suggestion of harms from temperatures well below 29°C. At the end of the season, yields are less sensitive to weather, but warm and dry conditions assist with drydown [1]. Rainfall is most beneficial in the hottest regions, reflecting combined contributions of hot and dry conditions to water stress [23], [24]. In the South, earlier planting means high temperatures are most detrimental earlier in the season than in the North.

The large disparity in predicted climate impacts is perhaps the most striking result. This is primarily driven by a lower average coefficient on extreme heat in the DL model; the LM learns a larger coefficient, and accounts for high yields in currently-hot areas by assigning them a large positive fixed effect. These effects are cause for some concern, since it is difficult to provide a physical basis accounting for them. The tendency of DL models to underestimate climate impacts is also cause for caution [17], [18], however, though attempts to address this increase estimates only slightly.

The highly simplified nature of these simulations, which assume constant warming across time and space and no effect on precipitation, mean both sets of results should be interpreted carefully. The analysis primarily suggests that climate impacts estimated from reasonable, well-performing models can diverge quite sharply. The precise geographic and intra-seasonal trends in climate change, and the structure of models used for estimation, are likely to have an outsize impact on predictions.

6 Conclusion

This work introduces a simple deep learning model, and demonstrates that it substantially improves prediction accuracy versus LMs and recovers nuanced relationships between weather and yields that are consistent with agronomic theory. This approach may be of particular benefit for crops and regions where hand-engineered features are not well understood. The findings on climate change, which suggest much smaller impacts on future yields than previous estimates, are particularly notable. While caution is required in interpreting these results, they do indicate that more flexible treatment of weather variables can produce dramatically different results. A key area for future work is to quantify the causal effects of these nonlinear features, and leverage them for more precise estimation of the effects of climate change on crop yields.

References

[1] Ethan E Butler and Peter Huybers. Variations in the sensitivity of us maize yield to extreme temperatures by region and growth phase. *Environmental Research Letters*, 10(3):034009, 2015.

[2] Jennifer Hsiao, Abigail LS Swann, and Soo-Hyung Kim. Maize yield under a changing climate: The hidden role of vapor pressure deficit. *Agricultural and Forest Meteorology*, 279:107692, 2019.

[3] Hyungsuk Kimm, Kaiyu Guan, Pierre Gentine, Jin Wu, Carl J Bernacchi, Benjamin N Sulman, Timothy J Griffis, and Changjie Lin. Redefining droughts for the us corn belt: The dominant role of atmospheric

vapor pressure deficit over soil moisture in regulating stomatal behavior of maize and soybean. *Agricultural and Forest Meteorology*, 287:107930, 2020.

[4] David B Lobell, Jillian M Deines, and Stefania Di Tommaso. Changes in the drought sensitivity of us maize yields. *Nature Food*, 1(11):729–735, 2020.

[5] Ariel Ortiz-Bobea, Haoying Wang, Carlos M Carrillo, and Toby R Ault. Unpacking the climatic drivers of us agricultural yields. *Environmental Research Letters*, 14(6):064003, 2019.

[6] AJ Rigden, N vD Mueller, NM Holbrook, Natesh Pillai, and Peter Huybers. Combined influence of soil moisture and atmospheric evaporative demand is important for accurately predicting us maize yields. *Nature Food*, 1(2):127–133, 2020.

[7] Wolfram Schlenker and Michael J Roberts. Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of sciences*, 106(37):15594–15598, 2009.

[8] Andrew Crane-Droesch. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(11):114003, 2018.

[9] Saeed Khaki and Lizhi Wang. Crop yield prediction using deep neural networks. *Frontiers in plant science*, 10:621, 2019.

[10] Saeed Khaki, Lizhi Wang, and Sotirios V Archontoulis. A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science*, 10:1750, 2020.

[11] Jie Sun, Liping Di, Ziheng Sun, Yonglin Shen, and Zulong Lai. County-level soybean yield prediction using deep cnn-lstm model. *Sensors*, 19(20):4363, 2019.

[12] USDA National Agricultural Statistics Service. NASS - Quick Stats. 11 2017.

[13] Wolfram Schlenker. Daily weather data for contiguous united states (1950-2016), 2018.

[14] Anonymous. Documenting climate impacts and adaptation with a remotely sensed dataset of us corn planting and harvest dates. 2024.

[15] Leo Breiman. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231, 2001.

[16] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[17] Tongxi Hu, Xuesong Zhang, Gil Bohrer, Yanlan Liu, Yuyu Zhou, Jay Martin, Yang Li, and Kaiguang Zhao. Crop yield prediction via explainable ai and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield. *Agricultural and Forest Meteorology*, 336:109458, 2023.

[18] Balsher Singh Sidhu, Zia Mehrabi, Navin Ramankutty, and Milind Kandlikar. How can machine learning help in understanding the impact of climate change on crop yields? *Environmental Research Letters*, 18(2):024008, 2023.

[19] Masahiro Ryo. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artificial Intelligence in Agriculture*, 6:257–265, 2022.

[20] James W Murdoch, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu. Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080, 2019.

[21] Daniel W Urban, Michael J Roberts, Wolfram Schlenker, and David B Lobell. The effects of extremely wet planting conditions on maize and soybean yields. *Climatic Change*, 130:247–260, 2015.

[22] Berta Sánchez, Anton Rasmussen, and John R Porter. Temperatures and the growth and development of maize and rice: a review. *Global change biology*, 20(2):408–417, 2014.

[23] David B Lobell, Marianne Bänziger, Cosmos Magorokosho, and Bindiganavile Vivek. Nonlinear heat effects on african maize as evidenced by historical yield trials. *Nature climate change*, 1(1):42–45, 2011.

[24] David B Lobell, Graeme L Hammer, Greg McLean, Carlos Messina, Michael J Roberts, and Wolfram Schlenker. The critical role of extreme heat for maize production in the united states. *Nature climate change*, 3(5):497–501, 2013.