Corn Yield Prediction in US Midwest Using Artificial Neural Networks

Somya Sharma and Snigdhansu Chatterjee[0000−0002−7986−0470]

University of Minnesota, Twin Cities, MN 55455 USA
{sharm636, chatt019}@umn.edu

Abstract. Climate change driven increment in temperature and variations in weather have affected agrarian economies throughout the world. Due to this temperature change-fueled uncertainty in agricultural yield, it becomes imperative to study the dependence of yield on meteorological factors. Deep learning architectures offer a way to clearly define this relationship through non-linear function approximations. In this study, we offer a comparison of deep learning with other popular data driven methods and outline a concrete dropout based Bayesian uncertainty estimation of yield predictions.

1 Introduction

A primary goal of smart and precision agriculture is to be able to predict the yield of agricultural produce at various time scales. This is especially important in view of the changing meteorological patterns owing to global warming, since climate and weather events affect food security all over the world. For example, extreme weather events in the US midwest affect crop yields, food price hikes and can lead to production losses – the effect being as high as 75% for some Minnesota counties [30].

In this paper, we propose using ***deep learning*** to predict county-level corn yield. The dataset comprises of county-level daily weather variables (each county corresponds with one training sample), and crop yield information for corn produce in Minnesota and Illinois for 2012. There are four major sets of features - daily maximum temperature, minimum temperature, average precipitation and geographic coordinates. Therefore, there are 365 features corresponding to maximum temperature, 365 minimum temperature readings and 365 average precipitation readings for each county. We explicitly model the spatial dependence in the data by using averaged climate variables from three nearest neighboring counties as additional features, and use a non-linear activation function. Using these 2200 features (1100 features from the county, additional 1100 from neighboring counties), we predict the end of year crop yield response for each county. Such spatial dependence and the non-linear relationships between the climatic variables and crop yield is learned in the *artificial neural network* (ANN) architecture using a Bayesian optimization approach. For comparison, we also study a few additional machine learning techniques, including penalized regression, support vector regressions with different kernels, random forest techniques, and a shallow ANN architecture with concrete dropout. Our proposed ANN method outperforms these rival techniques in terms of test data mean squared error (MSE) and R^2-value, as can be seen from Table 1.

Next, we do a deep dive into the shallow ANN model from Table 1, to assess the gaps that remain in the predictive processes after shallow model fitting. Here, we discover some interesting details based on the concrete dropout-based uncertainty quantification scheme. The uncertainties depend on the temperature and location, which suggests a deep architecture may be a necessary component for a model of sufficiently high predictive capability.

After a summary review of the current state of the literature in crop yield modeling in Section 2, we present our findings in Section 3. Additional discussions, including questions for future research are briefly presented in Section 4.

2 Related work

Historically, process-based biophysical models and classical statistical models have been employed for crop yield prediction. Process based models [17, 22, 25, 26, 32] study physiological and physical processes to simulate crop yield. Often, simpler statistical models [17, 23, 36] are used owing to their straightforward reporting of goodness of fit metrics. Many of these models may be constrained by their strict and often unrealistic assumptions to control multi-collinearity and spatio-temporal dependence [37]. Additionally, process-based and simple statistical models often miss or exclude non-linear terms, which may prevent them for being useful for yield predictions under extreme climatic conditions.

Machine learning methods present the opportunity to model agricultural data using more complex architectures, using fewer assumptions, and on larger datasets. Artificial neural networks [3, 4, 6, 8, 21, 27–29, 35], linear regression [3, 12, 19, 24, 29, 40], tree based models [7, 12, 16, 29, 33, 34, 38, 40] and support vector machines [11, 13, 34] are some of the most used machine learning algorithms [18, 20] for crop yield modeling. In particular, ANNs have been used for tasks like crop yield prediction, species recognition, weed detection or crop quality assessment in [3, 4, 6, 8, 21, 27–29, 35] and elsewhere, using a variety of complex features including satellite data.

Note that most process-based, machine learning and statistical methods fail to capture the spatio-temporal dependency in the data, since the default shallow learning architectures that are used typically correspond to independent data. Spatial and spatio-temporal statistical models may be used to capture such dependencies explicitly, but are very sensitive to the stringent assumptions made for such models, and the computations do not scale with data size. Ignoring positive spatio-temporal correlation may lead to considerable over-estimation of the uncertainty in prediction, and loss of statistical power for feature selection and risk bounds, which has severe consequences for downstream industries like that of crop insurance. More alarmingly, ignoring negative spatio-temporal correlation may lead to misleadingly narrow uncertainty bounds that may be centered at an inaccurate and biased estimate. In view of these issues, this paper focuses on using easily available within-season meteorological variables for rapid and efficient usability, but ensures that the ANN captures spatio-temporal dependencies as well as non-linear functional components.

3 Results from the data analysis

Table 1: *Comparison of different machine learning models on the test data.*

Model	Test MSE	R^2
Linear Regression (Lasso)	2.3432	0.7205
Random Forest	2.1113	0.7481
Support Vector Regressor (rbf kernel)	2.3	0.7246
Support Vector Regressor (polynomial kernel, degree: 8)	4.2943	0.4878
Concrete Dropout, 3-layer ANN	3.0001	0.6379
Neural Network	1.9224	0.7684

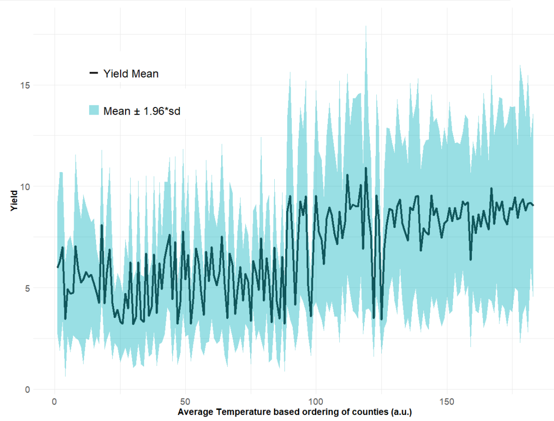

Fig. 1: *Yield and uncertainty estimates by county.*

We report in Table 1 several commonly used models that are used for data-driven inference about crop growth. First, we implement a LASSO [39] regression approach, which imitates the generalized linear regression models that are popularly used in many of the related domain science outlets. The regularization coefficient at 0.0781 is learned through 5-fold cross validation. Based on the LASSO regularization and resampling inference on deep, semi-parametric modeling (not shown here), we can infer a strong negative correlation between agricultural yields and harvest-period precipitation. Next, we use a random forest model with 100 tress and max depth 100 per tree. Next, a support vector regressor with radial-basis kernel and $C = 10$ and margin of error $\epsilon = 0.1$ is learned. To enable comparison with [36], we also fit a support vector model of 8^{th} degree, with $C = 1.0$. A shallow ANN model using three layers, with concrete dropout, is also fitted to the data. The proposed deep-ANN architecture is then used, which results in the best MSE and R^2 values on test data.

Unlike the conventional use of dropout to improve generalization power by sampling neurons during training, we can derive an empirical predictive distribution by using the layer-wise dropout relaxation during the testing process

[10]. We use this approach for the shallow ANN reported in Table 1, for a three-layer ANN. The layer wise dropout rates are set to trainable and learned via the standard back-propagation process along with other neural network parameters.

Figure 1 showcases the per county mean estimates and confidence interval of corn crop yield using this concrete dropout-based shallow ANN. The counties are ordered in ascending order of average yearly temperature. In Figure 1, variance of the predictive distribution is generated by randomly dropping neurons at test time using the learned dropout rate. The figure shows that for counties with higher temperatures, the uncertainty is also higher and confidence bands are larger. Similarly, Figure 2 compares the mean estimate with observed yield and provides an overview of uncertainty as well. Here, we notice that there is a spatial dependence in the residuals as well, which suggests that shallow architectures may not be adequate for fully capturing the complexities of crop yield data.

4 Discussion and Results

In our study we primarily employ a deterministic deep neural network that learns the non-linear relationship between meteorological factors and the corn crop yield. The architecture is learned through Bayesian optimization to minimize mean squared error. As compared to popular baseline models, the deep-ANN performs well in picking up the county specific variations in crop growth. Presumably, our deep-ANN model performs better because of two factors that other machine learning methods fail to capture: (i) nonlinear functional relationship between the features and crop yield values, and (ii) spatio-temporal dependencies in the data. Based on Table 1, it seems the latter cause is stronger than the former.

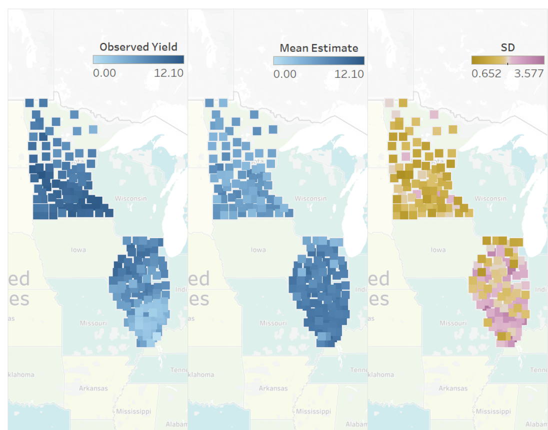

Fig. 2: *Yield and uncertainty estimates by county.*

The study that we present here is illustrative, to show the strengths and limitations of different architectures for modeling crop yields. In future, we will extend our deep-ANN architecture to include systematic uncertainty quantification and analysis. We will also use additional features like solar radiation and soil composition, and satellite data as and where available. Such additional features can only strengthen the proposed deep-ANN architecture, which is already performing better than rival techniques that are popularly used in this domain.

Bibliography

[1] CPC global temperature data provided by the NOAA/OAR/ESRL PSL, boulder, colorado, USA, from their web site at https://psl.noaa.gov/ https://psl.noaa.gov/

[2] USDA maize yield data https://www.nass.usda.gov/Statistics_by_Subject/index.php?sector=CROPS

[3] Ahamed, A.M.S., Mahmood, N.T., Hossain, N., Kabir, M.T., Das, K., Rahman, F., Rahman, R.M.: Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh. In: 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). pp. 1–6. IEEE (2015)

[4] Baral, S., Tripathy, A.K., Bijayasingh, P.: Yield prediction using artificial neural networks. In: International Conference on Advances in Communication, Network, and Computing. pp. 315–317. Springer (2011)

[5] Chatfield, K., Lempitsky, V.S., Vedaldi, A., Zisserman, A.: The devil is in the details: an evaluation of recent feature encoding methods. In: BMVC. vol. 2, p. 8 (2011)

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

[7] Everingham, Y., Sexton, J., Skocaj, D., Inman-Bamber, G.: Accurate prediction of sugarcane yield using a random forest algorithm. Agronomy for sustainable development **36**(2), 27 (2016)

[8] Fernandes, J.L., Ebecken, N.F.F., Esquerdo, J.C.D.M.: Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. International Journal of Remote Sensing **38**(16), 4631–4644 (2017)

[9] Frazier, P.I.: A tutorial on Bayesian optimization. arXiv preprint arXiv:1807.02811 (2018)

[10] Gal, Y., Hron, J., Kendall, A.: Concrete dropout. In: Advances in neural information processing systems. pp. 3581–3590 (2017)

[11] Gandhi, N., Armstrong, L.J., Petkar, O., Tripathy, A.K.: Rice crop yield prediction in India using support vector machines. In: 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE). pp. 1–5. IEEE (2016)

[12] Goldstein, A., Fink, L., Meitin, A., Bohadana, S., Lutenberg, O., Ravid, G.: Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge. Precision agriculture **19**(3), 421–444 (2018)

[13] González Sánchez, A., Frausto Solís, J., Ojeda Bustamante, W., et al.: Predictive ability of machine learning methods for massive crop yield prediction (2014)

[14] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
[15] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
[16] Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.M., Gerber, J.S., Reddy, V.R., et al.: Random forests for global and regional crop yield predictions. PLoS One **11**(6), e0156571 (2016)
[17] Jones, P.G., Thornton, P.K.: The potential impacts of climate change on maize production in Africa and Latin America in 2055. Global environmental change **13**(1), 51–59 (2003)
[18] van Klompenburg, T., Kassahun, A., Catal, C.: Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture **177**, 105709 (2020)
[19] Kouadio, L., Deo, R.C., Byrareddy, V., Adamowski, J.F., Mushtaq, S., et al.: Artificial intelligence approach for the prediction of robusta coffee yield using soil fertility properties. Computers and Electronics in Agriculture **155**, 324–338 (2018)
[20] Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D.: Machine learning in agriculture: A review. Sensors **18**(8), 2674 (2018)
[21] Lin, T., Zhong, R., Wang, Y., Xu, J., Jiang, H., Xu, J., Ying, Y., Rodriguez, L., Ting, K., Li, H.: DeepCropNet: a deep spatial-temporal learning framework for county-level corn yield estimation. Environmental Research Letters **15**(3), 034016 (2020)
[22] Liu, X., Chen, F., Barlage, M., Zhou, G., Niyogi, D.: Noah-MP-Crop: Introducing dynamic crop growth in the Noah-MP land surface model. Journal of Geophysical Research: Atmospheres **121**(23), 13–953 (2016)
[23] Lobell, D.B., Burke, M.B.: On the use of statistical models to predict crop yield responses to climate change. Agricultural and Forest Meteorology **150**(11), 1443–1452 (2010)
[24] Matsumura, K., Gaitan, C.F., Sugimoto, K., Cannon, A.J., Hsieh, W.W.: Maize yield forecasting by linear regression and artificial neural networks in Jilin, China. The Journal of Agricultural Science **153**(3), 399–410 (2015)
[25] McCown, R.L., Hammer, G.L., Hargreaves, J.N.G., Holzworth, D.P., Freebairn, D.M., et al.: APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. Agricultural systems **50**(3), 255–272 (1996)
[26] McDermid, S.P., Ruane, A.C., Rosenzweig, C., Hudson, N.I., Morales, M.D., Agalawatte, P., Ahmad, S., Ahuja, L., Amien, I., Anapalli, S.S., et al.: The AgMIP coordinated climate-crop modeling project (C3MP): methods and protocols. In: HANDBOOK OF CLIMATE CHANGE AND AGROECOSYSTEMS: The Agricultural Model Intercomparison and Improvement Project Integrated Crop and Economic Assessments, Part 1, pp. 191–220. World Scientific (2015)

[27] Mkhabela, M., Bullock, P., Raj, S., Wang, S., Yang, Y.: Crop yield forecasting on the Canadian prairies using MODIS NDVI data. Agricultural and Forest Meteorology **151**(3), 385–393 (2011)

[28] Pantazi, X.E., Moshou, D., Mouazen, A.M., Kuang, B., Alexandridis, T.: Application of supervised self organising models for wheat yield prediction. In: IFIP International Conference on Artificial Intelligence Applications and Innovations. pp. 556–565. Springer (2014)

[29] Rahman, M.M., Haq, N., Rahman, R.M.: Machine learning facilitated rice prediction in Bangladesh. In: 2014 Annual Global Online Conference on Information and Computer Technology. pp. 1–4. IEEE (2014)

[30] Ray, D.K., Gerber, J.S., MacDonald, G.K., West, P.C.: Climate variation explains a third of global crop yield variability. Nature communications **6**(1), 1–9 (2015)

[31] Ray, D.K., West, P.C., Clark, M., Gerber, J.S., Prishchepov, A.V., Chatterjee, S.: Climate change has likely already affected global food production. PloS one **14**(5), e0217148 (2019)

[32] Riha, S.J., Wilks, D.S., Simoens, P.: Impact of temperature and precipitation variability on crop model predictions. Climatic change **32**(3), 293–311 (1996)

[33] Romero, J.R., Roncallo, P.F., Akkiraju, P.C., Ponzoni, I., Echenique, V.C., Carballido, J.A.: Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires. Computers and electronics in agriculture **96**, 173–179 (2013)

[34] Ruß, G., Kruse, R.: Regression models for spatial data: An example from precision agriculture. In: Industrial Conference on Data Mining. pp. 450–463. Springer (2010)

[35] Ruß, G., Kruse, R., Schneider, M., Wagner, P.: Data mining with neural networks for wheat yield prediction. In: Industrial Conference on Data Mining. pp. 47–56. Springer (2008)

[36] Schlenker, W., Roberts, M.J.: 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)

[37] Sheehy, J.E., Mitchell, P.L., Ferrer, A.B.: Decline in rice grain yields with temperature: Models and correlations can give different estimates. Field crops research **98**(2-3), 151–156 (2006)

[38] Shekoofa, A., Emam, Y., Shekoufa, N., Ebrahimi, M., Ebrahimie, E.: Determining the most important physiological and agronomic traits contributing to maize grain yield through machine learning algorithms: A new avenue in intelligent agriculture. PloS one **9**(5), e97288 (2014)

[39] Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) **58**(1), 267–288 (1996)

[40] Zhong, H., Li, X., Lobell, D., Ermon, S., Brandeau, M.L.: Hierarchical modeling of seed variety yields and decision making for future planting plans. Environment Systems and Decisions **38**(4), 458–470 (2018)

Supplementary Material

Dataset

Depending on the temporal scale and objective of prediction, the predictions can either be pre - season forecasting before or during sowing, within-season predictions or long - term seasonal climate forecasts. Our study focuses on within - season prediction and presents the opportunity to forecast the end of year yield. The feature and yield spatial distribution is displayed in 3

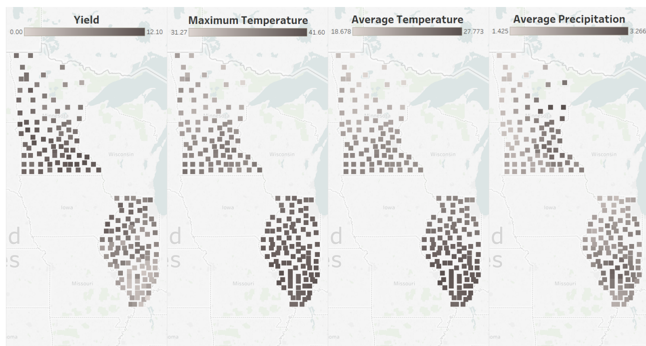

Fig. 3: Data Features and Yield Response

A negative correlation between the amount of variation in extreme temperature and yield is significantly evident for each county. Above, is a figure depicting the spatial distribution of maximum temperature, average temperature, average precipitation and end of year crop yield for the year 2012. The meteorological data come from the Physics Sciences Laboratory climate dataset [1] and yield data is derived from the USDA corn field crop yield data[2, 31]. K nearest neighbor approach is used to resample the half degree spatial resolution data (5 minute grid cells) to obtain average maximum temperature, minimum temperature and average precipitation values for each county and each day. This leads to one county level reading for each feature and output variable. Corn yields are measured in metric tons per hectare.

Neural Network and Hyper - parameter search

For a resource - efficient hyper-parameter search, we employ Gaussian Processes based optimization. This involves using expected improvement as the surrogate acquisition function as objective function for the hyper-parameter optimization. Eventually, over several trials with different hyper-parameter sets, the model can be expected to accurately represent the maximized original objective function.

$$\text{Expected Improvement}_n(x) = E_n[[f(x) - f_n^*]^+],$$

is the expected improvement for $(x_i, y_i)_{i=1}^n$ samples and $[a]^+ = max(0, a)$ [9].

Neural Networks have been successfully used in several previous studies for non linear response prediction and are also used in this study to model crop yield as a function of the daily variations in meteorological features and spatial coordinates. We use Bayesian hyper-parameter optimization to explore the hyper - parameter space. The number of layers, number of hidden units per layer, dropout rate for each layer, activation function for each layer, optimization method for gradient ascent, learning rate are the hyper - parameters that are varied to learn the best set. Over 10,000 trials to search for optimal hyper-parameters, we learn a 14 layer sequential deterministic Neural Network with ReLU activation and dropout after every layer. We use adaptive momentum for model optimization with a learning rate of 1e-5. Learning rate is decayed by $1/10^{th}$ after every 1000 iterations. Here, the Deterministic Neural Network model is a sequential Neural Network with 14 fully connected layers. The input is 1100 features with temporal temperature and precipitation information along with spatial coordinates corresponding to the centroid of each of the counties. Each layer has ReLU activation and drop out regularization. There are 1,106,483 trainable parameters.

DNN performs well in our study. The Deterministic Neural Network architecture is given below. It has 14 fully connected layers in a sequence with 1,106,483 layers.

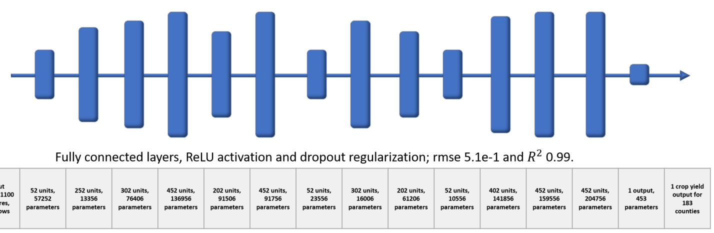

Fig. 4: DNN architecture with training set performance metrics

Apart from the varying drop out rate after each layer that improve the generalization power of the model, through Bayesian optimization, the search also yields an architecture that comprises of bottleneck hidden layers with relatively less hidden units at a few places in between bigger hidden layers [5, 14, 15]. This helps in ensuring that the model does not grow overly complex and captures any lower dimensional relations among the climate and crop growth.