What can crop modelers learn from machine learning models about corn, sorghum and soybean?

Armen Kemanian, Alexis Hoffman, Forest Chris

What can crop modelers learn from machine learning models about corn, sorghum and soybean?

Kemanian Armen[1] (kxa15@psu.edu), Hoffman Alexis[2], Chris Forest[2]

[1] *Plant Science, The Pennsylvania State University, University Park, PA, US*; [2] *Meteorology, The Pennsylvania State University, University Park, PA, US*

Background

An ingenious statistical analysis by Schlenker and Roberts (2009) of the county-level grain yield of cotton, corn and soybean in response to climate showed that these non-controlled experiments contain valuable and somewhat hidden information. Critically, these authors identified a temperature range over which grain yields increase (≈10 to 29°C for corn for corn) after which grain yields decrease sharply. Analogous analyses have been presented by Lobell et al. (2014) and Hoffman et al. (2017).
Our goal was to apply machine learning (ML) tools to data panels like those used by Schlenker and Roberts (2009) to reveal the relationship between grain yield and climate variables for specific phases of the crop cycle. We analyzed the rainfed yield of corn, sorghum and soybean in response to climate in the US using Random Forest (RF), a non-parametric machine learning (ML) tool (Breiman, 2001). Unlike other ML tools, RF allows gleaning the functional form between predicted and predictor variables.

Methods

For the period 1980-2016, we extracted county-level yield for corn, sorghum and soybean (≈30,000, 18,000 and 27,000 records, respectively) from the USDA-NASS database for 1631 US Great Plains counties. This data was paired with the 4-km resolution grid from MetData (Abatzoglou, 2011). We computed climate predictors (temperature = T, solar radiation = SR, precipitation = PP, vapor pressure deficit = D) for three growth phases and for the entire crop cycle for each species. The crop phases were establishment, critical window pre-post flowering and grain filling. Planting dates were based on temperature and varied by year and county. Each crop yield was modeled using RF with year and all climate variables in each phase and in the growing season as predictors.

Results and Discussion

The RF models captured 86%, 71% and 81% of the variance in yield for corn, sorghum, and soybean. The yield dependence on time was linear (90 kg ha-1 y-1 for corn and 29 kg ha-1 y-1 for sorghum and soybean). T, PP and D related to yield; the magnitude and thresholds varied by crop. Minimum T affected yield positively roughly until 19°C but decreased quickly below 14°C for sorghum. Maximum T thresholds after which yield decreased were ≈29, 32 and 29 for corn, sorghum and soybean. The cumulative precipitation levels that maximized yield were 628, 420 and 494 mm; corn was responsive to cumulative precipitation than to precipitation in any given phase, while sorghum and soybean responded mostly to precipitation in the grain filling phase. All crops responded negatively to the D, but the most striking response was that of corn, whose yield decreased sharply as the D increased during flowering. This response was likely independent of soil moisture as precipitation already accounted for the water supply.

Conclusion

This analysis provides a benchmark against which process-based models can be compared: were we to create yield x climate panels based on simulation models for the same region, the results of an RF analysis should be similar to those presented here. The RF allowed identifying critical T and D threshold that can guide research in this area. The sharp response to D of corn during grain filling and the positive response of sorghum to minimum T increases as well as a higher tolerance to high T provide clear food for thought for modelers. It remains to be seen if ML can provide information on expected production outputs in a changing climate.

…/…

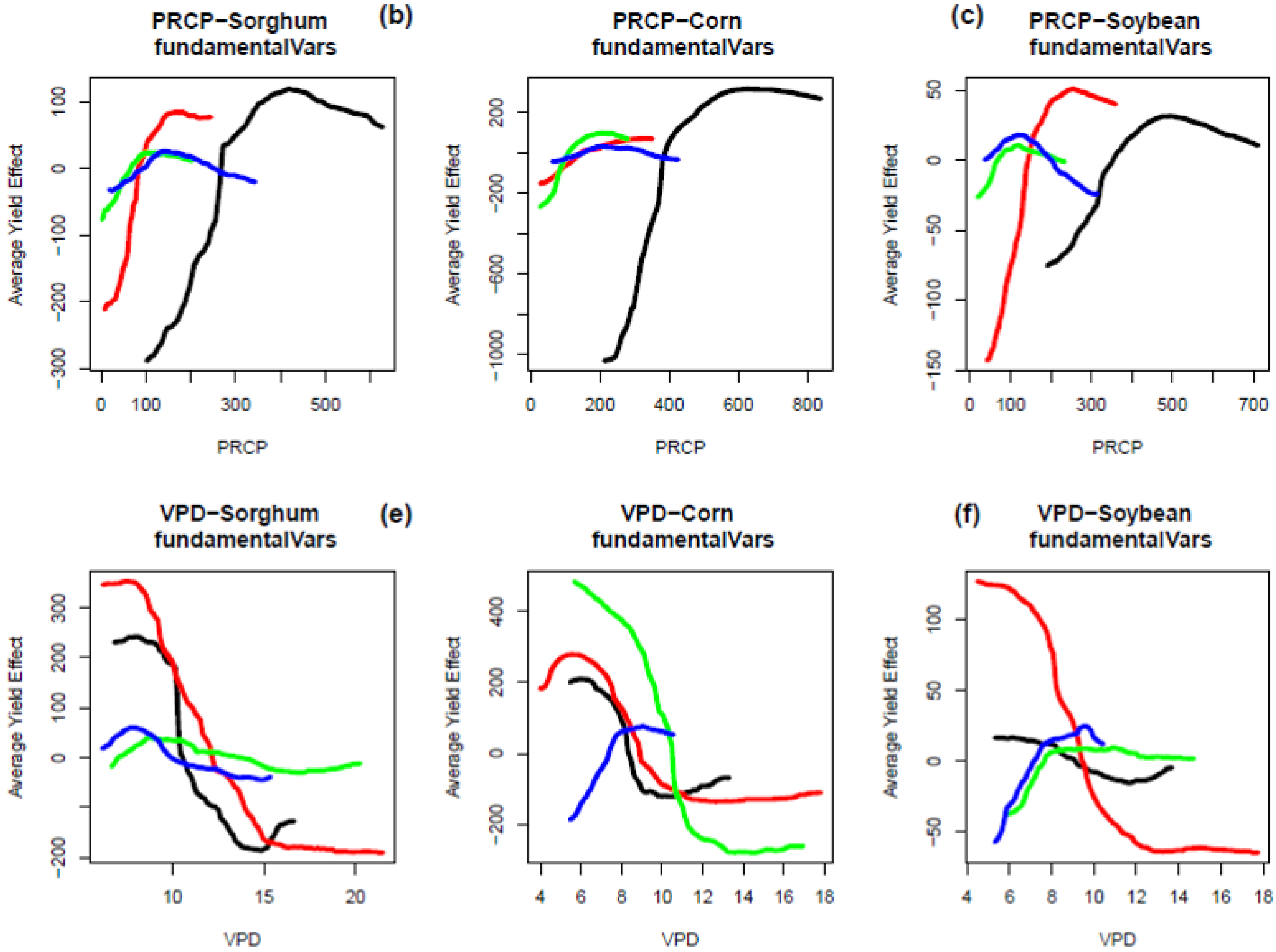

Figure 1. Partial dependence plot of the yield versus precipitation and vapor pressure deficit for sorghum (a, d), corn (b, e) and soybean (c, f). Line color code: blue = establishment, green = flowering, red = grain filling, black = growing season.

Keywords: machine learning, climate, corn, sorghum, soybean.

References:

1. Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32.
2. Hoffman, A.L., Kemanian, A.R. and Forest, C.E. (2018). Analysis of climate signals in the crop yield record of sub-Saharan Africa. Global Change Biology, 24, 143-157.
3. Lobell, D.B., Roberts, M.J., Schlenker, W., Braun, N., Little, B.B., Rejesus, R.M. and Hammer, G.L. (2014). Greater sensitivity to drought accompanies maize yield increase in the US Midwest. Science, 344, 516-519.
4. Schlenker, W. and Roberts, M.J. (2009). Nonlinear temperature effects indicate severe damages to US crop yields under climate change. Proceedings of the National Academy of Sciences, 106, 15594-15598.