A machine learning interpretation of the contribution of foliar fungicides to soybean yield in the north-central United States

Denis A. Shah, Thomas R. Butts, Spyridon Mourtzinis, Juan I. Rattalino Edreira, Patricio Grassini, Shawn P. Conley & Paul D. Esker

IN A BEAN POD:

- Foliar fungicide usage in soybeans in the north-central United States increased steadily over the past two decades.
- Foliar fungicides ranked 7th out of 20 factors in terms of relative importance explaining soybean yield.
- Using foliar fungicides in late-planted fields and in lower latitudes realized a larger yield benefit.
- Less than a 1.5 bu/ac yield penalty for not using foliar fungicides was observed in high-yielding environments.
- Except in a few production environments, yield gains due to foliar fungicides sufficiently offset the associated costs when soybean prices are near-to-above average.

INTRODUCTION

Soybean (*Glycine max*) is one of the major crops produced in the United States (U.S.). Success in growing soybean depends on multiple management decisions, which rest largely on the individual grower or crop manager. One of these choices is the use of foliar fungicide and/or insecticide. The decade from 2005 to 2015 saw the use of foliar fungicides in U.S. soybeans double on a per unit area basis, and almost triple in terms of total product applied across all so-treated fields. Foliar fungicide applications are not necessarily made in response to the actual threat or presence of diseases; prophylactic applications may be made to the perceived future possibility of disease (sometimes as an insurance spray) or for so-called plant health benefits (e.g., a “greening effect”). The accumulated body of evidence to date does show that foliar diseases are responsible for measurable financial losses. Yet at the same time, foliar diseases in soybean are, except in a few circumstances, rarely severe when compared to losses due to soilborne pathogens. When foliar diseases are absent or at low levels, the consensus from recent field trials is that the yield response to foliar fungicides (including the plant health benefit effect) are not sufficient to offset the interventional costs.

The increase in foliar fungicide use in U.S. soybeans does therefore seem to contradict the scientific research showing low economic returns when disease levels are low or absent. A partial explanation may be that the myriad of soybean crop management choices makes it impossible to account for complexity beyond three-way interactions in designed field trials which are by practical necessity focused on a few controlled main effects of interest. Moreover, such trials are conducted in a few locations at best, which raises questions about the scalability of inference beyond local conditions. Therefore, it is not uncommon for inferences made from research trials to conflict across studies.
A novel complementary approach to traditional field experiments, given their limited design and inferential space, uses grower-supplied data linked in a spatial framework to other data layers representing soil properties and weather. This approach leads to an observational database covering wide and diverse geographies, is broad in scope, and possibly capturing complex, realistic interactions among agronomic, environmental and crop management variables beyond those which may be represented in designed field trials. The challenge, however, is that the multidimensional observational space must now be queried for pattern recognition and for drawing inferences from those identified relationships. This usually requires a machine-learning (ML) approach rather than traditional statistical methods.

In this paper, a ML algorithm was used to fit a yield prediction model to a grower-derived database on soybean production practices in the north-central U.S. The model was then queried with the objective of understanding how foliar fungicides fit into overall soybean production practices in the north-central U.S. and their contribution to yield from an economic standpoint.

**METHODS**

Soybean grower-supplied agronomic practices and average yield for 2,738 non-irrigated soybean fields in the years 2014 to 2016 across 11 states in the U.S. north-central region (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin) were collected. The grower-supplied data were augmented with variables representing technology extrapolation domains (TEDs) which define regions with similar climate and soils; as well with soil properties data. This data structure was a fusion from different sources linked by GPS coordinates. Growers did not report on product name, chemistry, or rates of application for any of the pest control inputs they used (fungicidal, insecticidal, nematicidal, whether seed or foliar applied), and therefore the only level of detail available was whether such products were used or not.

For the machine learning modeling, the data matrix was split (80:20) into training (2,191 observations) and test (547 observations) sets. The training set was used to tune a random forest (RF) model with soybean yield as a continuous response to the 20 variables as predictors. The tuned RF model was evaluated by predicting yield on the test set, after which it was refit to the full data matrix. The RF model was then interpreted using model-agnostic approaches. Feature importance (FI) was summarized visually by plotting the median of FI, and the 5% and 95% quantiles.

For local model interpretation, the goal was to compute the contributions of the features based on the difference between the predicted yield for a single field and the global average, with an emphasis on the impact of foliar fungicide use in soybean fields. For any one observation, the Shapley values ($\phi$) values are an estimate of how much a predictor contributed to the difference between an individual field’s predicted yield and the predicted yield averaged across all fields in the data matrix. We studied the Shapley values within different subsets of fields in the data matrix.

Shapley values were studied within different subsets ($s$) of fields, consisting of different cohorts ($c$) described as follows. In subset 1 ($s_1$), cohorts were selected from each of the 12 TEDs with the most fields in the data, where within each of those TEDs the 1$^{st}$ cohort consisted of the 20 highest-yielding fields among those sprayed with foliar fungicides ($s_{1c_1}$) and the 2$^{nd}$ cohort consisted of the 20 highest-yielding fields among those which were not sprayed ($s_{1c_2}$). The four cohorts of subset 2 were the 100 highest-yielding fungicide-treated (sprayed) fields ($s_{2c_1}$), the 100 lowest-yielding sprayed fields ($s_{2c_2}$), the 100 highest-yielding unsprayed fields ($s_{2c_3}$), and the 100 lowest-yielding unsprayed fields ($s_{2c_4}$), among all fields. Subset 3 had two cohorts, chosen from all fields in the data: the 90th percentile for yield among sprayed fields ($s_{3c_1}$); and the 90th percentile for yield among the unsprayed fields ($s_{3c_2}$). A final subset ($s_4$) consisted of two cohorts, the first being the 210 fields which had been sprayed with foliar fungicides but not with foliar insecticides ($s_{4c_1}$). The second cohort of $s_4$ ($s_{4c_2}$) was a random sample of 210 of the 623 fields which had been sprayed with both foliar fungicides and foliar insecticides, with yields restricted to be within the range of yields in $s_{4c_1}$. 
For each subset, the $\phi$ values associated with using foliar fungicides were interpreted as follows. If foliar fungicide applications had no effect, then the $\phi$ value for that feature would be zero for the field. If the predicted yield for a sprayed field was greater than the global average yield, then a positive fungicide $\phi$ value was an estimate of how much of the yield increase (above the global average) was due to fungicide application. If, however, a sprayed field’s yield was below the global average then a positive fungicide $\phi$ value estimated how much the spray contributed to raising the yield in a situation in which other features contributed more heavily to a yield reduction (to below the global average). That is, the fungicide was not able to counterbalance the negative effects that other features had on yield. For any sprayed field, a negative fungicide $\phi$ value would indicate a yield reduction (loss) due to spraying, perhaps due to very high disease pressure or wheel damage. Finally, for unsprayed fields a positive $\phi$ value for the fungicide feature would counterintuitively indicate that yield benefited from not spraying, whereas a negative $\phi$ value for the fungicide feature would estimate how much yield was penalized by not applying a foliar fungicide.

The Shapley $\phi$ values associated with foliar fungicide use were used in a partial economic analysis to estimate the net profit (loss) realized by applying foliar fungicides to the soybean crop. Soybean price (price) was fixed at the price as of Jan 31, 2021 (US$15.68/bu). The combined cost of fungicide plus its application (chem.cost) was also held fixed, at US$25.05/ac.

**RESULTS AND DISCUSSION**

The surveyed, rain-fed commercial soybean fields were spread across the U.S. north-central region (Figure 1) with a latitudinal gradient evident for maturity group (MG). Among the 2,738 fields, 833 (or 30.4%) were sprayed with foliar fungicides. Of the 833 fields sprayed with foliar fungicides, 623 (74.8%) had also been sprayed with foliar insecticides.

Location (latitude; a surrogate for other unmeasured variables) and sowing date (day of year from Jan 01) were the two variables most associated with yield (Figure 2), consistent with the central importance of early planting to soybean yield. Soil-related properties (pH and organic matter content of the topsoil) were also associated with yield (Figure 2). Management-related variables such as foliar fungicide, insecticide and herbicide applications were of intermediate importance, and other management variables (row spacing, seed treatments, starter fertilizer) were on the lower end of the importance spectrum in predicting soybean yield (Figure 2). Insecticide and fungicide seed treatments were poorly associated with soybean yield increases as has been previously shown. The relatively lower importance of row spacing is consistent with previous analyses of this variable from soybean grower data. The dataset we analyzed did not contain enough observations to include artificial drainage as a variable, which has been shown to influence soybean yield, presumably by allowing earlier sowing.
Examination of the interactions among the variables showed that the yield difference between sprayed and unsprayed fields increased with later sowing, indicative of a greater fungicide benefit in later-planted fields (Figure 3). This would seem to conflict with the results of a recent meta-analysis in which soybean yields responded better when foliar fungicides were applied to early-planted fields \(^{27}\), but in that study, there was also the confounding effect of higher-than-average rainfall between sowing and the R3 growth stage. With respect to latitude, the difference in yield between sprayed and unsprayed fields decreased as one moved further north (Figure 3), suggesting that foliar fungicides were of more benefit when applied to the more southerly located fields, which do tend to experience more or prolonged conditions conducive to foliar diseases than the northern fields \(^{22,24}\).

Focusing on model interpretation at the local level, we examined the Shapley \(\varphi\) values associated with foliar fungicide applications for different subsets of fields within the data. Predicted yields within the two cohorts of subset 1 were mainly above the global average of 56.3 bu/ac (hence the single field predicted yield – the global average was greater than zero), except in TED 602303 (Figure 4), which corresponded to fields in North Dakota.
In most cases Shapley $\phi$ values for foliar fungicide use exhibited a positive contribution to the yield above the global average. If these cohorts of fields represented high-yielding environments within each TED, then foliar fungicide sprays contributed positively up to 4.5 bu/ac in the yield increase above the global average in $s_{c_1}$. However, among high-yielding fields in $s_{c_2}$, the penalty for not spraying was less than 1.5 bu/ac. This finding supports the contention that fungicide sprays are most worthwhile in high-yielding environments.

The Shapley $\phi$ values for fungicide use were well-separated among the four cohorts of fields of subset $s_2$ (Figure 5). The fields within $s_2$ were selected across the entire dataset and not by TED membership. The lowest-yielding fields ($s_{c_2}$ & $s_{c_4}$) were all below the global yield average, whereas the converse was true of the highest-yielding fields ($s_{c_1}$ & $s_{c_3}$). Among the lowest-yielding fields, foliar fungicides were mainly associated with a positive, but less than 3 bu/ac, effect on yield ($s_{c_2}$), and other factors were responsible for dropping a field’s yield to below the global average. Amongst the highest-yielding fields ($s_{c_1}$), foliar fungicides were associated with between 2.2 and 5.2 bu/ac of the yield above the global average. These Shapley $\phi$ values for the contribution of foliar fungicides are consistent with estimates of the yield response to foliar fungicides from a meta-analytic perspective. Given that the

**Figure 4.** Shapley phi values attributed to foliar fungicide use for two cohorts of fields within the 12 technology extrapolation domains (TEDs) with the most fields. Within each TED, the cohorts are the 20 highest-yielding fields among those sprayed with foliar fungicides and the 20 highest-yielding fields among those which were unsprayed.

**Figure 5.** Shapley phi values attributed to foliar fungicides for four cohorts of soybean fields. The cohorts are (i) the 100 highest-yielding fungicide-treated fields, (ii) the 100 lowest-yielding fungicide-treated fields, (iii) the 100 highest-yielding unsprayed fields, and (iv) the 100 lowest-yielding unsprayed fields. The insert table summarizes the minimum (Min), maximum (Max) and mean predicted yields (bu/acre) for each of the four cohorts. Point color represents whether fields were sprayed or unsprayed, whereas point shape represents whether fields were in the lowest-yielding or highest-yielding cohorts.
individual yields in $s_{c2}$ & $s_{c3}$ were 15 to 30 bu/ac above the global average, other location-driven factors such as early sowing (Figure 2) were the larger drivers of yield in these cases. However, there was only a negligible or small (< 1.5 bu/ac) penalty for not using foliar fungicides in high-yield situations ($s_{c3}$).

There was some overlap in the fields of $s_2$ and $s_3$ [where $s_3$ consisted of fields within the 90th percentile for yield among sprayed fields ($s_{c1}$); and the 90th percentile for yield among the unsprayed fields ($s_{c2}$)], at least where high-yielding fields were concerned. All fields in $s_3$ had predicted yields that were above the global average (Figure 6). Yield distributions of the two cohorts within $s_3$ were similar, with the cohorts having near-identical mean yields. Foliar fungicides contributed between 1.5 and 5.2 bu/ac to the yield increase above the global average, while the penalty (if there was one) for not using foliar fungicides was mainly confined to less than 0.75 bu/ac, indicating that among the fields of $s_{c2}$ spraying was unnecessary (otherwise the penalty would have been larger). Overlaying the estimated $\phi$ values for fungicide use with MG, sowing date and growing degree days showed that these high-yielding fields were mainly in MG II and III, that the fields tended to be planted early, and were restricted to GDD groups 03 and 04 (Figure 6), the latter factor being highly aligned with latitude. A formal comparison of the Shapley $\phi$ values across cohorts was not attempted because they potentially differed in their underlying variables despite similar yield distributions within the lowest- or highest-yield cohorts.

Intuitively, one may have expected the yield increase due to foliar fungicides to be about the same magnitude (about 1.5 bu/ac) as the yield penalty associated with not using fungicides. The larger yield gain versus the penalty may be due to synergistic interactions of foliar fungicides with other management factors. For example, foliar insecticides are likely to be applied along with foliar fungicides; conversely, fields that were not sprayed with foliar fungicides were unlikely to be sprayed with insecticides as well. Therefore, in subset 4 ($s_4$), we examined the Shapley $\phi$ values associated with fungicide use among all 210 fields in the data matrix which had been sprayed with foliar fungicides but not with foliar insecticides ($s_{c1}$), and compared them to the Shapley $\phi$ values for foliar fungicide use among another cohort of 210 fields ($s_{c2}$) which had been sprayed with both foliar fungicides and insecticides, where the fields of $s_{c2}$ were sampled to match the range of reported yields in $s_{c1}$. There was no discernable separation of the Shapley $\phi$ values between cohorts $s_{c1}$ and $s_{c2}$ (Figure 7), and the $\phi$ values were consistent with what had been observed with the other subsets of fields.
A partial economic analysis estimated the net realized profit associated with foliar fungicide use on the respective cohorts within subsets of fields. The profitability of foliar fungicides in the fields of $s_1c_1$ (20 highest-yielding sprayed fields within the 12 TEDs with the most fields in the dataset) is shown in Figure 8. Assuming a price of US$15.68/bu, fungicides were overwhelmingly profitable in all but four TEDs (403603, 602303, 403703, 303603) in which the average return (with respect to fungicide use) was less than US$3/ac. For these four TEDs, confidence intervals for the mean financial return per ha after accounting for fungicide costs indicated returns could be negative (loss), zero, or up to US$10.72/ac, depending on the individual field. Considering these were the highest-yielding fields within TEDs, there was the risk of losing money on fungicide sprays in these four TEDs. Obviously, environment mattered, and with the four TEDs listed above the most noticeable feature was their higher-latitude locations relative to fields in other TEDs. Among other things, higher latitude is associated with cooler weather and shorter accumulation of GDD. Underlying yield potential factors (early sowing, PAWR, GDD, AI) contributed to higher predicted yield. Higher-yielding environments were more likely to also realize a larger contribution of foliar fungicides to yield above the global average, thereby leading to the profitability of spraying. The financial return on spraying the fields in $s_2c_2$ (100 lowest-yielding fungicide-sprayed fields) was negative, except in a few individual cases. The mean net return due to foliar fungicides for $s_2c_2$ (100 highest-yielding fungicide-sprayed fields) was US$30.20/ac, whereas for $s_2c_1$ the return was -US$10.62/ac. Considering the two cohorts of $s_3$ (unsprayed and sprayed fields in the 90th percentile for yield), there was a small financial penalty to not using foliar fungicides in high-yield environments. Not spraying high-yield fields ($s_3c_2$) was associated with a mean loss of -US$4.12/ac. Yet, spraying high-yield fields ($s_3c_1$) was associated with a mean gain of US$26.55.

The soybean price required to at least break even on a (fixed) fungicide investment cost of US$25.05/ac was a nonlinear function of $\phi$. At a realized Shapley $\phi$ value of 1.5 bu/ac in response to foliar fungicides, soybean price would have to be at least US$16.85/bu to recover the costs of fungicides and their application, dropping to US$8.42/bu, US$5.62/bu, and US$4.21/bu for Shapley $\phi$ values of 3, 4.5 and 6 bu/ac, respectively. The percentage of U.S. soybean acreage treated with foliar fungicides rose from 1 to 11% between 2004 and 2015 (62), which is a yearly increase of 0.91%. Assuming the average gain of 3.3 bu/ac due to foliar fungicides among sprayed fields in the 90th percentile for yield ($s_1c_1$), we estimated a yield gain of 0.03 bu/ac/year attributed to the adoption of foliar fungicide. This translated to 6% of the estimated annual yield gain in U.S. soybean (0.5 bu/ac/year) attributable to foliar fungicide use in high-yield environments.
ACKNOWLEDGMENTS

We acknowledge the North-Central Soybean Research Program (NCSRP) for their support of this project. Partial support for the research done in this study was provided by the USDA National Institute of Food and Federal Appropriations under Project PEN04660 and Accession Number 1016474 (to PDE).

Adapted from: Shah, A.D., T. R. Butts, S. Mourtzinis, J. I. Rattalino Edreira, P. Grassini, S. P. Conley and P. D. Esker. 2021. An interpretable machine learning assessment of foliar fungicide contribution to soybean yield in the north-central United States. Scientific Reports 11:18769. https://doi.org/10.1038/s41598-021-98230-2.

REFERENCES (citation number from original manuscript)

5. Mourtzinis, S., Specht, J. E. & Conley, S. P. Defining optimal soybean sowing dates across the US. Sci. Rep. 9, 2800. https://doi.org/10.1038/s41598-019-38971-3 (2019).

7. Dorrance, A. E. et al. Effects of foliar fungicide and insecticide applications on Soybean in Ohio. Plant Health Prog. https://doi.org/10.1094/PHP-2010-0122-01-RS (2010).

6. Andrade, J. F. et al. Assessing the influence of row spacing on soybean yield using experimental and producer survey data. Field Crops Res. 230, 98–106. https://doi.org/10.1016/j.fcr.2018.10.014 (2019).

8. Mourtzinis, S. et al. Neonicotinoid seed treatments of soybean provide negligible benefits to US farmers. Sci. Rep. 9, 11207. https://doi.org/10.1038/s41598-019-47442-8 (2019).

13. Mourtzinis, S. et al. Sifting and winnowing: Analysis of farmer field data for soybean in the US North-Central region. Field Crops Res. 221, 130–141. https://doi.org/10.1016/j.fcr.2018.02.024 (2018).

14. Mourtzinis, S. et al. Assessing benefits of artificial drainage on soybean yield in the North Central US region. Agric. Water Manage. 243, 106425. https://doi.org/10.1016/j.agwat.2020.106425 (2021).

18. Orlowski, J. M. et al. High-input management systems effect on soybean seed yield, yield components, and economic break-even probabilities. Crop Sci. 56, 1988–2004. https://doi.org/10.2135/cropsci2015.10.0620 (2016).
19. Kandel, Y. R. et al. Analyses of yield and economic response from foliar fungicide and insecticide applications to soybean in the north central United States. Plant Health Prog. 17, 232–238. https://doi.org/10.1094/PHP-RS-16-0038 (2016).

20. Ng, S. J., Lindsey, L. E., Michel, A. P. & Dorrance, A. E. Effect of mid-season foliar fungicide and foliar insecticide applied alone and in-combination on soybean yield. Crop. Forage Turfgrass Manag. 4, 1–6. https://doi.org/10.2134/cftm2017.09.0067 (2018).

21. Bergman, K., Ciampitti, I., Sexton, P. & Kovács, P. Fungicide, insecticide, and foliar fertilizer effect on soybean yield, seed composition, and canopy retention. Agrosyst. Geosci. Environ. https://doi.org/10.1002/agg2.20116 (2020).

22. Bandara, A. Y. et al. Modeling the relationship between estimated fungicide use and disease-associated yield losses of soybean in the United States I: Foliar fungicides vs foliar diseases. PLoS ONE 15, e0234390. https://doi.org/10.1371/journal.pone.0234390 (2020).

23. Kyveryga, P. M., Blackmer, T. M. & Mueller, D. S. When do foliar pyraclostrobin fungicide applications produce profitable soybean yield responses?. Plant Health Prog. 14, 6. https://doi.org/10.1094/php-2013-0928-01-rs (2013).

24. Bandara, A. Y., Weerasooriya, D. K., Bradley, C. A., Allen, T. W. & Esker, P. D. Dissecting the economic impact of soybean diseases in the United States over two decades. PLoS ONE 15, e0231141. https://doi.org/10.1371/journal.pone.0231141 (2020).

25. Wrather, J. A. & Koenning, S. R. Estimates of disease effects on soybean yields in the United States 2003 to 2005. J. Nematol. 38, 173–180 (2006).

26. Wrather, J. A. & Koenning, S. R. Effects of diseases on soybean yields in the United States 1996 to 2007. Plant Health Prog. 10, 24 (2009).

27. Kandel, Y. R. et al. Meta-analysis of soybean yield response to foliar fungicides evaluated from 2005 to 2018 in the United States and Canada. Plant Dis. https://doi.org/10.1094/pdis-07-20-1578-re (2020).

30. Mourtzinis, S., Marburger, D. A., Gaska, J. M. & Conley, S. P. Characterizing soybean yield and quality response to multiple prophylactic inputs and synergies. Agron. J. 108, 1337–1345. https://doi.org/10.2134/agronj2016.01.0023 (2016).

32. Mourtzinis, S. et al. Corn, soybean, and wheat yield response to crop rotation, nitrogen rates, and foliar fungicide application. Crop Sci. 57, 983–992. https://doi.org/10.2135/cropsci2016.10.0876 (2017).

37. Humphries, G. et al. (eds) Machine Learning for Ecology and Sustainable Natural Resource Management (Springer, 2018).

40. Matcham, E. G. et al. Management strategies for early- and late-planted soybean in the north-central United States. Agron. J. 112, 2928–2943. https://doi.org/10.1002/agj2.20289 (2020).

42. Giesler, L. J. & Miller, J. J. Managing foliar diseases in soybean. (2017). https://extensionpublications.unl.edu/assets/html/g1862/build/g1862.htm.