2022

Field validation of a farmer supplied data approach to close soybean yield gaps in the US North Central region

José F. Andrade
University of Nebraska-Lincoln, jandrade@unl.edu

Spyridon Mourtzinis
Agstat Consulting

Juan I. Rattalino Edreira
University of Nebraska-Lincoln

Shawn P. Conley
University of Wisconsin-Madison

John Gaska
University of Wisconsin-Madison

See next page for additional authors

Authors
José F. Andrade, Spyridon Mourtzinis, Juan I. Rattalino Edreira, Shawn P. Conley, John Gaska, Herman J. Kandel, Laura E. Lindsey, Seth Naeve, Scott Nelson, Maninder P. Singhi, Laura J. Thompson, James E. Specht, and Patricio Grassini

Field validation of a farmer supplied data approach to close soybean yield gaps in the US North Central region

José F. Andrade[a,b], Spyridon Mourtzinis[c], Juan I. Rattalino Edreira[a], Shawn P. Conley[d], John Gaska[d], Herman J. Kandel[e], Laura E. Lindsey[f], Seth Naeve[g], Scott Nelson[h], Maninder P. Singh[i], Laura Thompson[j], James E. Specht[a], Patricio Grassini [a,] *

[a] Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583-0915, USA

[b] Catedra de Cerealicultura, Facultad de Agronomía, Universidad de Buenos Aires, IFEVA-CONICET, 1417, Buenos Aires, Argentina.

[c] Agstat Consulting, Athens, Greece

[d] Department of Agronomy, University of Wisconsin-Madison, Madison, WI, 53706, USA

[e] Department of Plant Sciences, North Dakota State University, Fargo, ND 58108, USA

[f] Department of Horticulture and Crop Science, The Ohio State University, Columbus, OH 43210, USA

[g] Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA

[h] Iowa Soybean Association, Research Center for Farming Innovation, Ankeny, IA 50023, USA

[i] Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA

[j] Nebraska Extension, University of Nebraska-Lincoln, Falls City, NE 68355, USA

*Corresponding author. E-mail: pgrassini2@unl.edu (P. Grassini)

Abstract

CONTEXT: Producer-reported data can be used to identify suites of management practices that lead to higher yield and profit. However, a rigorous validation of the approach in relation to its potential impact on farmer yield and profit is lacking.

OBJECTIVE: This study aimed to validate a producer-data approach on its capability to guide on-farm evaluation of management practices with greatest potential for increasing producer yield and profit. We show proof of concept using soybean in the North Central US region as a case study.

METHODS: We used a combination of regression tree analysis and a spatial framework to determine practices with highest influence on yield for specific climate domains across the region. These practices were used as a basis for designing an 'improved' management package for each domain. The impact associated with adoption of the 'improved' management package on producer yield, seed constituents, and profit was evaluated against a 'reference' treatment that follows farmer management *via* replicated on-farm trials across 100 sites over two crop seasons.

RESULTS AND CONCLUSIONS: Average yield was 278 kg ha^{-1} higher in the improved *versus* reference management, equivalent to a closure of the current exploitable yield gap by 40%. In turn, adoption of the improved management led to an average increase of \$76 ha^{-1} in net profit. Sensitivity analysis showed that adoption of the improved management packages should increase farmer profit across a wide range of grain price scenarios, with very small downside risk. Seed protein concentration was negatively associated with the positive yield advantage of the improved management, whereas seed oil concentration tended to increase.

SIGNIFICANCE: Analysis of producer data can accelerate discovery, evaluation, and adoption of suites of management practices that consistently lead to higher farmer yield and profit, which, in turn, would help speed up current rates of yield gain.

Keywords: soybean, yield, farmer data, profit, seed protein.

1. Introduction

Soybean production needs to increase by *ca.* 15% worldwide during the next decade to meet the growing demand for food, biodiesel, and livestock feed (OECD-FAO, 2019). United States (US) produces about one third of global soybean, with the North Central region accounting for 80% of US soybean production. On-farm soybean yield in the US during the last decade averaged 3.2 Mg ha^{-1} (2011-2020, USDA, https://www.nass.usda.gov). Soybean yield potential, as determined by the interactive effects of weather, soils, genetics, and water availability, has been estimated to average *ca.* 5 Mg ha^{-1} for the US North Central region, ranging from *ca.* 4 Mg ha^{-1} in short-season water-limited environments to 7 Mg ha^{-1} in irrigated favorable environments (www.yieldgap.org; Rattalino Edreira et al., 2017). The difference between yield potential and average farmer yield is known as the yield gap (van Ittersum et al., 2013). Full closure of the yield gap is difficult to achieve, as it would require a host of costly inputs, coupled with producer acquisition of a greater degree of sophistication needed for the implementation of key soil and crop management practices. Thus, striving to attain yields above *ca.* 80% of the yield potential may not be justifiable due to diminishing returns to additional inputs and labor and increasing negative environmental impact (Cassman, 1999; Cassman et al., 2003). In practice, 80% of yield potential is typically considered to be a realistic measure of the yield that could be attained by farmers with good access to inputs, markets, and technical information as in the case of crop producers in US. There, it has been estimated that the current exploitable yield gap for soybean in the US North Central region is *ca.* 0.7 Mg ha^{-1} (www.yieldgap.org; Rattalino Edreira et al., 2017). Understanding causes for the yield gap can help tune current management and, by doing so, increase regional production on existing cropland.

The most common approach for identifying yield-limiting and reducing factors in producer fields has been *via* replicated on-farm trials in which researchers selectively apply a set of inputs or management practices and compare them against a reference management across multiple locations and years (Villamil et al., 2012; Orlowski et al., 2016). However, this approach has several limitations. First, it is difficult to evaluate more than two factors at a time in replicated field experiments. Second, cost, labor, and time expended in the planning and oversight of multiple fields trials imposes limits on the number of treatments and environments that can be evaluated. Third, results derived from the comparison among different management practices will depend on the weather and soil background associated with each field. Finally, selection of management practices to be evaluated are, in many cases, determined based on perceptions from researchers from public and private sectors about yield constrains without field validation. Not surprisingly, many of these evaluations have led to the conclusion that little can be done to narrow the existing yield gap in a cost-effective way (Bluck et al., 2015; Marburger et al., 2016; Mourtzinis et al., 2016; Gregg et al., 2015; Seidel et al., 2015; Di Mauro et al., 2022).

An alternative method is to simply identify those management practices with largest impact on producer fields for a given environmental context *via* analysis of self-reported producer data (e.g., Grassini et al., 2011; 2015; Di Mauro et al., 2018; Ribas et al., 2021). This approach has the advantage of comparing the main effects of different management practices across the background of thousands of commercial fields and within the range of cost-effective management practices that are followed by producers (Rattalino Edreira et al., 2017; Mourtzinis et al., 2018). When complemented with crop modeling and remote sensing data, the underlying drivers of yield responses to different management practices can be identified and understood (Rattalino Edreira et al., 2020). The farmer-data approach can help orient evaluation of management practices that may increase producer yield and profit. While previous studies have suggested the potential of producer-data to accelerate discovery and adoption of improved management practices, an explicit field validation of this approach in relation to its impact on producer yield and profit *via* on-farm trials is pending.

This study aimed to validate a producer-data approach on its capability to guide on-farm evaluation of management practices with greatest potential for increasing producer yield and profit. The approach consisted of identification of suites of management practices *via* analysis of producer supplied data for specific environmental domains. These practices were subsequently evaluated for impact on yield, seed constituents, and profitability against current farmer management across 100 replicated on-farm trials conducted in the US North Central region over two crop seasons (2019 and 2020). Implications of the proposed approach to guide agricultural research and extension programs are discussed.

2. Methods

2.1. Identification of yield-enhancing management practices

We used the climatic zone (CZ) spatial framework developed for the Global Yield Gap Atlas as a basis for our analysis (www.yieldgap.org). Briefly, the CZ spatial framework delineates geographic domains based on (i) annual total growing degree-days, (ii) aridity index, and (iii) annual temperature seasonality

(Rattalino Edreira et al., 2018). Although the framework may not eliminate some residual variation in climate within each CZ, and does not consider variation in soil types, it helps in assembling cohorts of fields with similar biophysical backgrounds so that the main effects of management practices on yield can be more easily assessed (Rattalino Edreira et al., 2017; Mourtzinis et al., 2020). We focused on seven CZs within the US North Central, altogether accounting for 60% of total US soybean area **(Figure 1)**.

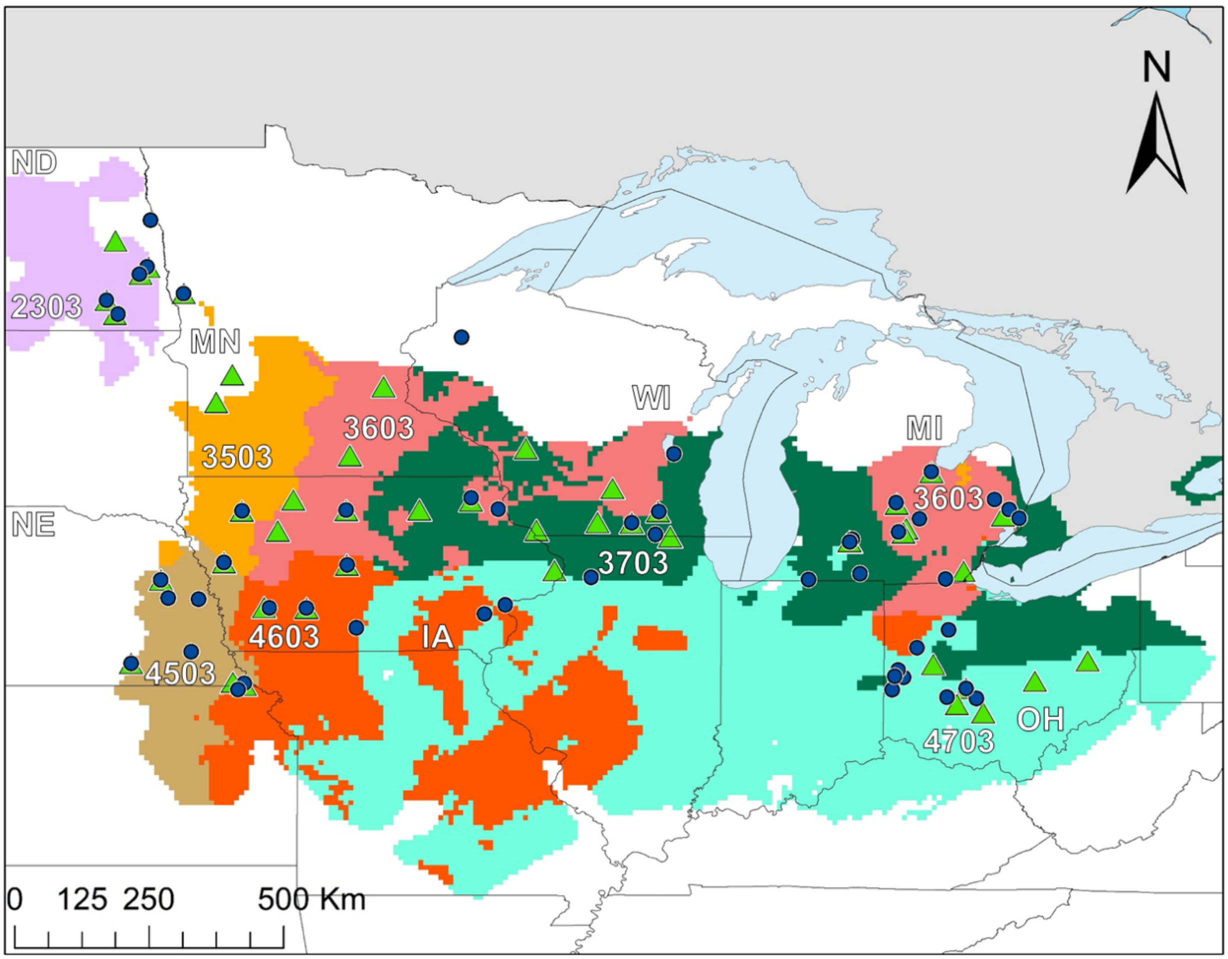

Figure 1. *Locations of the 100 on-farm trials conducted in 2019 (green triangles) and 2020 (blue circles) conducted in the North Central US region. The seven differently colored areas denote the seven climate zones (CZs) selected for our study. CZ values indicate a unique combination of annual total growing degree-days, aridity index, and annual temperature seasonality (www.yieldgap.org). Also shown are the state acronyms: North Dakota (ND); Minnesota (MN), Wisconsin (WI), Michigan (MI), Nebraska (NE); Iowa (IA); Ohio (OH).*

For each of the seven selected CZs, we have farmer data on soybean yield and management practices collected over four cropping seasons (2014-2017). Details on the collected data are provided elsewhere (Rattalino Edreira et al., 2017; Mourtzinis et al., 2018). Briefly, soybean producers were asked to provide data for several fields that portray their yield range. Requested data included field location, average field yield, and details on 16 crop management practices. Here we used conditional inference tree analysis to identify management practices associated with higher yields in each of the seven CZs (R development Core team, 2016). The algorithm selected the input management practices with strongest association with yield, measured by a p-value (alpha = 0.05 in this study). Then, a binary split was implemented in the selected input variable (node) and all steps are recursively repeated. As a result, a tree look-alike graph is displayed showing hierarchical associations between yield and crop management practices **(Figure 2)**. The terminal nodes indicate the average yield for each final subset of fields with similar background management practices. For all CZs, there were data from >200 field-year observations, which is considered sufficient to identify yield constrains following this type of analysis. Detailed description of the statistical analysis is provided elsewhere (Mourtzinis et al., 2018).

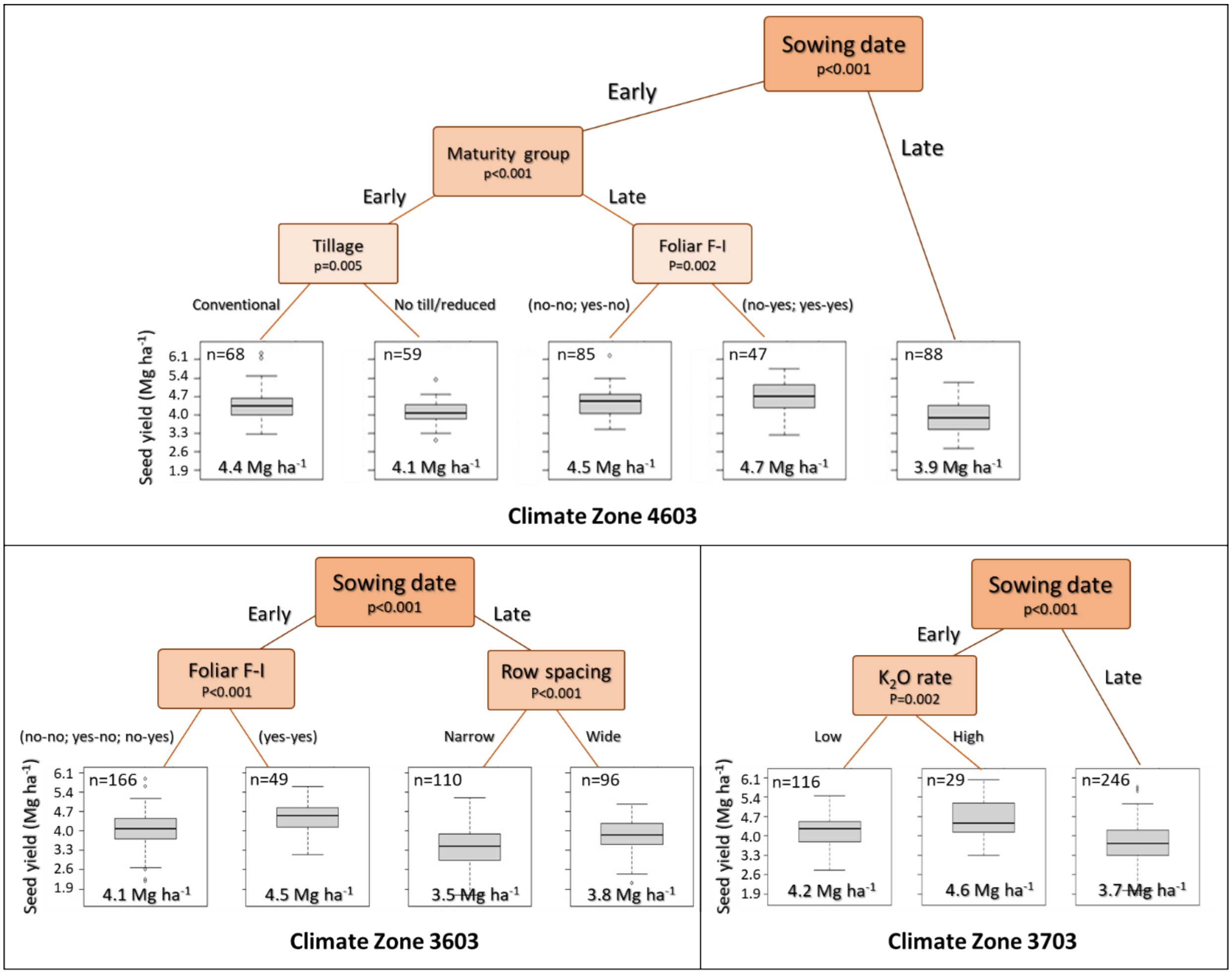

Figure 2. *Example of regression tree analysis for three climate zones (CZs) using famer reported data on commercial fields (2014-2017). Foliar F-I stands for foliar fungicide and/or insecticide (Mourtzinis et al., 2018). CZ values indicate a unique combination of annual total growing degree-days, aridity index, and annual temperature seasonality (www.yieldgap.org).*

2.2. On-farm validation

Based on the results of the foregoing regression-tree analysis, we selected a set of management practices to be evaluated *via* on-farm trials in each of seven CZs **(Table 1, Figure 1)**. When selecting practices to test, we did not consider those that would involve substantial a priori capital cost investments, such as installation of artificial drainage, or upgrades in field machinery. Instead, we focused on those practices that could be readily changed by producers, including sowing date, maturity group (MG), and application of in-season foliar fungicide and insecticide (FI). In addition, some management practices that, in our regression trees, were judged to be yield-neutral factors, were included in our evaluations, not only for confirmation purposes, but also to assess the degree to which the practice could be modified (without yield penalty) to reduce its implementation cost (for example, seeding rate). We did not attempt to quantify the individual impact of each individual practices on yield and profit, but rather performing a 'system comparison' in which the selected management practices were combined into an 'improved' treatment, which was subsequently evaluated against the typical farmer management, used as a 'reference' treatment. In all cases, the improved treatment involved an earlier sowing date in comparison to the reference treatment (average difference: -20 d). Experiments also included lower seeding rate in the improved treatment (from 65 to 53 seeds m^{-2}). Seeding rates were reduced proportionally less in CZ 2303 (from 75 to 67 seeds m^{-2}), as a higher plant population is recommended in this northern short-season environment to allow canopy closure at the beginning of the critical period for yield determination at R3 stage. In some cases, the improved treatment also involved application of an in-season foliar fungicide and insecticide (CZs 3503, 3603, 4503, and 4703), and use of a longer MG cultivar (CZ 4603 and CZ 2303).

To conduct the field trials, we collaborated with extension personnel and researchers located in seven states: Nebraska (NE), Iowa (IA), North Dakota (ND), Minnesota (MN), Wisconsin (WI), Michigan (MI), and Ohio (OH). Relative to the location of evaluation trials, we selected farmer fields that were representative of dominant tillage, water management (tiled or non-tiled), and crop sequence within each CZ. In each field, we evaluated the improved treatment against a reference treatment that reflected the typical producer management in that CZ **(Table 1)**. Each experimental unit consisted of a field strip (12-60 m wide × at least 91 m long), and there were four field strips per treatment, which were used as replicates. Field trials were conducted during two crop seasons (2019 and 2020), including a total of 100 replicated site-year trials. Measured seed yields at harvest were adjusted to 130 g kg^{-1} seed moisture content. Seed protein and oil concentration were determined after harvest for a total of 82 trials using near infrared reflectance spectroscopy (DA 7250 NIR analyzer, Perten Instruments, Inc., Springfield, IL, US).

Table 1. *Number of trials per climate zone (CZ), field background, and management practices associated with the reference and improved treatments.*

CZ number (number of trials)	Field background	Reference				Improved treatment
		Sowing date (DOY)	MG	Seeding rate (m^{-2})	Foliar F and/or I	
2303 (n = 12)	No tile drainage, conventional tillage, previous maize crop.	151	0.5	75	yes	Earlier sowing, longer MG, lower seeding rate
3503 (n = 6)	Tile drainage, conventional tillage, previous maize crop	160	1.8	65	no	Earlier sowing, FI, lower seeding rate
3603 (n = 27)	Tile drainage, conventional tillage, previous maize crop.	143	2.5	60	no	Earlier sowing, FI, lower seeding rate
3703 (n = 20)	No tile drainage, no-till or reduced-till, previous maize crop	143	2.3	57	no	Earlier sowing, lower seeding rate
4503 (n = 11)	No tile drainage, no-till or reduced-till, previous maize crop	139	3.5	65	no	Earlier sowing, FI, lower seeding rate
4603 (n = 9)	Tile drainage, no-till or reduced-till, previous maize crop.	141	2.3	57	yes	Earlier sowing, longer MG, lower seeding rate
4703 (n = 15)	Tile drainage, no-till or reduced-till, previous maize crop	149	3.4	65	no	Earlier sowing, FI, lower seeding rate

FI: in-season foliar fungicide and insecticide application around beginning of pod setting, which corresponds to R3 stage in the Fehr and Caviness (1977) stage system. Sowing date (as day of year), maturity group (MG), and seeding rate are average values across sites within CZs. CZ values indicate a unique combination of annual total growing degree-days, aridity index, and annual temperature seasonality (www.yieldgap.org).

2.3. Data analysis

We evaluated the effect of management, CZ, year, and their interactions on yield and seed composition using a combined analysis of variance (ANOVA) (Moore and Dixon, 2015). CZ × management × year combinations were not too different in relation with their variances (Fmax<7); hence, our combined ANOVA can be considered robust (Milliken and Johnson, 2009). The CZs, management treatments, and year were treated as fixed effects and mean differences between management practices were calculated. Although different sites cannot be directly compared, the CZ × year × management interaction can be used to evaluate consistency of the response to the improved management. For each site × year, two-tailed t-tests were used to evaluate differences in yield and seed constituents between the reference and improved management treatments. We used linear regression analysis to explore relationships between yield and seed constituents with other factors to understand variation in the response to the improved management treatment across site-years. Despite our efforts to conduct on-farm trials within the seven selected CZs, a few trials (*n*=4) fell outside them; those trials were assumed to be inside the nearest selected CZ for the analysis.

We performed partial economic analysis to calculate the additional profit derived from the improved management treatment in comparison to the reference. To calculate the net change in production cost, we only considered those inputs that were different between management treatments (*i.e.*, seeding rate and foliar application of fungicide and/or insecticide). Input prices varied little across years (Nebraska Crop Budgets; https://cropwatch.unl.edu/budgets); hence, we used a 7-year input cost data (2015-2021) to generate treatment prices: foliar insecticide (\$10 ha^{-1}), foliar fungicide (\$52 ha^{-1}), and foliar application cost (\$17 ha^{-1}). Different prices were used for fungicide and/or insecticide treated (\$154 ha^{-1}) and non-treated seed (\$133 ha^{-1}). To calculate the additional net profit, we multiplied the yield response to the improved treatment by the average input prices for the period 2015-2021 for soybean seed. Finally, to account for variation in net profit because of variation in soybean grain price, we ran a sensitivity analysis using the range of soybean prices during the past decade (2012-2021), which ranged from \$295 to \$595 Mg^{-1} (USDA, https://www.nass.usda.gov/Charts_and_Maps/Agricultural_Prices/pricesb.php).

Finally, we calculated the portion of the yield gap that was narrowed due to adoption of the improvement management. To do so, we used the long-term water-limited yield potential (Yw) and average farmer yield available in the Global Yield Gap Atlas (www.yieldgap.org). For each CZ, we computed the average Yw and calculated the attainable yield as 80% of the Yw. The exploitable yield gap for each CZ was calculated as the difference between attainable yield and average farmer yield. Comparison of the yield response to the improved management treatment and the exploitable yield gap let us quantify the portion of the yield gap that could be reduced *via* adoption of our improved management packages. Extra soybean production derived from adoption of the improved management was calculated by multiplying the yield response by the total soybean harvested area located within the seven CZs.

3. Results

3.1. Yield response to the improved management packages

According to the analysis of variance, CZ, treatment, year, and CZ × year were significant sources of variation in soybean yield **(Table 2)**. The CZ and CZ × year terms explained a relatively large portion of total variation in seed yield (21%), indicating that delineation of climatic domains was useful to reduce the error associated with environmental variability. In addition, the treatment × year, treatment × CZ, and treatment × year × CZ interactions were not significant for seed yield, suggesting that the response to the improved technological package was consistent across years and CZs. Overall, the improved treatment averaged 4,182 kg ha^{-1}, representing a 7% increase over the reference treatment (3,904 kg ha^{-1}) **(Figure 3)**. Although treatment × year interaction was not significant, the yield benefit derived from the improved treatment was comparably smaller in 2020 than in 2019 (+214 *versus* +369 kg ha^{-1}). Yield response was positive and significantly different from zero in 50% and 37% of the trials in 2019 and 2020, respectively (t-test, $p<0.01$; **Figure 3-insets**). Those few trials in which the improved treatment delivered a negative yield response were associated with specific site-year factors. For example, in a few cases, heavy rainfall events after sowing of the improved treatment led to soil crusting and poor crop establishment. In other cases, water excess early in the seasons confounded the treatment comparison.

Table 2. *Analysis of variance for seed yield, and seed protein and oil concentration based on data collected from 100 replicated field trials conducted in seven climate zones (CZ) spanning the US North Central region in 2019 and 2020.*

Source of variation	d.f.	Yield		Protein		Oil	
		%SS	F	%SS	F	%SS	F
CZ	6	13	20^{**}	1.2	1.4	8	12^{**}
Treatment (T)	1	2.4	23^{**}	0.2	1.4	0.7	6^{*}
Year (Y)	1	0.6	6^{*}	<0.1	0.3	11	104^{**}
CZ x Y	6	8	15^{**}	0.5	0.6	1.2	2
CZ x T	6	0.5	0.5	0.6	0.8	0.8	1
T x Y	1	0.1	2	<0.1	0.3	<0.1	0.1
CZ x T x Y	6	0.3	0.5	0.5	0.6	0.4	0.7
Residual	715	74		96		75	

Asterisks indicate statistical significance at *p<0.05 and **p<0.001.

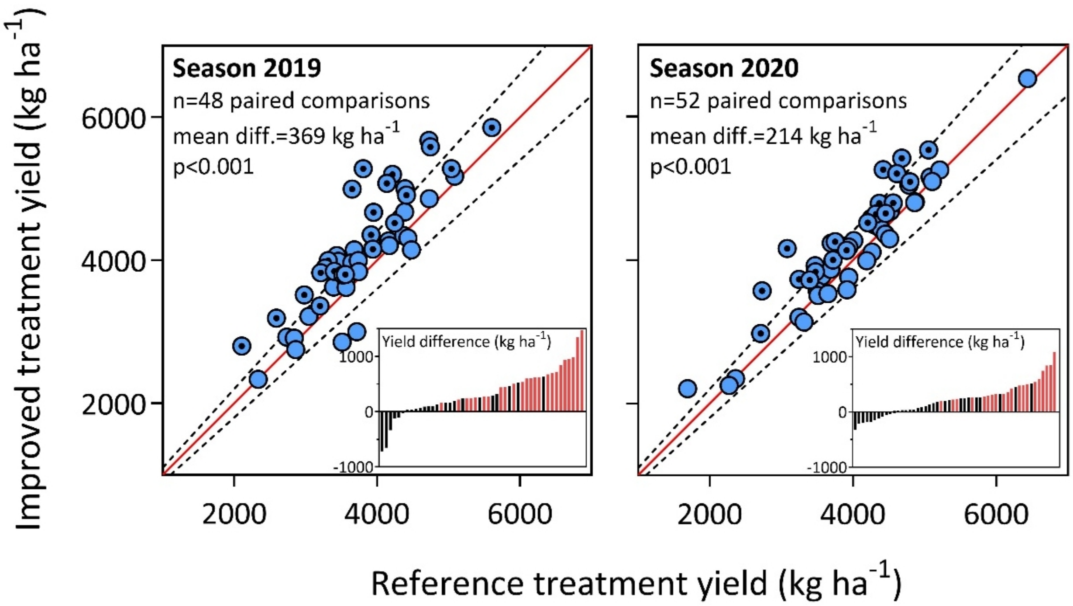

***Figure 3**. Yield comparison between improved and reference treatment across 48 fields in 2019 and 52 fields in 2020, distributed in seven climate zones of the North Central US region. Symbols with a dot inside refer to individual experiments with significant yield differences between treatments. The red line is the 1:1 line of agreement. The dashed lines show the ±10% deviation from the 1:1 line. Insets show the yield response difference between improved and reference treatments in each trial, ranked left to right by the sign (negative to positive) and magnitude of the response. Individual experiments with significant yield differences between the improved and reference treatments are shown with red bars.*

A more detailed analysis across all CZ-year combinations revealed that the average yield response of the improved treatment was positive and significant in 7 out of 14 CZ × year combinations (t-test, p<0.05)

(Figure 4). Across CZs-year combinations, the average yield response ranged from 96 to 547 kg ha^{-1} and in most CZs, the magnitude of yield response was larger in 2019 than in 2020. The latter effect could be partially related to differences in the sowing window between crop seasons: on average, sowing date was a week earlier in 2020 than in 2019. In contrast, the time interval between sowing of the improved and reference treatments did not change between years. Hence, we speculate that the positive effect of early sowing was greater when the sowing window was later, as it was the case in 2019.

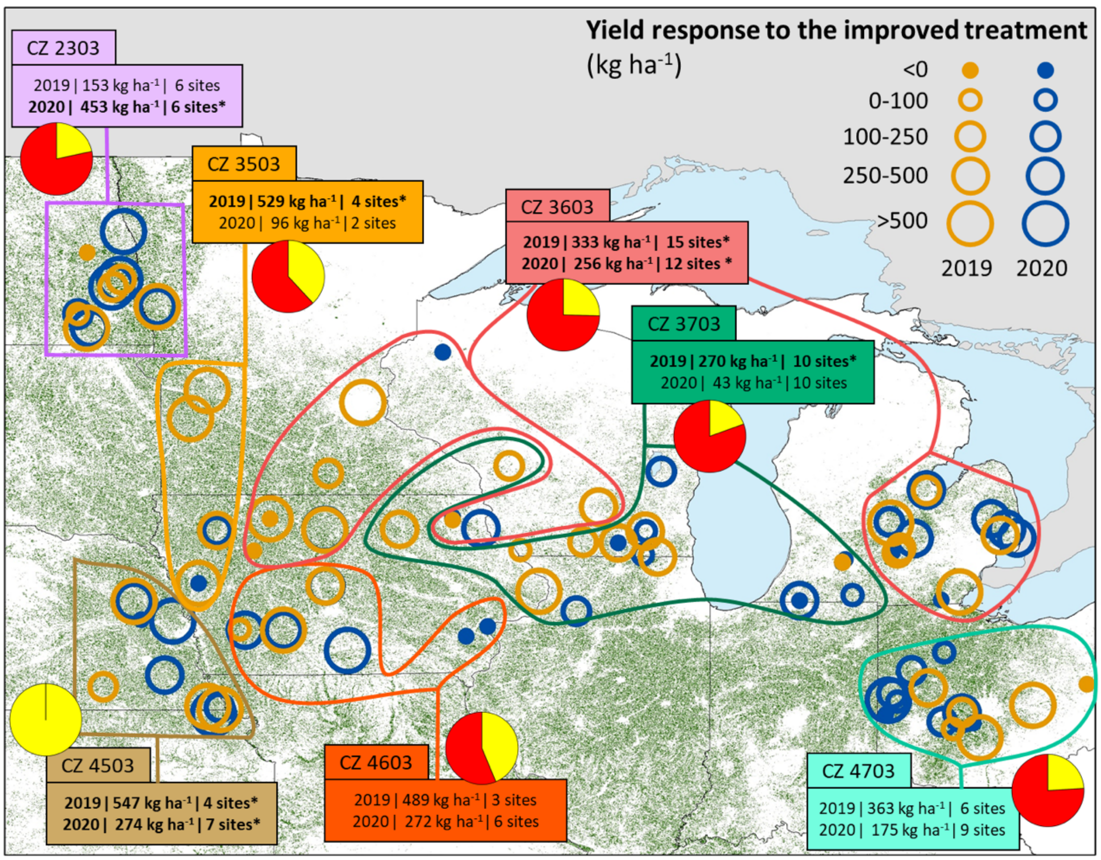

Figure 4. *Yield difference between the improved and reference management in replicated field trials in 2019 (orange) and 2020 (blue). Asterisks denote statistical significance ($p<0.05$) of the differences. Soybean hectarage density per unit area in 2020 is shown in green (USDA, 2021). Pie chart fractions depict the yield gap closure due to adoption of the improved management treatment (yellow), and remaining exploitable yield gap (red) in each CZ.*

It is of interest to analyze the degree to which the current exploitable yield gap was narrowed due to adoption of improved agronomic practices. We estimated that, on average, the exploitable yield gap was reduced by 40% with adoption of the improved management, although the reduction ranged from 20 to 100% amongst the seven CZs **(Figure 4)**. In general, yield gap closure due to adoption of the improved treatment was comparably larger in the western and central CZs compared with those located in the eastern portion of the region.

3.2. Changes in seed constituents due to adoption of the improved management

The ANOVA revealed a significant effect of treatment on seed oil concentration, with average concentration increasing from 22.3% to 22.7% in the reference and improved management treatments, respectively **(Table 2, Figure 5)**. An individual analysis of each of 82 trials in which seed oil was measured revealed that only 18 (22%) exhibited a statistically significant increase ($p=0.05$) in oil concentration in the improved management treatment, and one trial with a significant decrease. The overall effect of treatment on seed protein concentration was not statistically significant. Indeed, the analysis of individual experiments showed that seed protein concentration did not differ between treatments in 57 (70%) of the 82 trials. In fact, we found a slightly lower protein concentration in the improved *versus* reference treatment when the analysis was restricted to just the 2019 trials ($p<0.01$). Changes in seed protein concentration due to adoption of the improved management treatment were negatively correlated with changes in seed yield, with protein concentration decreasing 0.76% Mg^{-1} of yield increase **(Figure 6)**. This relationship was not apparent for seed oil concentration. Higher seed yield, higher oil concentration, and similar or slightly lower seed protein concentration in the improved management treatment led to 8% and 6% higher oil and protein yield ha^{-1}, respectively, in comparison to the reference management.

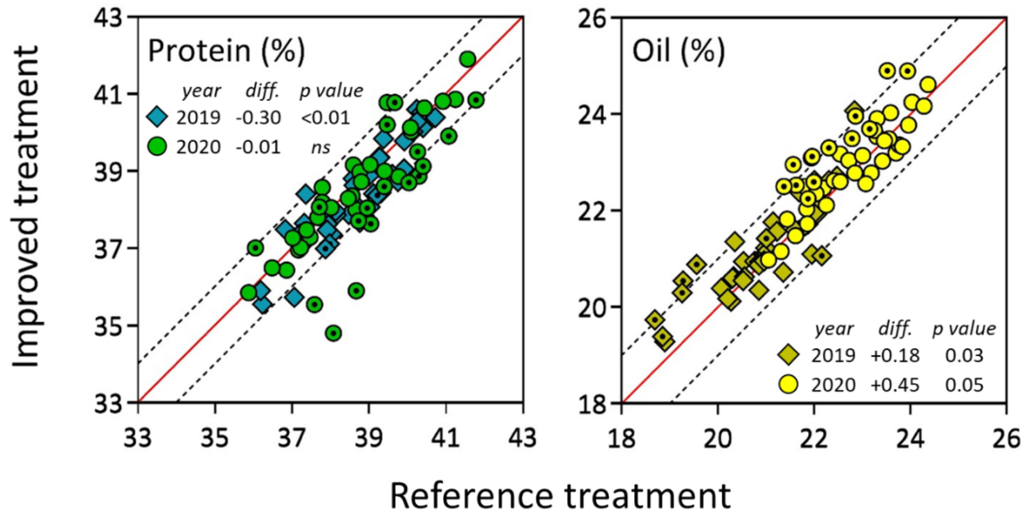

Figure 5. *Comparison of seed protein and oil concentration (%) between the reference and improve treatments. Symbols shown with a dot inside refer to individual trials with significant differences between treatments. Solid line indicates 1:1 between both treatments. Dashed lines indicate ±1% change in protein or oil concentration. Statistical significance for differences in seed constituents between treatments in each year are shown.*

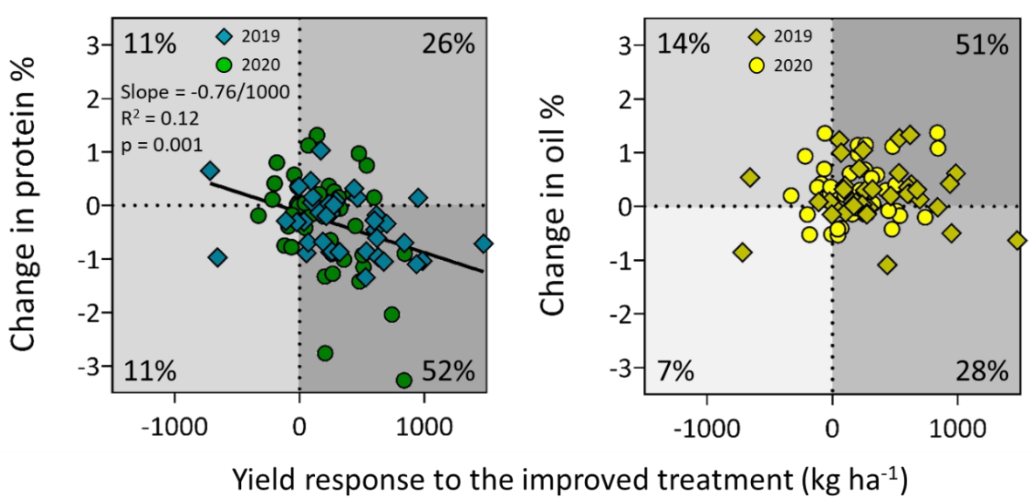

Figure 6. *Relationship between changes in seed protein and oil concentration between the improved and reference treatment and associated changes in seed yield. Slope, coefficient of determination, and associated p-value for the fitted linear regression are shown in the left panel. Also provided is the percentage of data points falling in each quadrant (corresponding to different combinations of negative and positive changes in seed constituents and seed yield).*

3.3. Change in net profit due to adoption of improved agronomic management

Yield increase, together with lower seed costs in some CZs, resulted in higher net profit in the improved *versus* reference management treatment **(Figure 7)**. Considering the median soybean price of \$356 Mg^{-1} during the past decade (2012-2021), and the experimental results from both seasons, the improved treatment resulted in an average +\$76 ha^{-1} extra profit (median = \$60 ha^{-1}). The positive economic impact derived from adoption of the improved management was high and fairly consistent across farms and years: the extra profit derived from the improved management was higher than \$25 ha^{-1} in 70% of the cases. We explored how the extra profit may change due to variation in soybean prices, considering the range from 5th to 95th percentiles during the 2012-2021 period. Our analysis shows that, across the range of grain price, the median additional profit would range from \$52 to \$97 ha^{-1} **(Figure 7)**.

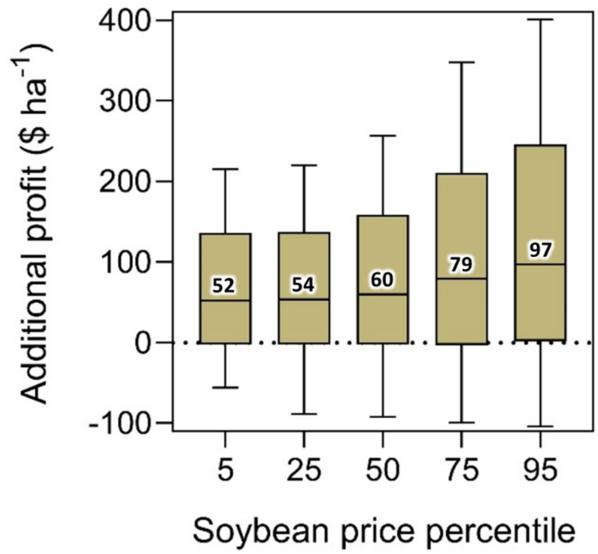

Figure 7*. Partial profit change (improved minus reference treatment) as a function of soybean price. The dashed line shows the zero-extra profit threshold. Additional profit was calculated across 100 trials in 2019 and 2020 using different prices per Mg -percentiles 5 (\$306), 25 (\$322), 50 (\$356), 75 (\$467), and 95 (\$536)- across the 2012-2021 period. Boxes indicate percentiles 25 and 75 of additional profit derived from the improved treatment; whiskers indicate percentiles 10 and 90. Median additional profit is indicated within boxes.*

4. Discussion

The combined analysis of producer-reported data and a robust spatial framework allowed us to design more productive *and* profitable soybean systems tailored for specific climatic domains. The use of a spatial framework delineating environments with different climate was key for grouping surveyed farmer fields based on their biophysical background while analysis of producer data helped us identify suites of management practices leading to highest yields in each environment. Likewise, the spatial framework allowed us to guide selection of fields for the on-farm validation, evaluate the experimental results, and scale them to measure the ex-ante impact associated with the improved management practices. The approach proposed here has the potential to accelerate the discovery, evaluation, and adoption of improved management practices leading to higher yield and profit, avoiding the problems associated with blanket recommendations or fuzzy expectations about commercial products (Mourtzinis et al., 2016; 2020; Di Mauro et al., 2022).

When evaluated in replicated field trials across the US North Central region, the improved management treatment, which was designed based on the analysis of the producer-reported data, resulted in higher seed yield (+278 kg ha^{-1}) and net profit +$76 ha^{-1} compared with the reference management (**Figure 3**). Our analysis shows that the positive impact on profit derived from the improved management was consistent across a wide range of grain prices, with little downside risk **(Figure 7)**. Scaled out over the US North Central region, where *ca.* 20.6 M ha are annually sown with soybean, the yield increase amounts to an additional annual soybean output of 5.8 MMT, which translates into ca. 2 billion US$ y^{-1} considering average grain prices. Our analysis also shows that seed constituents changed little with the improved management, generally increasing oil concentration, while maintaining or slightly decreasing seed protein concentration. The latter was expected, as seed protein concentration tends to decline with early sowing date (Mourtzinis et al., 2017, Helms et al., 1990), which was a common practice for improved treatments across CZs. Though the percentage changes in seed composition were small or nil, total protein and oil production (per unit of area) increased with improved management, which is consistent with previous studies (Assefa et al., 2019).

The yield increase that resulted from adopting our improved management was equivalent to 40% reduction of the existing exploitable yield gap in soybean **(Figure 4)**. At question is which other management practices should be adjusted to further narrow the exploitable yield gap. As explained previously, our improved management treatment did not include management practices that would have required changes in field equipment and/or large investments, as it would have been the case with row spacing, artificial drainage, and tillage (Mourtzinis et al., 2021, Andrade et al., 2019). Similarly, other management practices required to reduce the current yield gap have important trade-offs with costs and environmental impact. For example, Cafaro La Menza et al. (2017) showed that nitrogen supply can limit yield in highly productive environments; however, the amount of N fertilizer needed to overcome this limitation is, at present, not cost effect nor environmentally desirable. Still, we believe that continued further analysis of farmer data, together with careful evaluation in well-designed on-farm field trials, can help identify additional opportunities for enhancing yield and/or profit as well as helping uncover other potential trade-offs associated with changes in management practices.

Acknowledgments

We thank the North Central Soybean Research Program (NCSRP), Nebraska Soybean Board (NSB), and Wisconsin Soybean Marketing Board (WSMB) for supporting the project and the farmers and extension educators participating in the project for their excellent fieldwork.

References

Andrade, J.F., Rattalino Edreira, J.I., Mourtzinis, S., Conley, S.P., Ciampitti, I.A., Dunphy, J.E., Gaska, J.M., Glewen, K., Holshouser, D.L., Kandel, H.J., Kyveryga, P., Lee, C.D., Licht, M.A., Lindsey, L.E., McClure, M.A., Naeve, S., Nafziger, E.D., Orlowski, J.M., Ross, J., Staton, M.J., Thompson, L., Specht, J.E., Grassini, P., 2019. Assessing the influence of row spacing on soybean yield using experimental and producer survey data. Field Crops Res. 230, 98–106.

Assefa, Y., Purcell, L.C., Salmeron, M., Naeve, S., Casteel, S.N., Kovács, P., Archontoulis, S., Licht, M., Below, F., Kandel, H., Lindsey, L.E., Gaska, J., Conley, S., Shapiro, C., Orlowski, J.M., Golden, B.R., Kaur, G., Singh, M., Thelen, K., Laurenz, R., Davidson, D., Ciampitti, I.A., 2019. Assessing variation in US soybean seed composition (protein and oil). Front. Plant Sci. 10 (298).

Bluck, G.M., Lindsey, L.E., Dorrance, A.E., Metzger, J.D., 2015. Soybean yield response to Rhizobia inoculant, gypsum, manganese fertilizer, insecticide, and fungicide. Agron. J. 107, 1757–1765.

Cafaro La Menza, N., Monzon, J.P., Specht, J.E., Grassini, P., 2017. Is soybean yield limited by nitrogen supply? Field Crops Res. 213, 204–212.

Cassman, K.G., 1999. Ecological intensification of cereal production systems: yield potential, soil quality, and precision agriculture. Proc. Natl. Acad. Sci. U. S. A.96, 5952–5959.

Cassman, K.G., Dobermann, A., Walters, D.T., Yang, H., 2003. Meeting cereal demand while protecting natural resources and improving environmental quality. Annu. Rev. Environ. Resour. 28, 315–358.

Di Mauro, G., Cipriotti, P.A., Gallo, S., Rotundo, J.L., 2018. Environmental and management variables explain soybean yield gap variability in Central Argentina. Eur. J. Agron. 99, 186–194

Di Mauro, G., Salvagiotti, F., Gambin, B.L., Condori, A., Gallo, S., Pozzi, R., Boxler, M., Rotundo, J.L., 2022. Assessing the impact of high-input management for reducing soybean yield gaps on high-productivity farms. Field Crops Res., 278, 108434.

Fehr, W.R., Caviness, C.E., 1977. Stages of soybean development. Cooperative Extension Service, Agriculture and Home Economics Experiment Station. Iowa State University, Ames, Iowa.

Grassini, P., Thorburn, J., Burr, C., Cassman, K.G., 2011. High-yield irrigated maize in the Western U.S. Corn-Belt: I. On-farm yield, yield-potential, and impact of management practices. Field Crops Res. 120, 142–150.

Grassini, P., Van Bussel, L.G.J., Van Wart, J., Wolf, J., Claessens, L., Yang, H., Boogaard, H., de Groot, H., van Ittersum, M.K., Cassman, K.G., 2015. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. Field Crops Res. 177, 49–63.

Gregg, G.L., Orlowski, J. M., Lee, C.D., 2015. Input-based stress management fails to increase soybean yield in Kentucky. Crop, Forage & Turfgrass Management, 1, 1-7.

Helms, T.C., Hurburgh Jr, C.R., Lussenden, R.L., Whited, D.A., 1990. Economic analysis of increased protein and decreased yield due to delayed planting of soybean. J. Produc. Agric. 3, 367-371.

Marburger, D.A., Haverkamp, B.J., Laurenz, R.G., Orlowski, J.M., Wilson, E.W., Casteel, S.N., Lee, C.D., Naeve, S.L., Nafziger, E.D., Roozeboom, K.L., Ross, W.J., Thelen, K.D., Conley, S.P., 2016. Characterizing genotype × management interactions on soybean seed yield. Crop Sci. 56, 786–796.

Milliken, G.A., Johnson, D.E., 2009. second edition. Analysis of Messy Data Volume 1: Designed Experiments, vol. 1 CRC Press.

Moore, K.J., Dixon, P.M., 2015. Analysis of combined experiments revisited. Agron. J. 107, 763-771.

Mourtzinis, S., A. Gaspar, S. Naeve, Conley, S.P., 2017. Planting date, maturity, and temperature effects on soybean seed yield and quality. Agron. J. 109,1-10.

Mourtzinis, S., Andrade, J.F., Grassini, P., Edreira, J.I.R., Kandel, H., Naeve, S., Nelson, K.A., Helmers, M., Conley, S.P., 2021. Assessing benefits of artificial drainage on soybean yield in the North Central US region. Agric. Water Manag. 243, 1–8

Mourtzinis, S., Grassini, P., Rattalino Edreira, J.I., Andrade, J.F., Kyveryga, P., Conley, S.P., 2020. Assessing approaches for stratifying producer fields based on biophysical attributes for regional yield-gap analysis. Field Crops Res. Field Crops Res. 245, 107825.

Mourtzinis, S., Marburger, D.A., Gaska, J.M., Conley, S.P., 2016. Characterizing soybean yield and quality response to multiple prophylactic inputs and synergies. Agron. J. 108, 1–9.

Mourtzinis, S., Rattalino Edreira, J.I., Grassini, P., Roth, A.C., Casteel, S.N., Ciampitti, I.A., Licht, M., Kandel, H., Kyveryga, P.M., Lindsey, L.E., Mueller, D.S., Naeve, S.L., Nafziger, E., Specht, J.E., Stanley, J., Staton, M.J., Conley, S.P., 2018. Sifting and winnowing: Analysis of farmer field data for soybean in the US North-Central region. Field Crops Res. 221, 130–141.

OECD-FAO, 2019. Agricultural outlook 2019-2028. Food & Agriculture Organization.

Orlowski, J.M., Haverkamp, B.J., Laurenz, R.G., Marburger, D.A., Wilson, E.W., Casteel, S.N., Conley, S.P., Naeve, S.L., Nafziger, E.D., Roozeboom, K.L., Ross, W.J., Thelen, K.D., Lee, C.D., 2016. High-input soybean management systems affect soybean yield, yield components, and economic break-even probabilities. Crop Sci. 56, 1988–2004.

R Core Team, 2016. R: A Language and Environment for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Rattalino Edreira, J.I., Cassman, K.G., Hochman, Z., van Ittersum, M.K., van Bussel, L.G.J., Claessens, L., Grassini, P., 2018. Beyond the plot: Technology extrapolation domains for scaling out agronomic science. Environ. Res. Lett. 13, 054027.

Rattalino Edreira, J.I., Mourtzinis, S., Azzari, G., Andrade, J.F., Conley, S.P., Specht, J.E., Grassini, P., 2020. Combining field-level data and remote sensing to understand impact of management practices on producer yields. Field Crops Res., 257, 107932.

Rattalino Edreira, J.I., Mourtzinis, S., Conley, S.P., Roth, A.C., Ciampitti, I.A., Licht, M., Kandel, H., Kyveryga, P.M., Lindsey, L.E., Mueller, D.S., Naeve, S.L., Nafziger, E., Specht, J.E., Stanley, J., Staton, M.J., Grassini, P., 2017. Assessing causes of yield gaps in agricultural areas with diversity in climate and soils. Agric. For. Meteorol. 247, 170–180.

Ribas, G.G., Zanon, A.J., Streck, N.A., Pilecco, I.B., de Souza, P.M., Heinemann, A.B., Grassini, P., 2021. Assessing yield and economic impact of introducing soybean to the lowland rice system in southern Brazil. Agric. Syst. 188, 103036

Seidel, E.P., Egewarth, W.A., Piano, J.T., Egewarth, J., 2015. Effect of foliar application rates of Calcium and Boron on yield and yield attributes of soybean (Glycine max). Afric. J. Agric. Res. 10, 170-173.

USDA National Agricultural Statistics Service Information. Available at: https://www.nass.usda.gov.

van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap analysis with local to global relevance—a review. Field Crops Res. 143, 4–17.

Villamil, M.B., Davis, V.M., Nafziger, E.D., 2012. Estimating factor contributions to soybean yield from farm field data. Agron. J. 104, 881–887.