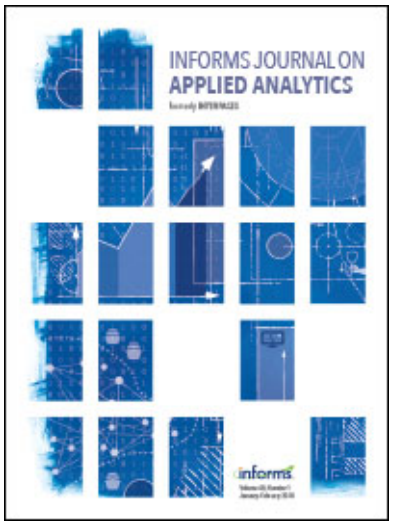


INFORMS Journal on Applied Analytics

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Genetic Gain Performance Metric Accelerates Agricultural Productivity

Joseph Byrum, Bill Beavis, Craig Davis, Greg Doonan, Tracy Doubler, Von Kaster, Ron Mowers, Sam Parry

To cite this article:
Joseph Byrum, Bill Beavis, Craig Davis, Greg Doonan, Tracy Doubler, Von Kaster, Ron Mowers, Sam Parry (2017) Genetic Gain Performance Metric Accelerates Agricultural Productivity. INFORMS Journal on Applied Analytics 47(5):442-453. https://doi.org/10.1287/inte.2017.0909

http://pubsonline.informs.org/journal/inte/

INTERFACES
Vol. 47, No. 5, September–October 2017, pp. 442–453
ISSN 0092-2102 (print), ISSN 1526-551X (online)

Genetic Gain Performance Metric Accelerates Agricultural Productivity

Joseph Byrum,[a] Bill Beavis,[b] Craig Davis,[a] Greg Doonan,[a] Tracy Doubler,[a] Von Kaster,[a] Ron Mowers,[a] Sam Parry[c]

[a] Syngenta Seeds, Inc., Slater, Iowa 50244; [b] Iowa State University, Ames, Iowa 50011; [c] Arizona State University, Phoenix, Arizona 85004

Contact: jrbyru@hotmail.com (JB); wdbeavis@iastate.edu (BB); craig.davis@syngenta.com (CD); greg.doonan@syngenta.com (GD); tracy.double@syngenta.com (TD); von.kaster@syngenta.com (VK); ronpmowers@gmail.com (RM); samuel.perry@asu.edu (SP)

https://doi.org/10.1287/inte.2017.0909

Abstract. The agricultural seed industry invests billions of dollars each year to improve our understanding of how best to unlock a seed's full potential. This investment brings a significant benefit to agricultural customers—the farmers who grow commodity crops, such as corn, soybeans, and wheat. Commodity farmers expect new crop varieties to be adapted to local conditions and have greater genetic potential for yield. We refer to the amount of increase in the genetic potential for yield as "genetic gain." The agricultural seed industry needs a universal, unbiased metric for genetic gain performance (GGP). Therefore, in 2010 we developed and implemented an algorithm that calculates an unbiased GGP metric that eliminates environmental factors (e.g., solar radiation, rainfall, and temperature) and is applicable at each stage of the product development pipeline. We subsequently used this metric during the variety development stage of our breeding projects to measure the impact of operational changes. We used weighted averages of GGP to retrospectively evaluate changes in genetic gain across 10 years of our breeding pipeline to quantify the benefit. We estimate that genetic gains are now 40 percent greater than the gains seen before implementation of the GGP in 2010. Our analyses show that the GGP metric has saved Syngenta approximately $250 million in varietal development costs, which would otherwise have been required to improve genetic gain by 40 percent. Syngenta scientists now use GGP to evaluate the genetic gain of all breeding projects. It serves as a valuable early-warning system. At the end of each growing season, we collect yield data and update the GGP database. This allows our scientists to perform an annual evaluation of genetic advances in each market segment. These assessments identify potential performance gaps likely to surface in the next growing season so that they can be avoided. Our successful development and deployment of a genetic gain metric is an important advance for both Syngenta and the entire agricultural industry.

Keywords: operational decisions • genetic gain improvement • genetic gain performance • algorithm • agriculture • varietal development

Introduction

~~The seed and agrochemical industry spends billions of dollars each year to develop new crop varieties with superior yield. Syngenta, a global seed and agrochemical company with over $13.5 billion in sales, spends millions of dollars every year in development.~~

At Syngenta, we work to develop new, superior crop varieties through a process of selective, conventional breeding and subsequent yield testing. Yield is the most important metric for our customers (farmers) because crop output directly impacts their profitability. Our customers expect new varieties to be adapted to local conditions and have a greater genetic potential for yield than previous varieties. We call the amount of increase in the genetic potential for yield in plant varieties "genetic gain," and we measure it in bushels per acre (bu/ac). Genetic gain is a key performance indicator for the seed industry.

In this paper, we describe how researchers at Syngenta developed and applied a novel algorithm, the first of its kind in applied plant breeding. This algorithm, implemented in 2010, offers a new, universal metric for genetic gain performance (GGP) of soybeans, separating nongenetic sources of variability from yield measurements. Specifically, GGP gives our variety development teams a way to more accurately measure the genetic performance during variety development by using applied mathematics and analytics to

identify and remove the bias of nongenetic factors. We describe the algorithm and present the results of six years of employing the model as a primary contributor to our operational decisions, such as experimental designs for comparison tests in future stages, which we must make throughout the year.

First, we outline the three phases of commercial soybean seed development. This process covers a six-to-seven-year period from variety design (generate genetic variance) through variety advancement (set favorable traits) to variety evaluation (selection of the most promising varieties). Next, we discuss the breeders' dilemma, isolating genetic gain (controllable by the breeder team) from environmental effects (uncontrollable by the breeding team). The operational challenge is to separate nongenetic influences of each year's growing season from the genetic components of yield. In the next two sections, we describe soybean markets, relative maturity, and yield to set the stage for the GGP metric development. We then describe the salient aspects of the metric development. In Appendix A, we present the important details and assumptions of the GGP metric. Each year, the GGP metric adjusts the yield for every experimental variety by relative maturity group for all experimental varieties evaluated in either first, second, or third regional field trials. We adjust the yield by using the estimate of the environmental effects based on the check varieties for each relative maturity. Finally, we describe in detail the results and conclusions of employing the GGP metric. The improvements to the operational decision process provided by the GGP have led to genetic contributions to yield in North America that range from +2.4 to +1.6 bu/ac per year. ~~We estimate that without GGP, Syngenta would have needed to invest an additional $250 million to realize the same degree of improvement to genetic gains.~~

~~The uniqueness of this work is not in the elegance of the mathematics. Rather, it is in the elegance of the application for an industry that has long needed such a metric.~~

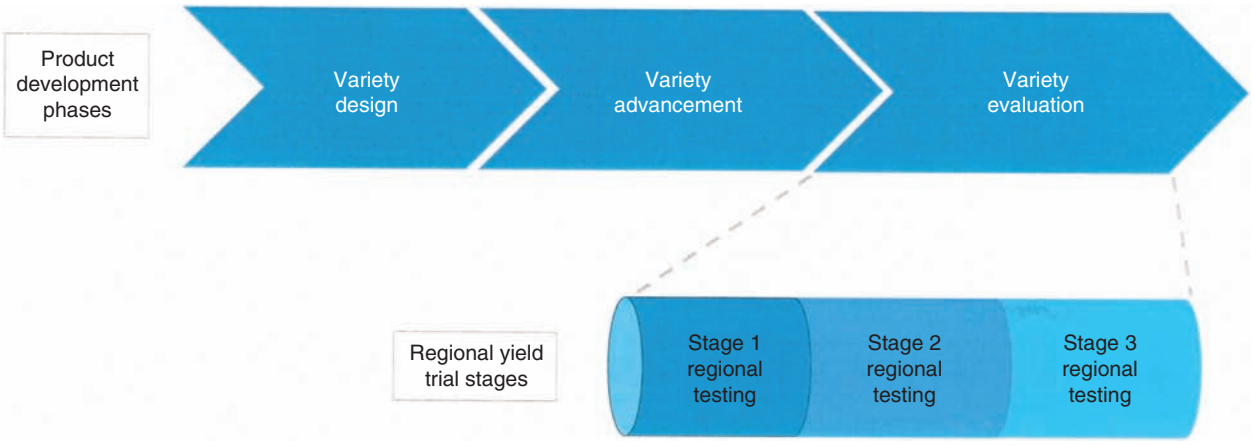
Soybean Seed Development Process

Syngenta's Soybean Product Development (SSPD) program is organized into teams. Each team is led by a highly trained and skilled plant geneticist—a plant breeder. SSPD breeding teams are geographically dispersed, and each team is responsible for the development of commercial varieties adapted to its specific geography.

SSPD breeding teams use a step-by-step product development process to create and develop new varieties in a breeding pipeline. The process is completed in six to seven years and consists of three phases: variety design, variety advancement, and variety evaluation (Figure 1).

1. *Variety design* (*one year*): The purpose of variety design is to create genetic variability. During this initial phase, the plant breeder mates two soybean varieties (parent lines) to produce a population of

Figure 1. (Color online) Syngenta's Soybean Product Development Pipeline Has Three Phases



Note. We applied our GGP metric to the three stages of regional trials within the variety evaluation phase.

progeny (offspring). ~~Half of the progeny's approximately 46,000 genes are inherited from each parent~~. As such, each progeny contains a unique combination of genes, resulting in a unique combination of characteristics (traits). ~~The plant breeder chooses the parent varieties for mating based on their performance relative to other soybean varieties in evaluation trials, or because they carry favorable traits.~~ Desirable progeny selected by the plant breeder advance to the next phase.

2. *Variety advancement* (*two to three years*): The purpose of variety advancement is to exploit the genetic variability created in variety design. During this phase, progeny that have the appropriate combination of favorable traits advance via self-pollination (mating of a plant with itself) for three or four generations. Self-pollination ensures that the favorable traits become permanently fixed within the new variety. The selected progeny become experimental varieties and advance to the next phase.

3. *Variety evaluation* (*three years*): The purpose of variety evaluation is to compare and select from among the experimental varieties that have come from variety advancement. During this phase, the SSPD breeding teams measure yield and other important traits in each of three trial stages. Trials are replicated across and within different locations. Commercial varieties are included in the trials to serve as benchmarks for performance. Only those experimental varieties that outperform current commercial varieties become new commercial varieties. As an industry, we have lacked adequate measures of genetic gain during this phase; this is where we apply GGP to make more favorable operational decisions throughout the year, such as experimental designs for comparison tests in future stages.

The Breeders' Dilemma

Measuring genetic gain presents plant breeders with a dilemma. Each phenotype (variety) that is measured for a trait (such as yield) has a genetic component (genotype) and an environmental component. Only the genetic component responds to selection by the breeder, which results in genetic gain. The problem is that plant breeders grow phenotypes under uncontrolled and unpredictable environmental conditions. For example, rainfall amounts and timing are never the same across the six to seven years required for variety development. If plant breeders can adjust for the environmental variation, they can isolate the genetic component and manipulate it for genetic gain.

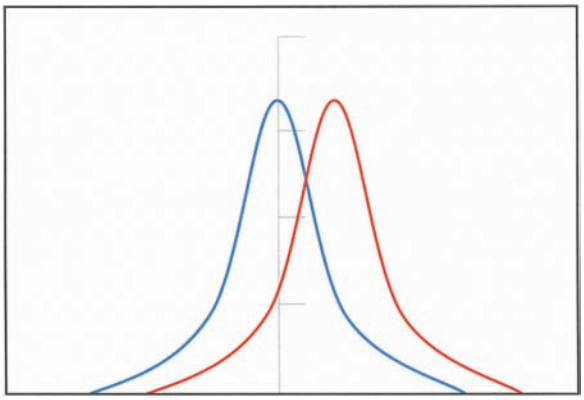
The concept of genetic gain (GG) is not new. Fisher (1930) first used GG to describe the roles of natural selection, genetic variability, and phenotypic (characteristic) variability in predicting evolution. The theory was quickly adopted by plant, animal, and microbial breeders and is often called "the breeder's equation." This theoretical model characterizes genetic gain as the product of heritability (the proportion of the phenotype variability due to genetic variation) and selection differential (the average amount of superiority of the progeny selected by the breeder to be advanced to the next generation).

The breeder's equation assumes heritability is known for any given trait, such as yield, but this is not the case. However, breeders can estimate heritability by growing the selected progeny and the original population in the same environment. This is not practical in a plant breeding program because it takes several years to develop the final selected progeny (new commercial varieties) from the initial breeding populations, as we discuss in the *Soybean Seed Development Process* section. Estimating heritability every six to seven years to determine genetic gain would not give us timely assessments to measure the yearly impact of operational decisions by the breeding teams.

A more useful measure of genetic gain would be one that assesses annual changes in genetic potential between each year's experimental varieties, rather than simply measuring the potential of the initial population and the final, selected varieties years later. For example, an assessment of annual genetic gain can be defined as the difference between the yields of two sets of experimental varieties grown in consecutive years (Figure 2).

Assume both distributions are normally distributed. If the difference in measured yield between the two distributions is due only to genetics, then we could use a simple parametric statistic, such as Student's t-statistic (Snedecor and Cochran 1989), to measure GG. The t-statistic would determine if genetic gains (or losses) between the average measured yield for the two sets of varieties are statistically significant. However, the measured yield is the result of both genetic and nongenetic

Figure 2. (Color online) This Notional Depiction of Distributions Represents Two Sets of Varieties Measured for a Trait of Interest, Such as Yield



components. Thus, a difference between distributions is still confounded by the different growing seasons in which the two sets of varieties are evaluated.

The operational challenge has been to separate nongenetic influences of each year's growing season from the genetic components of yield. ~~Sources of nongenetic variation can include environment, plot and row variation, crop management, and other factors.~~ The goal of this project was to meet that challenge and obtain unbiased yearly estimates of genetic gain.

Soybean Markets, Relative Maturity, and Yield

The soybean market area in North America is divided into segments based on the length of the growing season. Latitude is the primary determinant of the duration of the growing season. Daily temperatures and the timing of precipitation within that latitude also influence the growing season. Because of this, the seed industry groups varieties by relative maturity (RM) rather than simply by latitude. Doing so helps breeders ensure that each variety can be fully grown in the market segment to which it is best adapted.

The RM groups range from RM 000 (earliest maturity) to RM 7 (latest maturity). RM groups can be further subdivided to indicate gradations within the group by adding a decimal to the RM number. For example, varieties within RM Group 1 can range from 1.0 (earliest maturity within the group) to 1.9 (latest maturity within the group). The distribution of these different RM groups is shown in Figure 3 (Specht et al. 2014).

When varieties from different relative maturities are grown at the same location, the later-maturing varieties will usually out-yield the earlier-maturing varieties. Based on Syngenta's analysis of internal data, for each 0.1 increase in relative maturity, there is a corresponding average of 0.6 bu/ac increase in yield. For example, for varieties grown at the same location, RM 2.4 varieties yield 0.6 bu/ac more than varieties with an RM 2.3. This difference becomes greater as the spread in RMs between the varieties increases. Therefore, it is important to compare yields within a specific RM group.

Development of the Genetic Gain Performance Metric

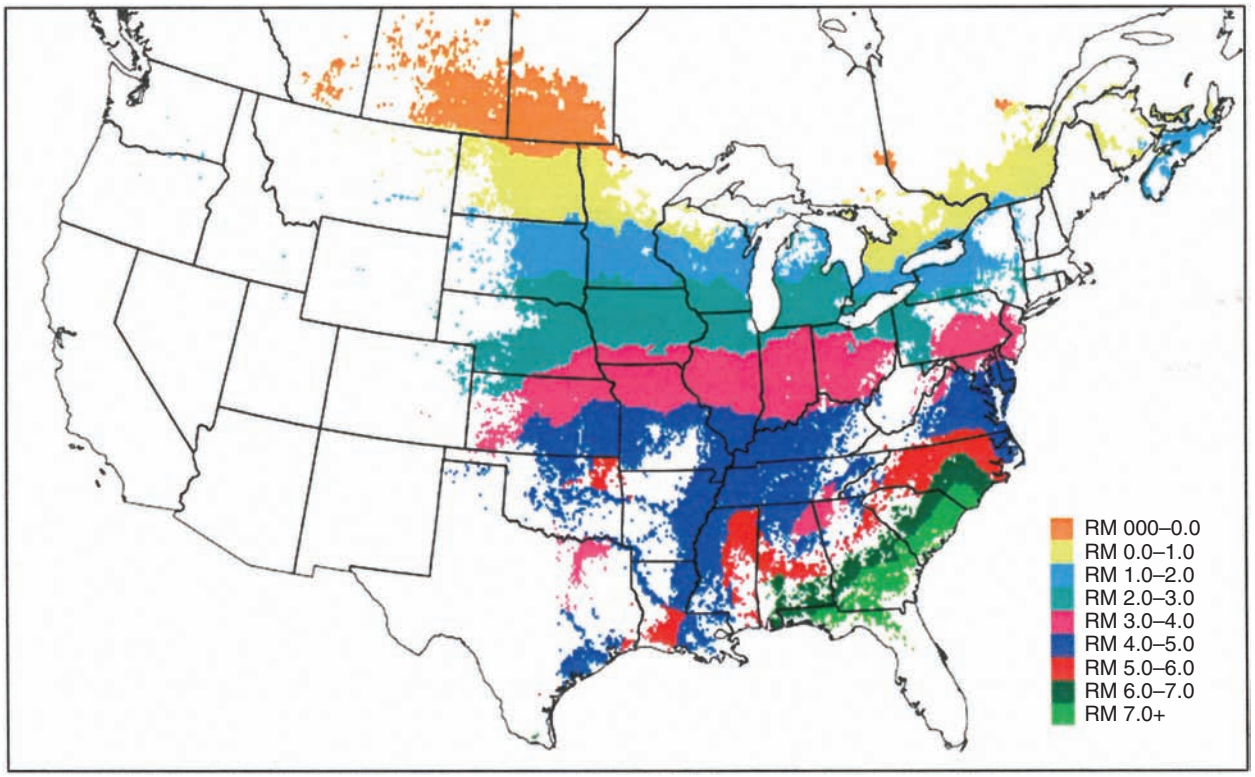
When we evaluate varieties, our objective is to assess the yield across a range of agronomic conditions that are representative of conditions encountered in real-world farm fields. Because field conditions are variable within each geographic site, we use an augmented lattice design (Cochran and Cox 1957) for all regional yield trials. In our lattice design, we assign experimental varieties to each block. We then augment each block with a set of check varieties that have similar RMs. In each block, we use at least four check varieties with similar RMs. The check varieties must be commercially successful and must have been tested for at least three years in yield trials. At each stage of regional testing, we plant hundreds of these yield trials in major soybean geographies across multiple testing locations. Each field trial matures through a set of environmental conditions that is unique to its location.

At harvest, we collect data for each variety. To do this, we use mechanical harvesters that automatically record the variety, yield, and moisture of each plot. We transfer the plot data to a centralized SSPD database and run routine statistical analyses on the data to assess the data quality.

In the final step of evaluation, we compute average varietal yields, assign RM values, and analyze other harvest characteristics, such as pod color and plant lodging (displacement of stems or roots from their vertical and proper placement).

In these trials, the yield we measure is a function of three basic parameters: varietal genetics, growing environment, and random variation. These three components are used in the GGP algorithm, which we

Figure 3. (Color online) Soybean Markets Are Segmented by Relative Maturity (RM) in the United States and Canada



Note. RM ranges from 000 in Canada to 7+ in the coastal regions of Florida, Georgia, and South Carolina.

describe in Appendix A. We provide an example of the convergence portion of the algorithm in Appendix B. Briefly, we adjust the yield for every experimental variety grouped by relative maturity each year for all experimental varieties evaluated in either first, second, or third regional field trials. We adjust the yield using the estimate of the environmental effects, based on the check varieties for each RM group.

Implementation of the GGP Algorithm

With the implementation of the algorithm in 2010, we were able to measure the impact of operational decisions on GGP during the Evaluation Phase (Figure 2). Previously, we were unable to track changes in genetic gain and intervene with operational improvements on an annual basis.

Syngenta retains yield data from past regional trials in the SSPD database. We can use the data retained before 2010 in the algorithm to estimate genetic gain prior to implementation of the GGP. This retrospective analysis allows us to evaluate changes in genetic gain before and after implementation of the GGP. Our data show that the use of GGP has helped to significantly improve the effectiveness of operational decisions as experimental varieties advance through the regional trials.

Evaluating the Effectiveness of GGP

To evaluate the effectiveness of GGP, we obtained RM and yield data from the SSPD database. We collected these data for all experimental and check varieties tested in regional trials from 2006 to 2015. We partitioned the relevant metadata associated with these trials by regional trial stage and relative maturity. Estimates of the environmental effects for years 2006–2015 were calculated with the GGP algorithm, which we describe in Equations (A.1)–(A.4) in Appendix A. Those estimates were then used to calculate adjusted yields of the experimental varieties, Equation (A.5), also shown in Appendix A.

In our analysis, we combined the values of adjusted yields for all varieties in a given growing season's

regional trial, using weighted averages for the groups in the metadata. We used the weighted averages to evaluate annual genetic gains by relative maturity within each regional trial stage. For the remainder of this paper, we discuss the impact of implementing the GGP algorithm on operational decisions to improve genetic gain for RMs, which define key market segments (Figure 3).

Results

Estimated Environmental Effects

To estimate environmental effects, we obtained more than 400,000 yield measurements of commercial varieties representing all RM groups. We obtained these measurements from all regional trials conducted from 2006 to 2015. The measurements were needed to estimate environmental effect for each growing season, even though the data were not evenly distributed among RMs or growing seasons (Figure 4). The ordinates of the graphs are the number of measurements (the count). The variable distribution among RM groups and years reflects differences in the intensity of yield testing among RM groups (Figure 3). The increase in counts also reflects the greater emphasis on obtaining estimates of environmental effect after 2010.

Estimates of environmental effects tend to be larger in absolute value the years before 2010, and larger for RM groups greater than 4.0 (Figure 5). The ordinates of the graphs are the estimated environmental effects (bu/ac). Except for the RM range 4.0–5.0, we observed that the larger estimates of environmental effects tend to coincide with smaller numbers of commercial varieties used in the algorithm. In years with small-magnitude estimates of environmental effects (i.e., 2008, 2010, 2011, and 2012), the estimates are both positive and negative. In contrast, in years with large-magnitude estimates of environmental effects (i.e., 2006, 2009, 2013, and 2014), estimates tend to be consistently positive or consistently negative across all RM groups. Estimates of environmental effects in 2007

Figure 4. (Color online) The Number of Yield Measurements Varies by Year and Relative Maturity (2015 Not Shown)

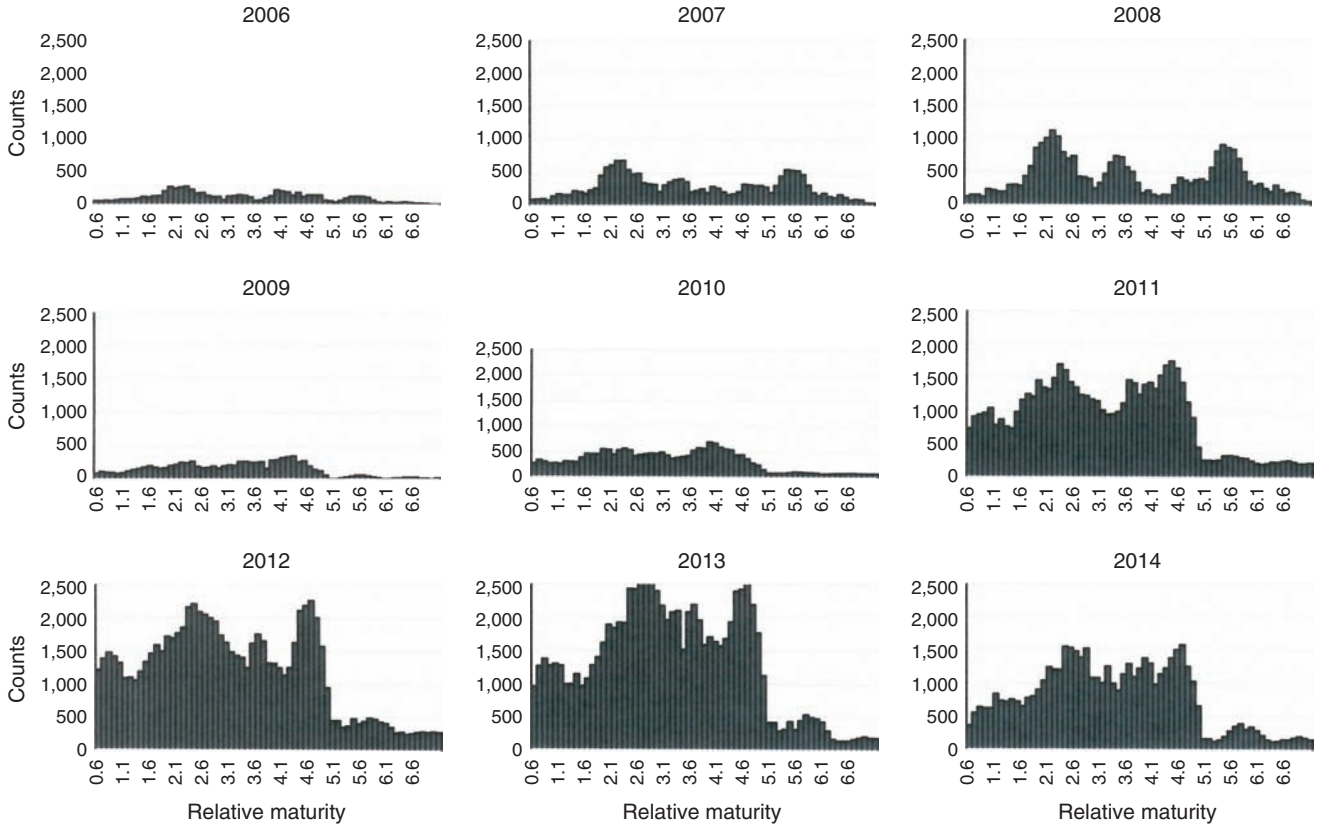
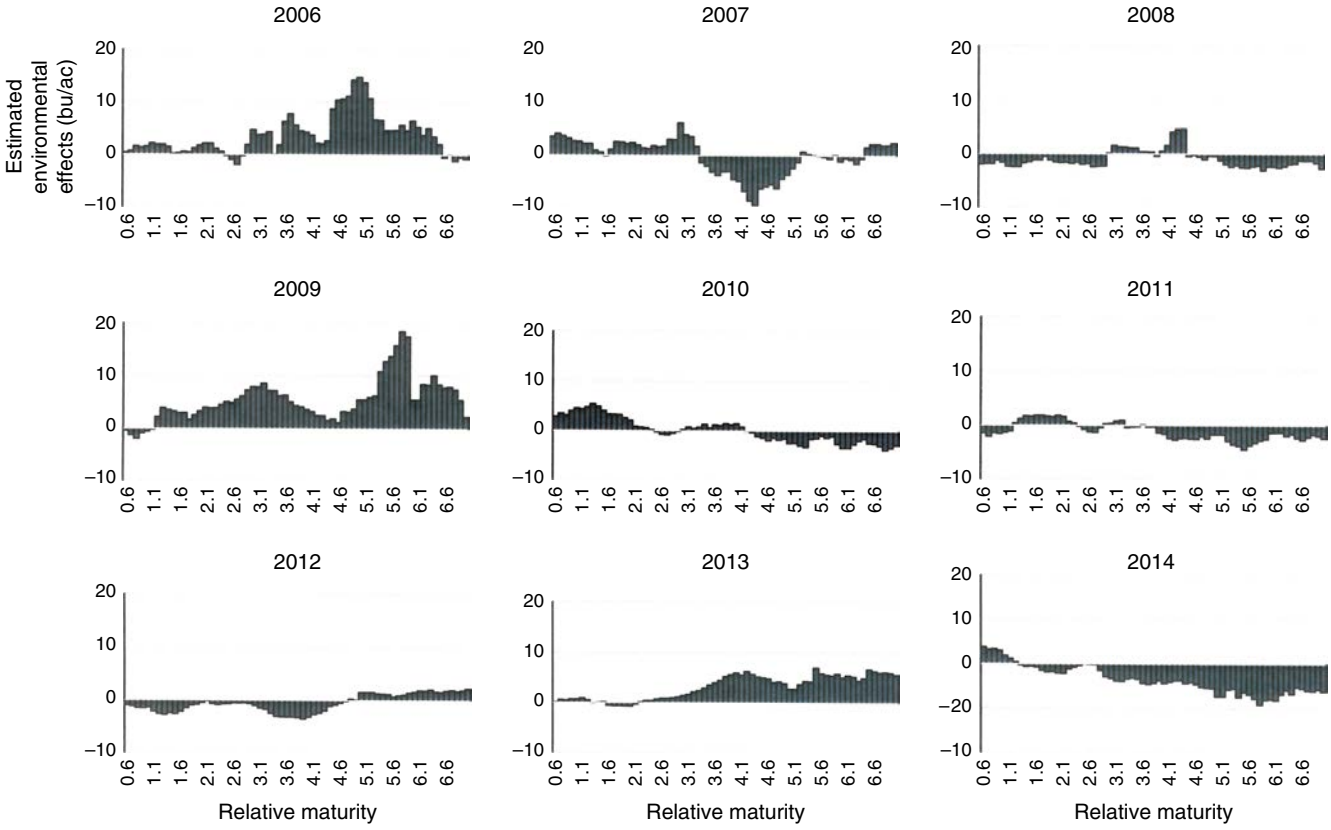


Figure 5. (Color online) The Estimated Environmental Effects (bu/ac) Vary by Year and Relative Maturity Within Years (2015 Not Shown)



tended to have relatively large negative values in the mid-range of RM groups.

Adjusting Yield for Environmental Effects

From 2006 to 2015, we evaluated experimental varieties in field trials:

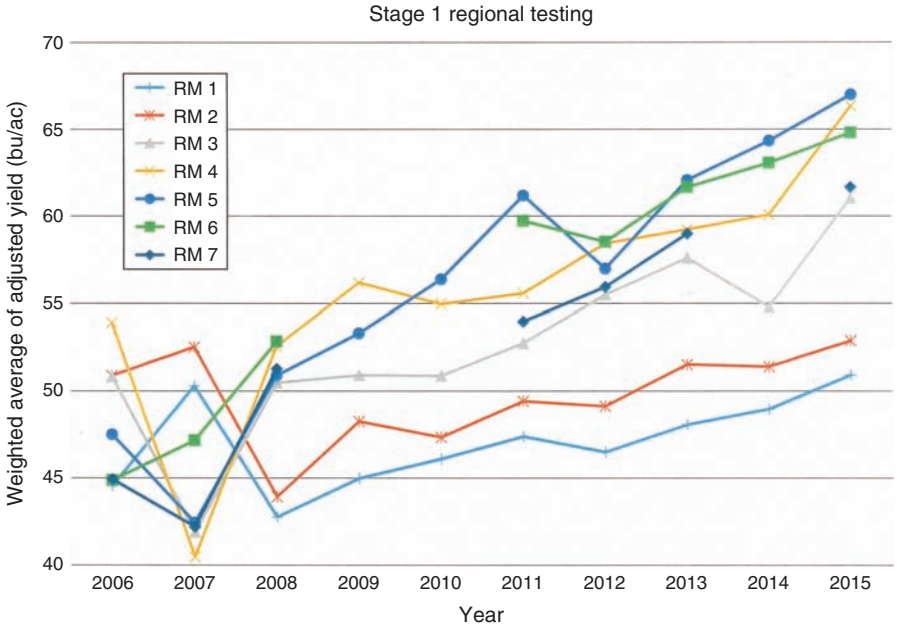
- 436,235 experimental varieties in Stage 1 regional field trials;
- 807,241 experimental varieties in Stage 2 regional field trials;
- 16,730 experimental varieties in Stage 3 regional field trials.

Of the varieties evaluated, three percent of the varieties in Stage 1 regional trials and four percent of the varieties in Stage 2 regional trials lacked RM values; therefore, we removed them from the analyses. We calculated the weighted averages of adjusted yield for the remaining varieties for each RM group by regional field trial (Stages 1–3). To examine the impact of the new metric within and across the development pipeline, we compared the results from Stages 1, 2, and 3 regional trials within the variety evaluation phase.

Among the varieties in Stage 1 regional yield trials for every RM group, we found no evidence of genetic improvement before we implemented the new metric to help make operational decisions (Figure 6). In addition, varieties in all RM groups exhibited extreme fluctuations until 2009. The discontinuities for RM Groups 6 and 7 in 2009 and 2010 are the result of not having enough data to estimate genetic gain. After we implemented the metric, we found that genetic improvement accelerated across all RM groups as we improved our operational decisions. The acceleration was more pronounced in later-maturity groups (4–7) and less pronounced for earlier-maturity groups (1–3).

Similar to Stage 1 regional trials, there is no evidence of genetic improvement in Stage 2 for any RM group before we implemented the new metric to help us make operational decisions (Figure 7). Varieties in all RM

Figure 6. (Color online) Stage 1 Adjusted Yields (Genetic Gain) Show the Favorable Effect of Making Improved Operational Decisions Beginning in 2010



groups exhibited extreme fluctuations until 2009. The discontinuities for RM Groups 6 and 7 in 2009 through 2013 are the result of not having enough data to estimate genetic gain. Those discontinuities have been excluded from this discussion. After we implemented the metric, varieties in the second set of regional field trials showed steady genetic gains for RM Groups 1, 2, and 3. In RM Groups 4 and 5, these varieties showed somewhat erratic, but positive, genetic improvement.

Stage 3 regional trials are critical because they represent the end point of variety evaluation. Successful varieties in these trials are candidates for commercialization. As with the earlier two stages of regional field trials, there is no evidence of genetic improvement for any RM group before we implemented the new metric (Figure 8). Note that there are no estimates of weighted averages for adjusted yield for any RM group in 2010, because the SSPD program was in a strategic transition. Those estimates were omitted from the analyses because there were insufficient data to estimate genetic gain for 2010.

Conclusion

In 2009, the SSPD teams began to develop and evaluate advanced analytic methods for crop development, based on methods developed in general operations research (OR) industries and deployed in manufacturing and financial systems. Curiously, after our agricultural industry initially adopted methods for genetic improvement in the 1950s (Robertson and Rendel 1950, Robertson 1957) and for farming systems (Boles 1955), OR approaches were widely ignored. There have been only a few recent notable exceptions (Xu et al. 2011, Canzar and El-Kebir 2011). However, Syngenta's SSPD program has developed and proven a suite of tools, which Byrum et al. (2016) describe, to optimize aspects of the breeding program.

Our retrospective analysis of the use of the genetic gain metric for yield revealed the following.

- Before implementation of the metric, we were unaware of the genetic contributions to yield at each testing stage.
- After implementation of the metric, we were able to track progress each year and intervene with operational improvements.
- These improvements to the operational decision process resulted in genetic contributions to yield in North America that ranged from +2.4 to +1.6 bu/ac per year.

Figure 7. (Color online) Stage 2 Adjusted Yields (Genetic Gain) Show the Favorable Effect of Making Improved Operational Decisions Beginning in 2010

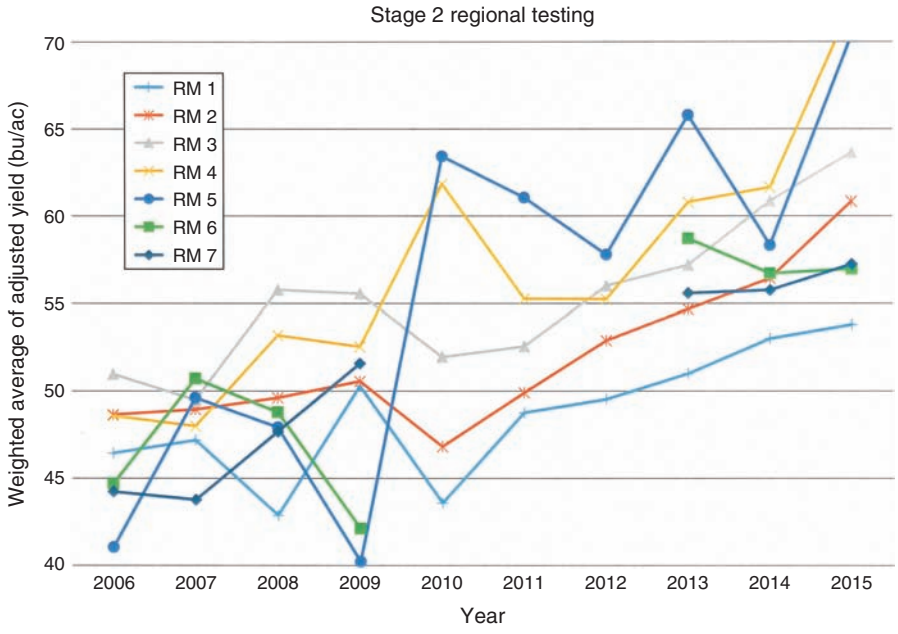
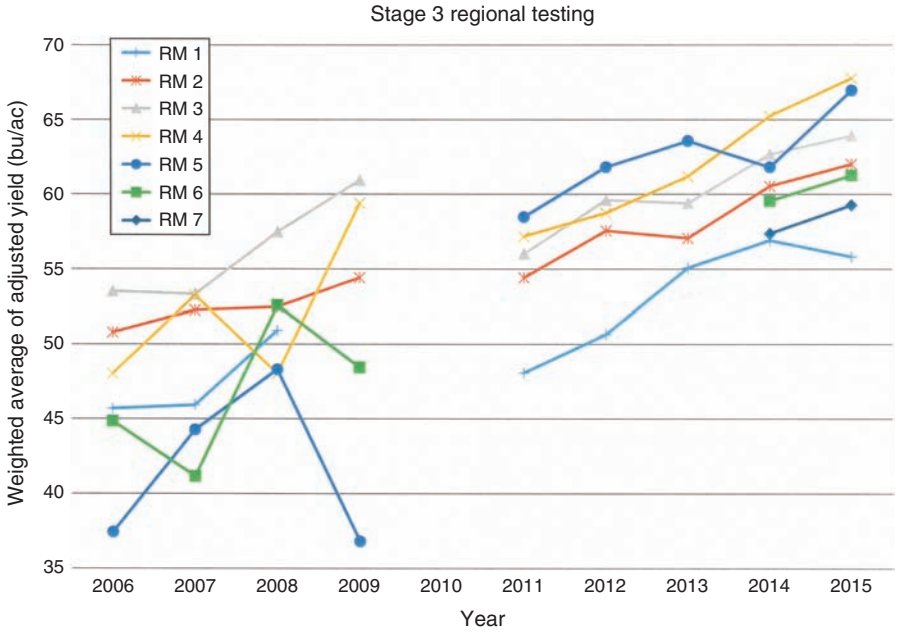


Figure 8. (Color online) Stage 3 Adjusted Yields (Genetic Gain) Show the Favorable Effect of Making Improved Operational Decisions Beginning in 2010



Syngenta's estimated genetic gains are now 40 percent greater than the gains seen prior to the implementation of the GGP in 2010. Without GGP, Syngenta would have needed to invest an additional $250 million to achieve the same 40 percent improvement in genetic gain. These additional investments would have been in two areas. First, we would have needed to double the size of the SSPD program at an estimated cost of $218 million between 2011 and 2015. Second, we avoided $31.5 million over the same period by increasing the probability of successfully delivering commercial varieties from 65 to 85 percent.

Today, Syngenta scientists continually evaluate the genetic gain performance of their projects. The successful application of the GGP metric enables them to obtain insights on where problem areas exist and proactively enact operational changes to address them. Our successful development and deployment of a genetic gain metric is accelerating agricultural productivity and helping Syngenta meet the challenge of feeding an ever-increasing global population.

Appendix A. GGP Algorithm

Environmental Adjustment to Yield

To determine environmental adjustments for yield, we average the yield values at each testing location. This gives us a single yield for each check variety c at every location. To simplify the calculations, we assume that the environmental effect at any given location is the same for all check varieties within an RM group. Note that we include check varieties in every block for all three stages of field trials. Because of this, we can use those check varieties to estimate a growing season's environmental (nongenetic) contributions to yields:

$$y_{it}^{(c)} = y_0^{(c)} + E_t + \varepsilon_{it}^{(c)}, \qquad \varepsilon_{it}^{(c)} \sim N(0, \sigma^{(c)}), \tag{A.1}$$

where

- $y_{it}^{(c)}$ is the ith measurement of yield, for check variety c, in year t;
- $y_0^{(c)}$ is the presumed yield, for check variety c, under a multitude of growing conditions; see Equation (A.2);
- E_t is the variation in the yield due to environmental conditions in year t;
- $\varepsilon_{it}^{(c)}$ represents random variability due to conditions not accounted for by $y_{it}^{(c)}$ and E_t (such as plot and row variation and crop management);
- As a random parameter, the expectation for $\varepsilon_{it}^{(c)}$ is assumed to be zero.

We can now estimate the presumed yield for check variety c, from the average difference between the measured values and the components due to environmental variation:

$$y_0^{(c)} = \frac{\sum_{it}(y_{it}^{(c)} - E_t)}{\sum_t N_t^{(c)}}, \tag{A.2}$$

where $N_t^{(c)}$ is the total number of measurements for check variety c, in year t.

The environmental effect E_t can be estimated using an expectation-maximization (EM) algorithm (Dempster et al. 1977). To determine this estimate, we apply the EM algorithm to yields for check varieties grown across environments and regional testing stages:

1. Set the initial environmental effect to zero: $E_t^{(0)} = 0$.
2. Estimate $y_0^{(c)}$ for all check varieties, using Equation (A.2).
3. Estimate all $\varepsilon_{it}^{(c)}$ values using

$$\varepsilon_{it}^{(c)} = y_{it}^{(c)} - y_0^{(c)} - E_t. \tag{A.3}$$

4. Because the expectation for $\varepsilon_{it}^{(c)}$ is assumed to be zero, we can now obtain a subsequent estimate $E_t^{(k+1)}$ using the updated equation:

$$E_t^{(k+1)} = E_t^{(k)} + \frac{\sum_{ic} \varepsilon_{it}^{(c)}}{\sum_c N_t^{(c)}}. \tag{A.4}$$

In essence, we have averaged the variability among all check varieties in year t. The deviation of their average provides a correction to the estimate of the environmental effect of year t.

5. We repeat steps 2–4 until the E_t values converge in the following sense: When implementing GGP, we stop this iterative process when the correction terms—the second term on the right side of Equation (A.4)—are $<10^{-3}$. At this point, the average residual variability among sites within a year will be less than 10^{-3}. Appendix B shows an example of this convergence.

Realized Genetic Gains

We can now compute a universal metric that represents the genetic gain for every experimental variety. To do this, we adjust the yield for every experimental variety n with relative maturity r in year t for all experimental varieties evaluated in either first, second, or third regional field trials. We adjust the yield using the estimate of E_t, based on the check varieties for each relative maturity r:

$$y_{\text{adj}} = y_t^{(n)} - E_t^{(r)}. \tag{A.5}$$

Appendix B. Example of Convergence Algorithm Developed in Appendix A

We randomly chose 2013 for RM Group 3.7 to exemplify the convergence of the algorithm in Table B.1. The final estimated environmental adjustment (E_t) of 5.37 bu/ac in the table corresponds to the same year and RM group in Figure 5.

Table B.1. This Example Illustrates the Convergence of the Correction Term to Obtain a Final Estimated Environmental Adjustment (E_t) for RM 3.7 in 2013

Iteration	Estimated environmental adjustment E_t	Correction term $(\sum_{ic} \varepsilon_{it}^{(c)} / \sum_c N_t^{(c)})$
0	0	N/A
1	4.5154	4.5154
2	5.1945	0.6791
3	5.3227	0.1282
4	5.3528	0.0301
5	5.3624	0.0096
6	5.3666	0.0043
7	5.3691	0.0024
8	5.3706	0.0015
9	5.3716	0.0010
10	5.3723	0.0007
11	5.3728	0.0004

Note. The iterative process stops when the correction term is $<10^{-3}$.

References

Boles JN (1955) Linear programming and farm management analysis. *J. Farm Econom.* 37(1):1–24.

Byrum J, Davis C, Doonan G, Doubler T, Foster D, Luzzi B, Mowers R, et al. (2016) Advanced analytics for agricultural product development. *Interfaces* 46(1):5–17.

Canzar S, El-Kebir M (2011) A mathematical programming approach to marker-assisted gene pyramiding. Przytycka TM, Sagot MF, eds. *Algorithms in Bioinformatics* (Springer, Berlin), 26–28.

Cochran WG, Cox GM (1957) *Experimental Designs*, 2nd ed. (John Wiley and Sons, New York).

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc. Series B* 39(1):1–38.

Fisher RA (1930) *The Genetical Theory of Natural Selection* (Oxford University Press, London).

Robertson A (1957) Optimum group size in progeny testing and family selection. *Biometrics* 13(4):442–450.

Robertson A, Rendel JM (1950) The use of progeny testing with artificial insemination in dairy cattle. *J. Genetics* 50(1):21–31.

Snedecor GW, Cochran WG (1989) *Statistical Methods*, 8th ed. (Iowa State University Press, Ames).

Specht JE, Diers BW, Nelson RL, Francisco J, de Toledo F, Torrio JA, Grassini P (2014) Soybean. Smith S, Diers B, Specht J, Carver B, eds. *Yield Gains in Major U.S. Field Crops, CSSA Special Publication* 33 (American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI), 311–356.

Xu P, Wang LZ, Beavis WD (2011) An optimization approach to gene stacking. *Eur. J. Oper. Res.* 214(1):168–178.

Joseph Byrum is a scientist and a business leader at Syngenta. He leads global soybean breeding, trait research, and product development. For the past 20 years, he has been actively involved in global soybean research spanning genetics, genomics, breeding, physiology, compositional analysis, grain origination, and end-user applications to benefit growers.

Joe is the chief architect of the Yield Engineering System (Y.E.S.), an ambitious plan to use mathematical principles to optimize decision making in the breeding process that has delivered $287 million in cost optimization, doubled the effectiveness of the breeding program, and advanced Syngenta's Good Growth Plan goals in a measurable way. Joe has a PhD in genetics from Iowa State University in Ames, and an MBA from the University of Michigan in Ann Arbor.

Bill Beavis is a professor and G.F. Sprague Chair for Population Genetics at Iowa State University. He also served as interim director of the Plant Sciences Institute from 2009–2014 and as director of graduate education in plant breeding since 2014. Before joining the faculty at Iowa State, Bill developed and applied novel statistical genetic methods at Pioneer Hi-Bred from 1986–1998 and was the chief science officer at the National Center for Genome Resources in Santa Fe, New Mexico, from 1998–2007.

Most often cited for his discovery of bias in estimates of genetic effects, a.k.a. the "Beavis Effect," Bill's current research interests include developing accurate predictive models and optimization of breeding processes for purposes of genetic improvement. He collaborates with OR engineers in developing optimization methods for genetic improvement projects. Bill has a PhD in plant breeding and statistics from Iowa State University, Ames.

Craig Davis joined Syngenta Seeds, Inc., in August 2004. He has served on several project teams, but it was while leading the Genetic Gain Project, whose objective is to double the rate of genetic gain in the global soy program, that he recognized the need for applying advanced mathematics and systems engineering to plant breeding. He was a senior technical advisor on two of the advanced modeling projects that contributed to the 2015 Edelman Award winning entry by Syngenta. Craig was also the soy contact for three major data mining projects conducted with a data mining contractor, resulting in the development of two tools that are undergoing validation and implementation.

Before joining Syngenta, Craig was a soybean breeder with Golden Harvest for 20 years. Craig's career was inspired by his father, who was a successful plant breeder in his own right. Craig has a PhD in plant breeding and genetics from Purdue University, West Lafayette, Indiana.

Greg Doonan is the global head, Oilseed Genetic Projects for Syngenta, based in Slater, Iowa. In the 17 years he has been with Syngenta, Greg has held a variety of roles in research and development that include platform and protocol development for high throughput systems and leading soybean trait introgression and development for several years. Through his experience, Greg brings a broad knowledge of genetics, project management, and product development to his current role.

Before joining Syngenta, Greg held a position with the University of Iowa, in which his work supported the human genome project. Greg has an MS in plant breeding from Iowa State University, Ames.

Tracy Doubler has been actively involved in agricultural crop research for over 25 years. He is the head of Seeds

Trialing and Nurseries U.S. for Syngenta based in Slater, Iowa. Tracy joined Syngenta via the Garst Seed Company in 1996. During his time with Syngenta, Tracy held several positions including genotyping lab manager, soy genetic information manager, global head of soy genetic projects, and head of North American soybean breeding projects before transitioning to his current role in January 2017.

Tracy has a PhD in agronomy and plant breeding from Iowa State University, Ames.

Von Kaster has served in various R&D capacities with Syngenta and its legacy companies for over 30 years. He brings considerable experience in the commercial development of genetically modified traits in maize and soy. He is a soybean project lead for Syngenta.

Von earned his Project Management Professional credentials from the Project Management Institute in 2008. He has managed several R&D projects at Syngenta that resulted in the successful commercialization of several key products. Von has a PhD in entomology from Iowa State University, Ames.

Ron Mowers served as a statistician for Syngenta and its legacy companies for over 30 years before retiring in 2016. He is an independent statistical consultant. During his time at Syngenta, Ron provided statistical consulting and expertise to many critical R&D projects. He also managed the U.S. and South American Seeds Biometrics groups for Syngenta. He led the doubled haploid team that increased production from 20 K to 200 K corn inbreds over a five-year period.

Before joining Syngenta, Ron was a senior statistician for Garst Seed Company from 1983–2005, an assistant professor at Louisiana State University from 1981–1983, and an instructor at Carl Sandburg College from 1973–1975. Ron has published 30 papers and secured four patents in his professional career. Ron has a PhD in statistics and agronomy from Iowa State University, Ames.

Sam Parry is a faculty associate at Arizona State University, teaching human factors engineering. He is also a consultant for the Institute for Defense Analyses and a senior fellow with ARTIS Research Inc., a scientific research group dedicated to improving understanding of collective political and cultural violence, risk assessment, and modeling through field-based research. He was a professor of OR at the Naval Postgraduate School from 1974–1998.

Sam has significant experience in industrial engineering and business analytics. He led project teams that conducted cost effectiveness and statistical analysis that resulted in international sales of Apache attack helicopters. He also developed and implemented several multivariate scoring and evaluation systems to determine the best allocation of discretionary funds for CINCPACFLEET that was adopted by the Commander of the Pacific Fleet. More recently, Sam has applied his vast experience in OR as a consultant with Syngenta, focusing on the modeling of genetic and environmental effects on soybean growth and development.

Sam has a PhD in OR and systems engineering from the Ohio State University in Wooster.