Optimizing Variable Rate Seeding in NYS

Soybean 2018 Final Report

Savanna Crossman

Impact Data

Change to Baseline Farm Data

The intent of this research has been to gain understanding of the spatial factors that interact with planting population to influence yield. This grant has helped to continue the creation and validation of an algorithm (model) that will increase profitability and productivity in corn and soybean fields. The project has been using random forest modeling, a technique that requires constant data input to further improve the model's prediction accuracy. The model was first debuted in 2016, and is currently still being developed in 2019. This grant allowed for the development of the 2019 model prescriptions and thorough analysis of 2018 results, which has added valuable data for the long-term goal of this research. As the scope and value of this project extends beyond the scope and timeline of this grant, results are presented in the context of the overall project.

The model prescriptions aim to increase in gross revenue and gross savings by increasing yield, or achieving a similar yield with lower seeding rates. By prescribing less seeds per acre, there is a cost savings to the grower. Seed is the most expensive input on the farm and impacts the amount of other inputs, such as fertilizer or nitrogen, which are needed. As the inputs are not currently included in the model, the net revenue per acre calculation is as follows:

Net Revenue Per Acre = (Yield bu/ac x Sale Price/bu) – (Seeds/ac*Cost of Seed)

In the 2016 season the model performance varied greatly field to field, but generated a gross increase in revenue of $2,710.05 on the acres that received the model. In 2017 the model performance improved from the previous year and generated a gross revenue increase of $8,656.51. In 2018 the model performance decreased to a gross loss of $1,430.58, and only improved net revenue in two fields. (Table 1)

Table 1. Model Generated Savings by Year

Farm	2016 gross savings	2017 gross savings	2018 gross savings
Swede	N/A	$5,861.00	N/A
Lott	$562.55	$1,546.84	(-)$568.706
DuMond	$2,599.38	$466.89	(-)$220.38
Holloway	N/A	N/A	(-)$641.49
Arliss	(-)$451.88	$781.78	N/A
Total	$2,710.05	$8,656.51	(-)$1,430.58

It is important to note that the 2018 season was an especially difficult one for growers due to the extreme weather that the region experienced. Planting was delayed by several weeks, which resulted in growers planting soybeans in fields that were intended for corn. This was followed by drought conditions mid-summer during pollination, and then a snowy and late harvest.

The potential statewide financial impacts of the model results are explored in the profitability, competitiveness, sustainability improvements section.

Profitability, Competitiveness, Sustainability Improvements

The preliminary results from the 2014-2018 data collection demonstrated the potential financial impacts these practices can have for growers across the State. In New York State, the general rule is that grain corn is planted at 35,000 plants/acre and soybeans are planted at 160,000 plants/acre. Producers typically use the same rate for every field across their farm.

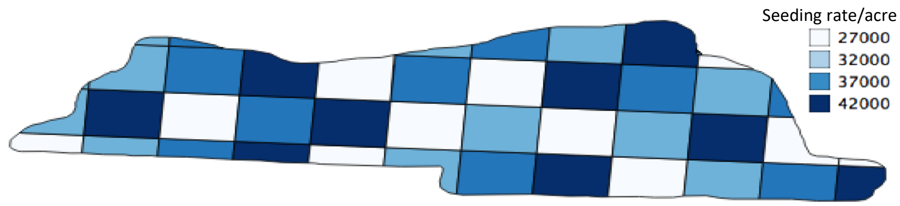

Figure 1. Experimental Block Design Planting Rate Prescription

The experimental design is intended to represent the low-end and high-end of typical grower practice in the

State (Figure 1). Four seeding rates were selected for each crop, and randomized in two-acre blocks across the fields. These random blocks need to be grown, and re-randomized, for two years in the same crop before the model can generated with confidence. Time, crop rotations, and access to high resolution soil sampling are all factors that contribute to the delay of testing the model. Due to this, a select number of fields are eligible for the model each season, though that number will snowball each season. For this section, the participating fields will be broken into two categories; Model Validation and Random Block.

Random Block

All fields in the project were planted with at least one year of the project's experimental block design, and during this stage are considered under the category of 'random block'. The planting prescription consisted of four different seeding rates randomly placed in two acre blocks across each field (Figure 1). This random block data was input into the random forest model to generate the 2018 prescriptions.

Model Validation

The Model Validation fields are smaller in number as there were several requirements that a field must meet before the model prescriptions could be tested. The field must have had at least one, ideally two years, of the same crop planted with the project's randomized block experimental design. Additionally, the field must have had high resolution grid soil sampling data that was collected within the last five years. Due to these requirements that were defined by Dr. Michael Gore of Cornell, a limited number of fields were ready for testing at the start of the 2016 year. As the research progresses each season, a larger number of fields will fall into this 'Model Validation' category each year.

The model design includes grower practice flat rate strips, alternating with strips of the model optimized rates, and randomized flat rate blocks. This unique design allows for comparison of model performance against grower flat rate in the same field, under the same conditions. It also allows for a lower financial risk to the grower during the testing phase, as only a third of the field would be in model testing. Starting in 2017, the analysis team add the strips of random blocks to allow for the data from a validation year to be used to further refine the model for the next season. (Figure 2)

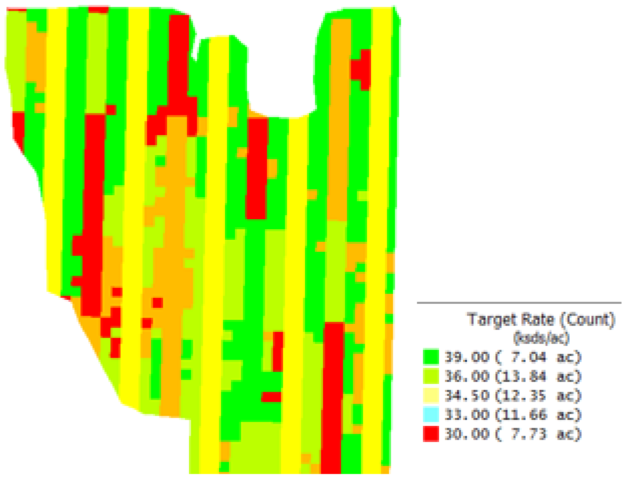

Figure 2. 2017 Model Validation Design

The 2016 planting season marked the first year of model validation. The model was tested on five fields in 2016 and only four made it to grain harvest due to severe drought stress. The results revealed that in three of the fields, there was not a significant difference in the revenue produced by the model. While the average yield of the model was significantly less, the model was able to achieve similar revenue per acre by using lower seeding rates (Table 3).

Table 3. 2016 Model Validation Results by Field

Soybean						
Field	**Variety**	**Flat Rate Average Yield (bu/ac)**	**Variable Rate Average Yield (bu/ac)**	**Flat Rate Average Revenue ($/ac)**	**Variable Rate Average Revenue ($/ac)**	**Difference**
MC 3	AG2035	46.6	44.9*	$401.62	$396.79	-($4.83)
AH01	P24T05	34.1	33.6	$432.25	$441.19	$8.94
Corn						
Field	**Hybrid**	**Flat Rate Average Yield (bu/ac)**	**Variable Rate Average Yield (bu/ac)**	**Flat Rate Average Revenue ($/ac)**	**Variable Rate Average Revenue ($/ac)**	**Difference**
Overhill	P0157	169.5	165.5 *	$572.89	$567.63	-($5.26)
Beach 2	P0216	172.9	167.5*	$1,045.36	$1,022.46	-($22.90)
Beach 2	P0533	146.1	154.4*	$854.58	$918.74*	$64.16

*Denotes statistical significance

The 2017 results showed stark improvement from those of 2016, demonstrating the power of multiple years of data combined with more typical weather patterns. In five out of seven fields tested, the model outperformed the grower flat rate with statistical significance in both yield and revenue (Table 4). Across all of the tested fields the model averaged an increase in revenue of $20.35/ac with a range of (-)$11.20/ac – (+)$100.70/ac.

Table 4. 2017 Model Validation Results by Field

Field	Crop	Flat Rate Average Yield (bu/ac)	Variable Rate Average Yield (bu/ac)	Flat Rate Average Revenue ($/ac)	Variable Rate Average Revenue* ($/ac)	Difference ($/ac)
Swede 37	Soy	53	64.8	$436.60	$537.30	$100.70
MC 2	Corn	217.9	214.4	$793.60	$782.40	-($11.20)
M 3	Corn	152.8	165.9	$526.30	$565.20	$38.90
AH 15	Corn	191.4	187.2	$685.70	$678.00	-($7.70)
AH 01	Corn	193.5	196.3	$691.10	$699.50	$8.40
Ball 1	Corn	146.1	148	$482.20	$491.10	$8.90
OverH	Corn	199.9	199.5	$731.00	$740.10	$9.10

The 2018 season was a trying one for growers across the State. Planting began in mid-May, resulting in planting dates a full month later than typical, which pushed the reproductive stages into the heat of the summer. Many trial fields were never planted as the growers were not able to traffic the fields until late in the spring. The late summer experienced drought conditions, and was followed by early snows and rain, which prevented timely harvests. Several trial corn fields were chopped as forage silage due to the poor ear set, poor stand establishment, and crops that would not fully mature. It appears that the extreme weather was a major factor in generating a large variability in yield across a field.

The results demonstrated a large spread in the data, even in the flat rate strips. This can be partially explained by the variability of moisture experienced throughout the season, and how that impacts plant development. For example low spots or high clay soils would have been difficult areas for the crop to establish, but may have provided necessary moisture during the heat of the summer. The lack of statistical significance for seemingly large net revenue differences demonstrates the large amount of variability in the data. In other words, the yield varied greatly within any given treatment, meaning larger differences in net revenue were needed to achieve a statistically significant treatment effect.

It is also worth noting that due to the conditions, the 2018 trials ended up being primarily soybean fields. The 2018 soybean results are indicating that in a year of adverse weather conditions, a fixed flat rate may have been able to compensate more adeptly than variable rate. In the results analysis, please see the attached 2018 results, it was clear that when the model prescribed rate exceeded the flat rate, profitability went down.

Table 5. 2018 Model Validation Results by Field

		Average Yield (bu/ac)			Average Net Revenue ($/ac)			Difference
Field	**Crop**	**Flat Rate**	**Random**	**Model**	**Flat Rate**	**Random**	**Model**	**($/ac)**
SW1	Soybean	57.11	57.13	57.93	397.31	403.1	399.08	1.77
SW3	Soybean	56.69	60.92	56.72	386.49	389.49	401.74	15.25*
KS1	Soybean	56.41	53.52	54.46	385.86	371.69	365.87	-19.99*
Big	Corn	223.58	218.22	214.14	790.64	738.47	765.09	-25.55*
Schoharie	Corn	215.59	209.7	213.05	758.42	745.13	739.06	-19.36*
AH01	Soybean	69.5	65.89	66.39	492.21	469.31	478.23	-13.98*
AH15	Soybean	75.23	73.81	74.96	534.96	524.07	529.31	-5.65
MC2	Soybean	72.36	70.28	71.78	513.43	517.39	506.67	-6.76
MC3	Soybean	79.41	78.33	79.27	570.06	576.4	572.48	2.42

* indicates statistical significance at alpha = 0.05

As the project continues to develop into a tool that can be utilized more widely, we are hopeful that industry wide change will come to New York State. As identified by the *Digital Agriculture Report* compiled by Van Es et al (2016), approximately 10.5% of NYS corn and soybean growers are fully equipped with the technology required to participate in this project. The total acreage harvested in 2014 was 680,000 grain corn acres and 327,000 acres of soybean (NASS & USDA, 2015). Using these numbers, there is potential for approximately 100,700 acres to adopt the model this coming 2020 spring. If all of these fields were to receive the model generated prescriptions and see the average result of 2016-2018, the result could be a $722,170.05 increase in total revenue for those acre in one year:

NYS 2014 Total Acres Harvest Commodity Corn and Soybean: 1,007,000 acres
Percent of NYS farms currently equipped to implement the model: 10.5%
Approximate acreage eligible for model implementation: 105,735 acres
Model average revenue increase ($8.02 in 2016, $20.35/ac in 2017, and -$7.98/ac in 2018): $6.83/ac
Projected total revenue increase in one year if all eligible acres adopted the model: $722,170.05

The hope would be that results such as these, which demonstrate the economic advantage of the model, would encourage more producers to become fully equipped with precision technologies. There is tremendous potential for cost savings to producers across the State. With the implementation and further testing of the model we can expect to see continued measurable changes in revenue across the participating farms.

A peripheral impact of this project relates to improvement in sustainability and environmental improvements. The project cost-shared approximately 1800acres of 1/2acre grid soil sampling on ten different operations. Grid soil sampling data can be used by the grower in a variety of ways. The most common use of grid sampling data is to generate variable rate fertilizer prescriptions. This ensures a more efficient use of fertilizer which benefits the grower economically, but also benefits the environment by reducing nutrient loss into waterways. Grid soil sampling data can also be incorporated into a nutrient management plan which provides further accountability that the growers are complying with environmental standards.

2019 Model Improvements

The most significant improvement to the model in 2019 was the addition of code to address the issue of marginal profit differences. For example, the following is the predicted yield level at the four different seeding rates for a given 10ft x 10ft cell:

(Experimental seed rate)		(Predicted yield level)
27k seed	is	244.2880
32k seed	is	251.1016
37k seed	is	251.7057
42k seed	is	251.9406

Based on the 2018 code, the model would chose '42k seed' as the best value in this situation since it has the highest predicted yield value among 4 categories. However, in this case, we only earn marginal 0.8bu/ac corn by planting 10,000 additional seed from 32 ksds/ac to 42 ksds/ac. ($4 marginal profit per acre but a $22 additional seed cost).

The previous code does not consider this 'marginal profit' problem, and would select the absolute 'highest revenue' based on 'total revenue - total cost' throughout the field. This may have led to an overestimate the 'prediction seed level' which would result in loss of profit with this prediction model. This would overestimation may especially impact profitability in years of adverse weather such as 2018.

The experimental design was also altered to increase the number of observations within a field and thus increase the power of the data set for statistical analysis. Previously the randomized block trials comprised of 2acre plots that were then reduced to subplots of 10ft x 10ft. To increase the randomization and reps, the randomized blocks were alter to randomized strips that were the width of the planter by 160 ft.

The model fields were altered from an alternating pattern of grower practice, randomized rates, and model rates to a randomized treatment design. This increased randomization helps to equally distribute any experimental error equally across treatment categories.

Knowledge Gain

Sixteen New York State growers participated in the project from 2014-2018 and eleven of these growers participated in the grid soil sampling program. This resulted in 1800 acres, ~100 site years, of data sets that included experimental planting prescriptions, topographic information, ½ acre grid soil sampling, NRCS soils information, planting and yield data.

Summary Data of Yield Driving Factors

These data were analyzed to examine the specific relationships between each factor and yield and to then create a variable rate planting prescription for each field. The results of the analysis revealed the top three most frequent drivers of yield were population, soil type, and organic matter in that order. The yield drivers that explained the greatest average variation in yield in descending order were soil type, planting rate, and calcium. On average, the analysis could explain 48.0 percent of the variation in yield with the soil and topographical features that were measured. The range was 12.2 percent to 78.6 percent demonstrating the drastic field to field differences across farms. (Figure 3)

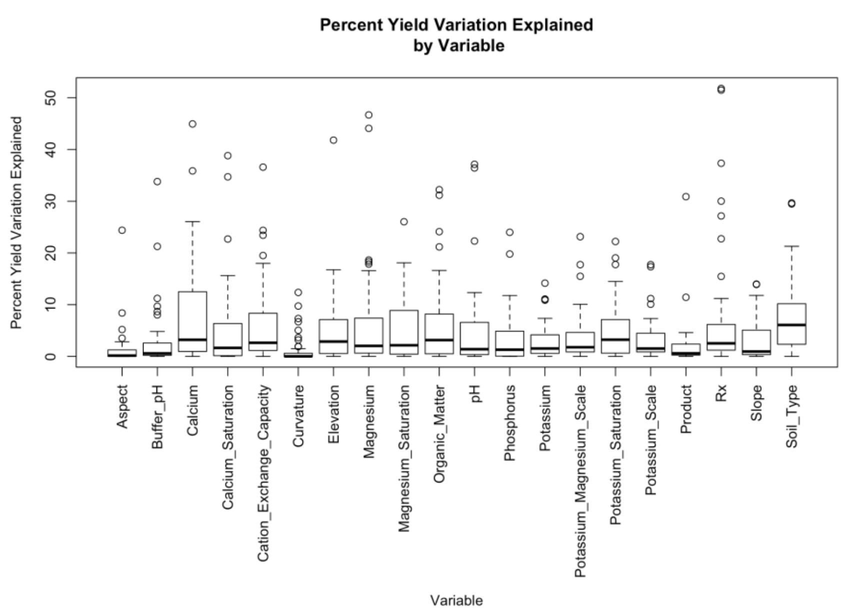

Figure 3. Percent Variation Explained by Variable Across All Site Years

The box plot gives the impression that planting rate is not a major driver in the bulk of the fields, as you can see from its many outliers. Those outliers are pulling the mean up. So in a way, although it fell in the top three for mean percent yield variation explained, it's not such an important factor in most of the fields. However, growers are not using variable rate technology to farm the averages. Variable rate technology is used to farm the extremes, to know and exploit the outliers. The outliers are what the grower might be interested in focusing on in his or her field since they are driving most of the variation in yield in individual fields.

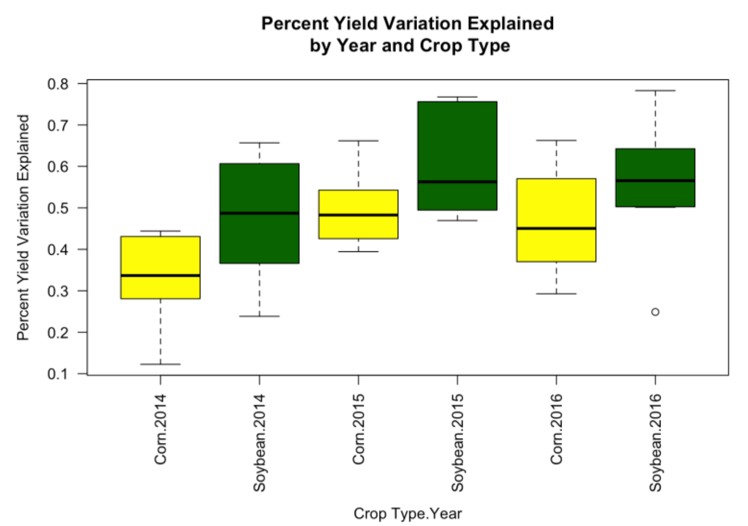

Figure 4. Percent Yield Variation Explained by Crop & Year

Across the 50 site-years, soil and topographical features were more predictive of yield in soybeans than in corn. In corn, on average we could explain 43.9 percent of the variation in yield with soil and topographical features. In soybeans, on average we could explained 54.1 percent of the variation in yield with soil and topographical features. These results might suggest that the soil and topographical features we measured drive yield in soybeans more so than they do in corn, and that perhaps there are other variables driving yield in corn that were not currently being measured, such as nitrogen. (Figure 4)

Summary Data of Revenue Maximization

For maximizing revenues in soybean, the extent of variability in the optimized designs was not consistent across site-years. In addition, for field locations that were assessed in more than one year, the cost:price ratio-based designs were less consistent across years for soybean than compared to corn. This inconsistency could be due to variations in environmental conditions across site-years (Wells, 1993), differential sensitivities to planting rate among varieties sown (Agudamu et a., 2016), or other factors not explored.

Variations in optimal planting rate designs were likewise observed for fields that were evaluated under the same crop in multiple years. The proportion of the field assigned to the same planting rate when using each year's data to build the optimized designs was low to moderate. This was particularly true for soybean: optimizations for the same field developed from different years' data shared roughly 30 percent identity in assigned rates, which is comparable to the expectation of similarity under when rates are assigned at random.

Further long-term testing is needed to more accurately assess the extent to which year-to-year variation impacts optimized designs for both corn and soybean.

Weather

Linear regressions where conducted to evaluate how weather could impact seeding rates. Previously, the model only considered soil factors (nutrients, soil type, etc.) and topography (slope, elevation etc.). Intra-annual weather variation during the growing season (wet springs or a dry August) does not evenly impact yield across a field. This can be due to topography, soil texture (sand, silt, and clay), drainage, and other factors—for example, some parts of a field may pond while others drain more quickly.

To better incorporate these types of effects, the model was adjusted to include daily measurements of precipitation, temperature, and vapor pressure deficit using PRISM climate data. PRISM data is based on weather stations and is the official dataset of the USDA. We included PRISM data in relation to key corn life cycle stages—emergence, pollination, and ear fill. The average dates corresponding to these life cycle stages were developed in consultation with several AAA growers. Corn variety was incorporated, as was the influence of the geographic location within a field (e.g. latitude and longitude).

It was found that in 2017 these weather factors accounted for only ~6% of the yield variation observed. As 2017 was categorized by growers as having typical weather conditions, this relatively low influence on yield is supported.

Meta-Data Model Evaluation

Current commercial variable rate seed algorithms are used to create prescriptions in fields that have never been tested in. Many academics, such as Dr. David Bullock, University of Illinois, believe that the main problem with existing models is their *limited inference space*; it is difficult to run a field trial "over there," and know what the implications are for management "over here"—even if "here" and "there" are across a road, let alone across a continent. All the categories of decision-tool analytical engines suffer from key limitations, which very much affect their practical efficacy. *Ultimately, the limitations come from a lack of data from sufficiently varied input application strategies, collected close enough to the location of the site for which management information is sought.*

While local testing is likely to always provide the most accurate optimized rates, the ability to extrapolate optimization models onto untested fields is of great interest in the commercial realm. Overall, using the fitted random forest regression models to predict across site-years yielded low to moderate prediction accuracies. The site-years with the highest mean prediction accuracies across all other site-years were Mc3_2015 and

Du3_2015 with accuracies of 0.11 and 0.20, respectively (please see attached document 'Manuscript'). Low accuracies are perhaps expected given the variation across site-years in the direction and magnitude of the relationships between predictor variables. However, there may be alternative approaches for building the most useful training sets for a given untested field. For example, there was a significant negative correlation between the across site-year yield prediction accuracies and the Euclidean distances between site-years in terms of random forest regression variable importance.

Provided yield, topographical, and soil data are available for an untested field, random forest regression models predicting yield could be fit, and the resulting variable importance measures could be utilized to identify the most similar site-year evaluated under the randomized block design to use as a training set. Additionally, data from multiple site-years could be combined in an optimal fashion to maximize prediction accuracy. Heslot et al. (2013) provides an innovative approach for this type of training set optimization in the context of plant breeding that identifies and removes less predictive site-years from the complete set of site-years used to train a combined model.

Another key limitation of machine driven predictions is the lack of agronomic assumptions. The next phase of model building should incorporate assumptions, such as growth limiting ranges for phosphorus, and weights for more important yield driving factors. As we are among the first in academia to conduct long-term variable rate seeding research, there many approaches should be explored in future work with this project.

Grower Education

It is also important to recognize that this project also brought grower level knowledge gain in terms of technology proficiency. Approximately two thirds of the growers who participated this project needed assistance learning how to upload and run variable rate prescriptions in 2014. At the end of the season, many of these same growers struggled with downloading data from the display and transferring it electronically to the project.

The Project Coordinator visited each participating farm and gave in person demonstrations of the process. Anytime the growers needed assistance, the Project Coordinator provided assistance over the phone or in person. By the end 2018 all but two of the participating growers were able to upload and run prescriptions, download data off the display, and transfer it electronically to the Project Coordinator with little to no assistance.

In 2018 many growers began using cloud-based data services such as Fieldview or Climate, which significantly increased the efficiency of data transfer between parties. Being proficient in the use of the precision technology was a significant hurdle to overcome, and one which undoubtedly brought whole farm improvement.

Summary Data

The results of this long-term study provide important insights into the relationship between planting rates, yields, and revenues that may serve to inform management decisions. The differences in variability between the designs optimized for yield and revenue for corn suggest that the marginal gains in yield from planting at higher rates may only be profitable when the ratio of the cost of seed and the price of corn is favorable to growers. As this ratio becomes less favorable, the optimized designs trend towards fixed or near-fixed rates consisting of the lowest planting rates. Growers should therefore factor the current seed costs and commodity prices into the decision to plant corn at a variable rate.

For soybean, however, fixed or near-fixed rates appeared to provide the maximum predicted yields for most site-years. Planting rate explained a much larger percentage of the variation in soybean yield in the linear modeling context as compared to corn, and it likewise exhibited a high level of variable importance in random forest regression. In terms of maximizing yields, these results suggest that it may be more beneficial to sow soybean at a fixed rate, though identifying the appropriate rate given the selected variety and field conditions appears to be important, as the dominant assigned rate was not necessarily the same across all site-years.

Outreach

The project staff recognized early in the start of this grant that outreach would be crucial to the project's success. We utilized many facets of outreach aimed at both academic, professional, and grower audiences. Please see the attached spreadsheet for a compiled list of outreach efforts.

Industry Changes

Please see the Knowledge Gain section, subheading "Meta-Data Model Evaluation", for a discussion of the performance of a meta-data modeling that is currently implemented in the commercial industry. The results from our approach suggest that the current industry approach to variable rate seed prescriptions is not supported by our results.

This project has brought to light the importance of hybrid and seeding rate selection. On a farm level, this means that growers are thinking more critically about their individual fields and attempting to match hybrid and seeding rate to the unique conditions each field presents. The project has demonstrated the value of high resolution soil sampling and encouraged widespread adoption of the practice. Collecting precision Ag data on these aspects for analysis will help farms better understand their interaction and relationship to yield and given soil, climate, and resource conditions. Most significantly, however, this project has given growers the ability to take this knowledge and utilize it to make economically beneficial management decisions.

Many growers across the state are equipped with variable rate technology, however, they are not equipped with the tools to turn their precision agriculture into decision agriculture. Growers felt they had a mountain of data that they were ill equipped by the industry to utilize. Growers such as Rodman Lott, Scott Arliss, Jason Swede, Todd DuMond, and Jim Begley see the value that this project brings as evidenced by a large rotation of fields involved in the research with plans of utilizing the model on a wide scale on their operation.

Rodman Lott has been involved with the project since its inception in 2013. When asked how the project has changed how he thinks about precision ag on his farm he responded, "I have realized that implementing variable rate applications of seed and fertilizer is no simple task. Hence the need for a research team dedicated to understanding the usefulness of precision ag technologies."

While this project has provided substantial knowledge gain, the larger impact will occur as the model testing continues. On a farm by farm level, this model will continue to turn grower data into realized revenue. If the DuMond farm was able to implement the model on all 5,000 production acres that farm alone could potentially see a $40,000 increase in revenue per year. As previously explored in the *Profitability, Competitiveness, Sustainability Improvements section*, there is potential for $722,170.05 per year increase in revenue in among New York State growers.

Farm Success Stories

Jason (Jay) Swede has seen significant success with the field he has kept in the project for four years. In the early days of the project, Jay's 60acre soybean field received the randomized block design and simple analysis. This simple analysis showed the possibility of increasing his revenue per acre by ~$12/ac in 2014 and ~$40 in 2016 utilizing flat rates. The model was tested on his field for the first time in 2017 and the results reveal a $100/acre increase in revenue. This results demonstrated that the model, and variable rate technology, can far outperform flat rate planting.

Photos, Presentations, Charts, Publications

Outreach Accountability Spreadsheet

Final Report Summary Statement

Variable rate seeding technology has the potential to deliver substantial cost-savings with the right algorithm driving the technology. New York State growers have struggled for years to implement the technology with algorithms that were not developed in the challenging climate, and topography of the area. This project has partnered the academic and applied agricultural communities to create a localized precision decision making tool. Still in its infancy, the algorithm has demonstrated its capabilities to integrate high resolution soils, topography, planting, and harvest data into profitable seeding rate prescriptions. Farms are able to effectively utilize their technology and data, keeping farms competitive and profitable. This initiative provides high quality farm scale research specific to New York State, with a long term commitment to understanding the best practices for precision agriculture use and implementation.