Optimal Design of Experiments for Precision Agriculture Using a Genetic Algorithm

Amy Peerlinck and John Sheppard
Gianforte School of Computing
Montana State University
Bozeman, MT, USA
{first}.{last}@montana.edu

Julie Pastorino
Department of Computer Science
University of Michigan
Ann Arbor, MI, USA
jpast@umich.edu

Bruce Maxwell
Institute on Ecosystems
Montana State University
Bozeman, MT, USA
bmax@montana.edu

***Abstract*—Variable Rate Application (VRA) is a popular technique in Precision Agriculture used to decrease the amount of fertilizer applied to a specific field while increasing profitability, effectively also reducing environmental impact. VRA tries to determine the rate of fertilizer to apply to different parts of a field based on a variety of factors, such as precipitation, elevation, and previous years' yield. To determine the appropriate variable nitrogen application rate for a field, experiments have to be conducted that provide data on how certain parts of the field react to specific nitrogen rates. In this research, a VRA of nitrogen is applied to fields of winter wheat in Montana where these experiments require the creation of a prescription map, which creates a grid of the field. The goal of the experiments is to vary nitrogen rate application, to determine how these nitrogen rates affect yield and protein production. However, when creating these prescription maps large jumps between consecutive cells' nitrogen rates often occur, putting strain on the farming equipment. To reduce the number of jumps while maintaining even distribution of nitrogen rates across different yield and protein bins, a Genetic Algorithm (GA) is used for optimization. The GA uses a multi-objective fitness function aiming to minimize jumps and maintain stratification. The results show that the GA is effective in meeting these goals for the fields studied.**

I. Introduction

The application of fertilizer is an essential part of farming with a large influence on yield and net return. Site-specific fertilizer studies have shown that one can increase net return by analyzing field specific data to vary the application rate across a field. This is known as variable rate application (VRA) [8]. VRA is an important component of Precision Agriculture (PA), which aims to increase yield and net return by applying new technologies to the farming process. PA found its roots in the 1980s with the idea of improving fertilizer application by changing the treatment based on the area of the field, but the right equipment was not available to carry out the required experiments [16]. With the continuous improvement of farming equipment, this became less of an obstacle, and VRA experiments have become common practice. With this improvement, VRA has been used successfully in several different studies to improve efficiency of fertilizer application, thus reducing costs, confirming the initial theory [5], [17].

Even with these proven gains, a lot of farmers still use uniform rate application, meaning they apply a single rate across an entire field. This can be explained partially by the lack of precise knowledge on what influences the choice for a specific fertilizer rate and that there are several site specific factors often not being taken into account when creating variable rate prescription maps. To encourage farmers to implement VRA, more feasibility studies and a more in-depth analysis of the process would be beneficial. It is important to convince the agricultural community to adapt to this strategy, as it will decrease nitrogen application, benefit environmental safety, increase food production, and reduce overall production cost [2], [21].

Our overall research goals are focused on developing strategies for improved crop production at fine scales. One component of this research involves examining variable nitrogen rate application for winter wheat in Montana to optimize profits through experimental treatments based on field specific data. Yield and protein information from previous years are taken into account to generate a variable rate nitrogen prescription map. The available yield and protein values are discretized into bins, and each nitrogen rate is distributed evenly across each bin through random, stratified assignment. This results in a randomly stratified nitrogen prescription map (see Figure 1 for an example), which can then be used to design an experiment allowing for proper analysis on how different nitrogen rates affect parts of the field with different protein and yield values.

An experimental process is repeated each year to gather temporal data on the fields to improve predictive ability of yield, protein and net return (although, without necessarily applying random application over the entire field). Aside from this data, environmental factors and plant specific information such as precipitation, normalized difference vegetation index (NDVI), slope, elevation, and topographic position index (TPI) are taken into account for the yield and protein prediction and subsequent economic analysis.

In our studies, the yield and protein prediction process has been optimized through the use of spatial data in combination with neural networks and other machine learning techniques [15]. The economic analysis uses a probability distribution of crop prices, yield values, and nitrogen cost to calculate the net return for a specific cell in a given year to perform field profit maximization. Predictive models are updated annually using a Bayesian updating process. This spatio-temporal probabilistic

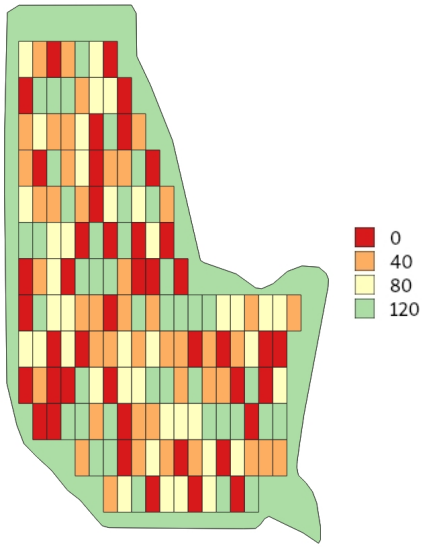


Fig. 1: Unoptimized prescription map of field davidsonmidwest. Fertilizer levels in terms of pounds of nitrogen per acre are shown in the legend.

Bayesian framework approach was shown to increase net return by \$23–\$25/ha [6]. An overview of this process can be seen in Figure 2, where the bolded stage indicates where the nitrogen rate prescription map generation is located.

Ultimately, two types of prescription maps need to be generated. So far, we have been describing the need for prescriptions to perform the necessary experiments to derive models for optimizing crop production. In fact, generating these prescriptions are the focus of this paper; thus, this work falls generally in the area known as "design of experiments" [4]. The other type of prescription results from performing the optimization based on these models. The resulting prescriptions then have the goal of yielding the optimal crop production, rather than generating data for modeling building.

When creating the experimental design prescription maps, as a result of the randomization process, there may be a large difference in the amount of nitrogen to be applied from one cell to the next, putting strain on the farmer's equipment. To lessen the wear on the machines, these differences or jumps need to be minimized, resulting in a more gradual prescription map. This problem and an example of a more desirable prescription is illustrated in Figure 3.

In this paper, we investigate the application of a genetic algorithm to generate random, stratified nitrogen applications while simultaneously attempting to minimize the magnitudes of the jumps between the cells in the field. This means that we have $\sum_{i=2}^{b} \binom{c-((i-2)c/b)}{c/b}$ combinations to consider, where c is the total number of cells and b is the number of nitrogen bins. As the number of cells in a field vary between 120 and 300 cells, we generally have at least 4 nitrogen rates, but there can be as many as 6. Then in the "easiest" case where $c = 120$ and $b = 4$, we have $\binom{120}{30} + \binom{90}{30} + \binom{60}{30} \approx 1.7 \times 10^{28}$ combinations. As we can see, an exhaustive search is impractical. For our GA, we use completed prescription maps as individual chromosomes in our population. A multi-objective fitness function is then applied, taking jumps as well as stratification into account. The specific implementation is explained in more detail in Section III.

II. Related Work

Genetic algorithms (GAs) have seen limited use in Precision Agriculture and do not appear to have been applied to the specific problem at hand. However, there have been some successful applications of GAs to other aspects of PA. For example, several PA studies have implemented GAs to optimize model parameter settings [11], navigation rules [13], and a matrix representation of a water budget [19] to be used in other processes. In order to explain variability and uncertainty in crop yield based on soil sampling density and weather data better, Pachepsky and Acock [14] use a GA to perform stochastic imaging of available soil water capacity. The results are then used to estimate variability and uncertainty in crop yield for a soybean crop model.

Also interesting in the field of PA is the use of wireless sensor networks to monitor plant health, climate, and other field specific features. Such networks provide useful information to the farmer, but they require a lot of energy. Several studies have used a GA to find the optimal sequence of sensors to improve energy consumption [1], [7], [9].

A subfield of PA is weed detection and crop management, which tries to eliminate weeds from crops. Neto et al. [12] consider methods using GAs for plant classification based on foliage. Successfully identifying whether a plant on a field is a weed through aerial images can make it easier to remove unwanted weeds from a field. Their method performs leaf extraction by clustering leaf fragments obtained from pictures and creating a chromosome out of a leaf fragment and its neighbors. A GA is applied to reconstruct the fragments of non-occluded, individual leaves. These reconstructed leaves can then be used for plant identification and mapping to improve weed control and crop management. In the same vein, Tang et al. [18] perform weed sensing based on color image segmentation, where the GA is used to identify specific regions in a Hue-Saturation-Intensity color space to segment plants from background.

Crop management also deals with identifying sick plants; in a recent study hyperspectral images were used to identify charcoal rot in soybean stems. A GA was used to select specific wavelengths from the hyperspectral image to improve classification [10]. Similar work is also being done in selecting wavelengths and generating associated filter parameters for multi-spectral imaging in produce classification [20].

III. Methodology

As mentioned, the problem addressed in this paper is the design of a random, stratified experiment for collecting data in Precision Agriculture. Our goal is to specify experimental prescriptions maps that maintain stratification over a space of prior yield and protein production in winter wheat while

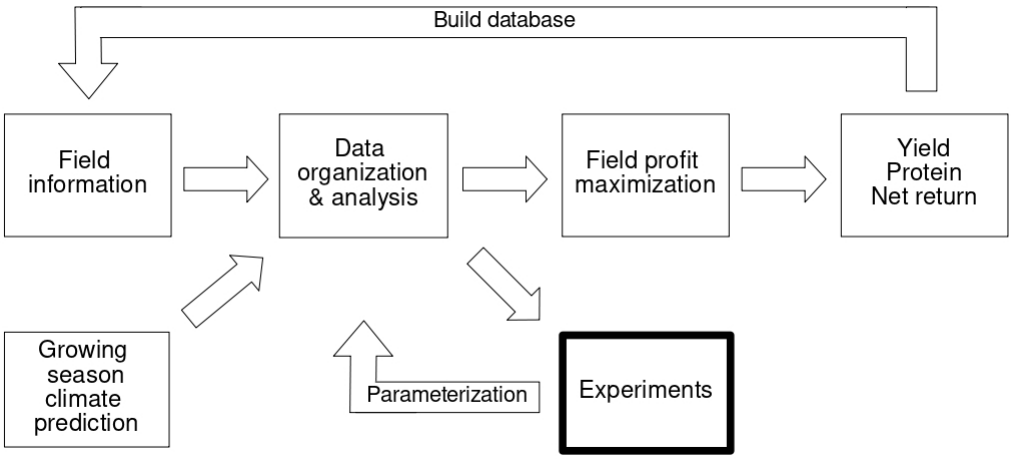
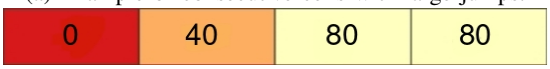


Fig. 2: Flowchart of the optimization process for field profit maximization. The "Experiments" stage is highlighted as this is the main focus of this research paper. It is the stage where the prescription maps are generated to set up the VRA experiment. Resulting data is then stored and analyzed in the economic analysis for profit maximization.

(a) Example of consecutive cells with large jumps.



(b) Example of consecutive cells with small jumps.

Fig. 3: Example of four consecutive cells in a field with large and small jumps. Red, orange, yellow and green correspond to 0, 40, 80 and 120 pounds of nitrogen/ha respectively.

simultaneously minimizing the magnitudes of jumps in nitrogen application over the map. To facilitate measuring jump magnitudes, a previous "as applied" map is used to determine the intended route the the spreader/sprayer is to follow. We hypothesize that we can apply a genetic algorithm to generate experiment prescriptions that effectively maintain stratification and minimize jumps in nitrogen rate application. In this study, we test this hypothesis by considering a variety of genetic operators and by examining the effects of these operators on overall fitness as well as individual impact on stratification and smoothness.

Specifically, we apply a GA to optimize a fertilizer experiment prescription map, which dictates the nitrogen application rate for each cell on a field. In actual use, the farmers decides which k nitrogen rates they wish to apply and how the field is to be subdivided (Figure 1). The ultimate goal is to minimize jumps (Figure 3) while maintaining stratification. For our GA, we use completed prescription maps as individual chromosomes in our population with each cell in the map as a gene and its corresponding nitrogen rate as the gene's allele. Each nitrogen rate maps to an index that is used to calculate the jump score. A multi-objective fitness function is then applied, taking both jumps and stratification into account. [3].

The stratification strategy tries to ensure that each nitrogen rate is represented evenly in each of the bins. Let bin denote a specific yield-protein combination, where yield has been divided into q bins and protein has been divided into p bins. Thus, we have a total of $k = q \times p$ bins. Then $\#cells_{bin}$ corresponds to the number of cells in the field that map to a specific bin. We then compute target stratification as

$$tstrat_{bin} = \frac{1}{k}\#cells_{bin}$$

since our goal is to distribute the nitrogen evenly over exactly k bins. The initial prescription maps are generated by computing bins based on the field's previous year yield and protein data.

We consider two different methods for discretization into bins: 1) by looking at the actual yield (yd) and protein (pro) values (called equal width binning), or 2) by splitting on the data points themselves (called equal sample binning). The first method looks at the minimum and maximum yield and protein values and creates an even split of these values based on the desired number of bins. Without loss of generality, consider yield. Then based on the number of bins q, we calculate an offset as

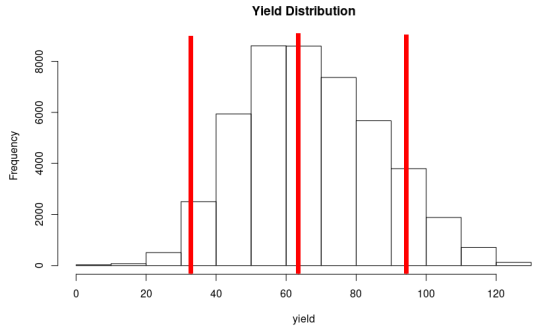
$$\mathit{offset}_q = \frac{1}{q}(max_{yd} - min_{yd}).$$

Thus, we get bin boundaries at

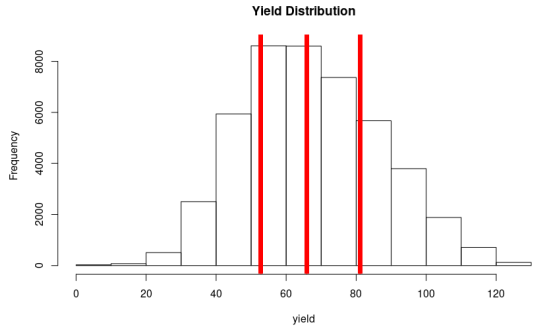
$$min_{yd}, \ldots, min_{yd} + j \cdot \mathit{offset}_q, \ldots, max_{yd}.$$

Equal sample binning, where we split on m data points, does not take the yield or protein values into account but aims to distribute an even number of points into each bin (i.e., m/q points for yield bins and m/p points for protein bins). The differences in binning strategies are illustrated in Figure 4.

The initial population consists of n prescription maps, where each map is a chromosome such that its c cells are genes in the chromosome. Each cell belongs to a bin combination $i \in \{1 \ldots k\}$ and is assigned a nitrogen rate with index $N \in \{0 \ldots r-1\}$. Once the population has been generated



(a) Equal Width Binning



(b) Equal Sample Binning

Fig. 4: Example of different bin discretization types using a histogram representation of the yield values. The vertical red lines indicate bin boundaries using each discretization type.

and evaluated, tournament selection is performed, choosing a predefined number of pairs by selecting the best map (lowest fitness score) from a chosen number of individuals from the current population. For each of these pairs, two-point crossover is performed by randomly selecting two indices and swapping the cells between these indices to create two new child prescription maps.

Finally, mutation is applied to the offspring to maintain diversity in the population. Two mutation approaches were implemented. Swap mutation chooses two random indices and switches the values of these two cells, and scramble mutation is performed by selecting all cells between two randomly chosen cells and performing a random permutation. These new maps then replace the maps in the original population with the worst fitness score.

There are several parameters that can influence the performance of the GA: the population size, the number of offspring to create, the number of candidates in tournament selection, and the crossover and mutation rates. The mutation (0.05, 0.10, and 0.15) and crossover (0.90, 0.92, 0.95, 0.98) rates were tuned simultaneously, where each combination of the two was tested. Population size (200, 400, and 800), tournament size (2, 3, 5, 10, and 20), and offspring (20, 40, 80, 100) were tuned individually; the best result was used while tuning the other parameters. After tuning, all experiments were run using

TABLE I: Chosen values for all hyper parameters. The parameters are population (Pop), offspring created (OS), crossover rate (CR), mutation rate (MR), and tournament size (TS).

Parameter	Pop	OS	CR	MR	TS
Value	400	40	0.9	0.1	3 or 20

tournament sizes of 3 and 20 to compare the behavior of the GA in more detail. The final values for each of the hyper parameters are shown in Table I.

Fitness is determined by creating a multi-objective minimization function based on jump and stratification scores. The jump score sums over the absolute difference in nitrogen levels between adjacent indices along the as-applied map of nitrogen rates.

$$\Delta jumps_i = |N(map_i) - N(map_{i+1})|$$

where $N(map_i)$ corresponds to the nitrogen index of cell i. A jump difference less than or equal to 1 for the ith cell is not added into the jump score, as this is the most desirable rate change between cells. Each individual jump score is then normalized to be within a [0,1] range:

$$F_{jumps} = \frac{\sum_{i=1}^{c-1} \Delta jumps_i}{max_{jumps}},$$

where $max_{jumps} = (r-1) \times (n-1)$ is the normalization factor. The maximum value is based on the worst case scenario where each consecutive cell goes from the minimum to the maximum nitrogen rate or vice versa (Figure 3a).

The stratification score looks for an even distribution of nitrogen rates across cells belonging to the same bins:

$$F_{strat} = \frac{\sum_{l=1}^{k} |tstrat_l - astrat_l| - min_{strat}}{max_{strat} - min_{strat}},$$

where $tstrat_l$ is the target stratification and $astrat_l$ is the actual stratification of the same bin. The maximum stratification is determined by the worst case scenario. This occurs when each cell has the same nitrogen rate, indicating that each bin only has one nitrogen rate in which it puts all its cells; therefore, every other nitrogen cell count for that bin will be set to 0.

To calculate the maximum stratification, a matrix is created of dimensionality $r \times k$, where rows represent nitrogen rates and columns represent bins. For example, assume there are three nitrogen rates and three yield and protein bin combinations for a field with 45 cells. If equal sample distribution is applied the resulting target stratification is $tstrat = 5$ for each bin, as $45/9 = 5$. Therefore, when every cell in the map is set to the first nitrogen rate, this is the resulting matrix:

$$\begin{bmatrix} 15 & 15 & 15 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This means that for the first row $r = 1$, each bin's stratification difference can be defined as:

$$\Delta strat_{1k} = r - 1 \times tstrat,$$

and for $r \neq 0$ each bin's stratification difference is:

$$\Delta strat_{rk} = tstrat,$$

which occurs $r-1$ times (i.e., for every nitrogen rate that is not the default nitrogen rate, in this case $r=1$). Both these differences occur k times, once for each bin. Summing these two stratification differences, therefore, results in:

$$max_{strat} = 2 \times tstrat \times k \times (r-1).$$

Minimum stratification can be obtained by determining the number of cells remaining after the nitrogen rates have been distributed evenly. Consider the following example: instead of 45 cells as in the previous example, suppose there are 47 cells, 16 in the first two bins and 15 in the third bin. This means that even if all nitrogen rates are distributed evenly, there will be two cells assigned an "extra" nitrogen rate, either the same nitrogen rate or two different nitrogen rates, making perfect stratification impossible. To complete the stratification, the number of remaining cells is divided by the number of nitrogen rates r for each yield and protein bin combination. To find the total minimum stratification, this intermediate minimum stratification has to be calculated and summed for each bin to determine the minimum possible stratification:

$$\min_{strat} = \sum_{i=1}^{q} \sum_{j=1}^{p} (\frac{\#cells_{ij} \bmod r}{r}).$$

Target stratification is calculated by counting the number of cells belonging to each bin $l \in \{0 \dots k\}$. This is done to determine how many cells each nitrogen rate should have for that specific bin.

$$tstrat_l = \frac{\#cells_l}{r}$$

The actual stratification for a yield and protein bin is calculated by counting the number each nitrogen rate occurs in each of the bins. For example, if we have 45 cells with three nitrogen rates and three total bins, we know the target stratification is 5 if each of the bins contains 15 cells. Suppose that the first bin of 15 cells has three low N rates, three medium N rates, and nine high N rates. The expected stratification is then subtracted from each of these counts (i.e. $|3-5|$, $|3-5|$, and $|9-5|$). The resulting values are summed to calculate the actual stratification for that bin.

The final fitness function is a weighted combination of the normalized jump and stratification scores, where $w \in [0,1]$:

$$F_{map} = (w \times F_{jumps}) + ((1-w) \times F_{strat}).$$

The GA stops running when the set maximum number of generations has been reached. In practice, the algorithm is also set to stop if the farmer deems the jumps to be low enough.

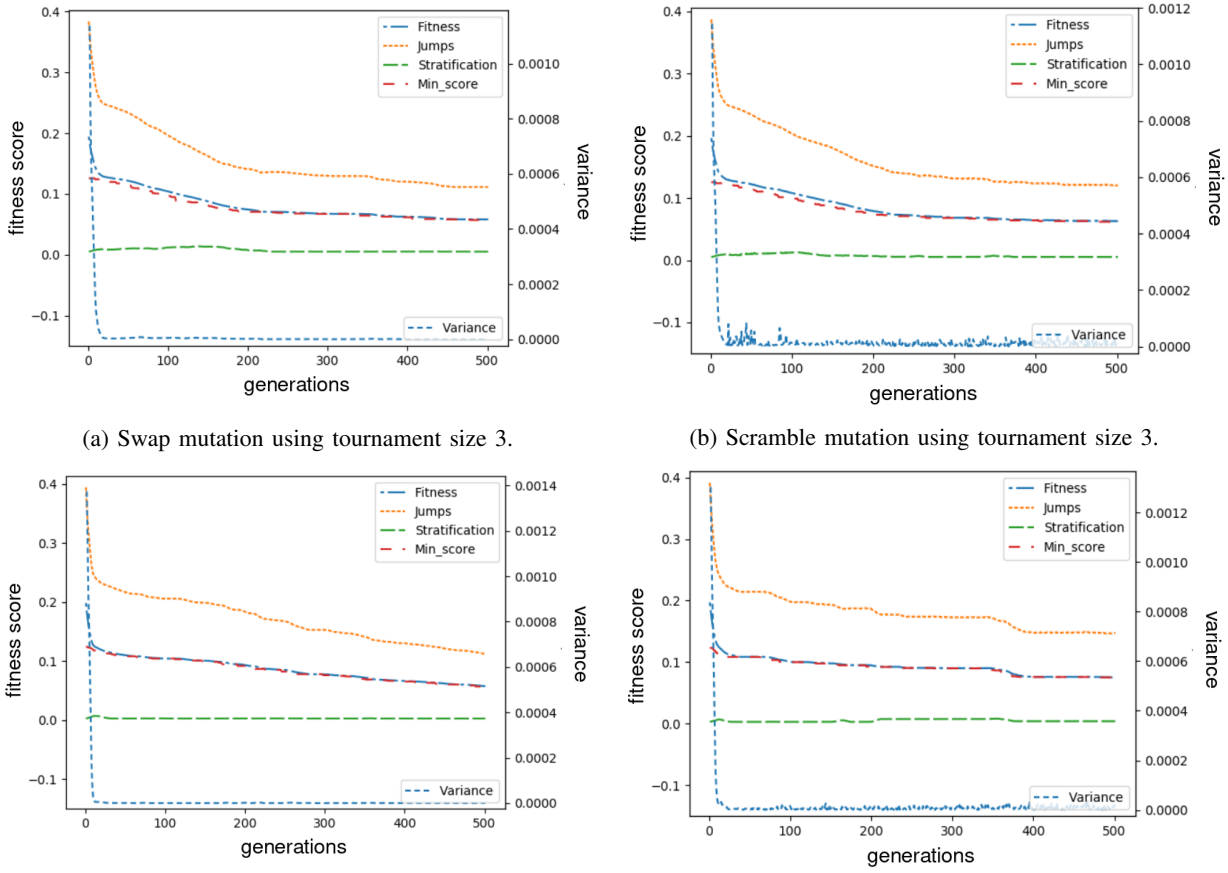
IV. Results and Discussion

In Figure 5 the jump and stratification scores give equal weight to the final fitness score. The plots show that the GA moves towards convergence, which is the desired result. However, as farmers generally do not have the luxury to wait for hours for a GA to run to find an improved prescription map, i.e. with much lower jumps, increasing the emphasis on the jump score could lead to a more practical use of the GA. Furthermore, using scramble seems to explore more of the search space as there is a slightly larger change in the jump and stratification scores for the population. More importantly, there are much larger changes in variance of the population across the generations. When comparing the behaviour of the GA with a different number of individuals for tournament selection, using a smaller number of individuals seems to reach more rapid convergence than using a larger number. As increasing the number of individuals increases selective pressure, this result seems counter-intuitive.

Figure 6 shows the results for 500 generations of the GA using swap and scramble mutation respectively and using the two different discretization methods, each using a weight of $w = 0.75$. The results again indicate convergence for both the scramble and swap mutation methods. It is interesting to note that the swap mutation seems to find lower jump scores than scramble. This might indicate that scrambling changes the maps too much, producing offspring that do not reduce the jump or stratification score. This would also explain why convergence is slower, as it would take longer to find better prescriptions. Swap mutation, on the other hand, makes smaller adjustments, thereby possibly providing maps that better maintain the overall stratification while exploring a minimization in jumps.

In all cases, there is a substantial drop in variance of the fitness scores early on in the process. The initial variance is small to begin with but becomes almost negligible after a few generations. However, a clear change in variance is evident when applying scramble mutation. This indicates that the population fitness becomes very similar early on. Considering the initial prescriptions are being created with the goal of laying out a randomly stratified prescription map for nitrogen rates based on yield and protein bins, it makes sense that there would not be much variance in the overall fitness. Once the jumps start to drop, the fitness scores would become even more similar. The plots show that the largest drop in jump score also occurs early in the process, making the drop in variance a logical consequence. The resulting fitness scores are rather small because of the way the jump and stratification scores are calculated. The maximum jump and stratification values are much higher than what most prescription maps result in, as these are worst case scenarios. The large difference between the maximum (worst case) and actual scores thus results in a small fitness score, potentially explaining the low change in variance, as well as the small difference between the average and minimum fitness score.

The average total fitness score results for 10 runs of the GA are shown in Table II. A paired t-test is performed to confirm that the scores for scramble and swap mutation are statistically different from each other for all fields and both binning methods at the $\alpha = 0.05$ level. The results using 3 and 20 individuals for tournament selection are significantly different at the $\alpha = 0.05$ level. Furthermore, the results



(a) Swap mutation using tournament size 3.

(b) Scramble mutation using tournament size 3.

(c) Swap mutation using tournament size 20.

(d) Scramble mutation using tournament size 20.

Fig. 5: Sre 1314 results for 500 generations of the GA using the two different mutation types and equal sample binning, where $w = 0.5$. The left y-axis shows the fitness score values (including jumps, stratification and the best score in the population), while the right y-axis details the variance value.

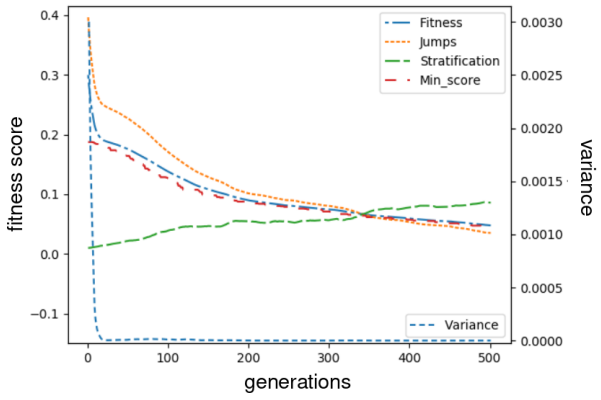
show that the fitness scores for swap mutation are consistently lower, and that the GA achieves a lower fitness score for both discretization methods. Based on these results, scores from a single run of the GA for the original prescription maps and the "best" maps, obtained through swap mutation, are shown in Table III. An illustration of an optimized prescription with minimized jumps can be found in Figure 7 under the assumption the spreader travels along the north-south axis. This illustrates a map where the jump score decreased substantially, and the stratification score only went up slightly for the final prescription map, which was the desired result.

V. Conclusion and Future Work

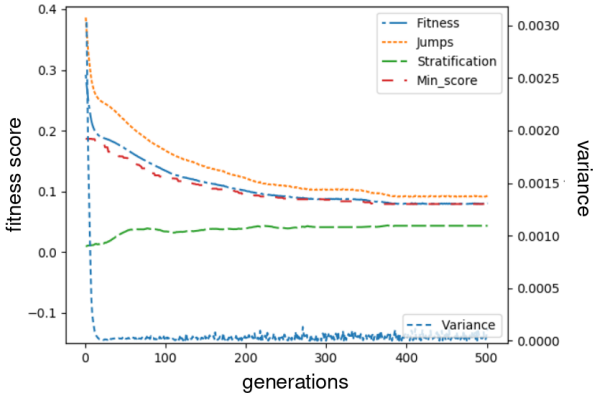
Variable rate application is an important aspect of Precision Agriculture, trying to reduce the amount of fertilizer needed across a field while simultaneously increasing yield, protein production, and net return. To be able to determine the amount of nitrogen to be applied adequately, field-specific experiments have to be performed. These experiments involve the creation of a nitrogen prescription map that overlays a grid on the field at hand and looks at previous year yield and protein values within each cell to determine what bin the cell belongs to. Nitrogen rates are then assigned randomly to cells with the goal of having an even distribution of nitrogen values across each bin, thus providing a good experimental basis to examine the influence of nitrogen rates on different parts of the field. However, these prescription maps often contain large jumps between consecutive cells, putting strain on the farming equipment.

In our experiments, a GA was applied successfully to minimize the jumps between cells, while maintaining stratification, thus supporting our hypothesis. The results show that by setting the weight of the jumps score to have a larger influence than the stratification score, a better solution in terms of jumps can be found. The results also indicate that using swap mutation improves results. Overall, the GA performs well for this particular application and achieves the set goal.

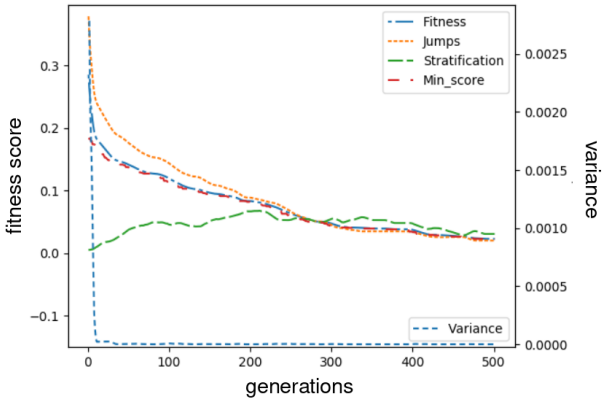
As future work, we would like to compare different GA operators in more detail, as well as explore how changing the weight of the multi-objective fitness function further influences results. We will also continue to explore the impact of additional constraints in generating prescription maps, such as



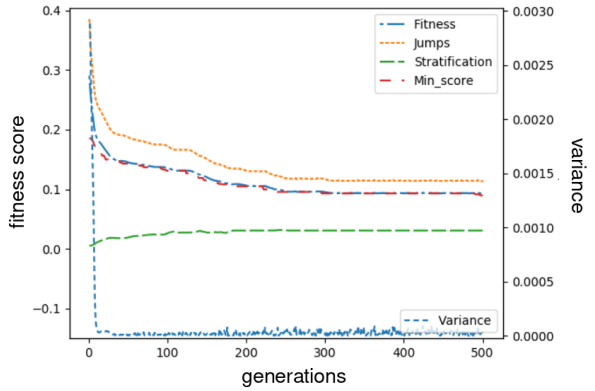
(a) Swap mutation using tournament size 3.



(b) Scramble mutation using tournament size 3.



(c) Swap mutation using tournament size 20.



(d) Scramble mutation using tournament size 20.

Fig. 6: Sre 1314 results for 500 generations of the GA using the two different mutation types and the two different discretization methods, where $w = 0.75$. The left y-axis shows the fitness score values (including jumps, stratification, and the best score in the population), while the right y-axis details the variance value.

TABLE II: Average fitness score of the best maps after 10 runs of the GA for scramble and swap mutation, using equal width and equal sample binning, on three different fields. The jump weight is set to $w = 0.5$.

		sec35mid		**davidsonmidwest**		**sre1314**	
		3	**20**	**3**	**20**	**3**	**20**
Equal Width	**Swap**	0.0282	0.0207	0.0493	0.0195	0.0582	0.0578
	Scramble	0.0520	0.0401	0.0385	0.0601	0.0627	0.0756
Equal Sample	**Swap**	0.0342	0.0309	0.0425	0.0213	0.0679	0.0525
	Scramble	0.0364	0.0468	0.0489	0.0578	0.0613	0.0744

TABLE III: Fitness, jump, and stratification scores for the initial map and the final map after one run of the GA using equal sample binning. Results are shown for three different fields, two different tournament sizes (3 and 20), and the jump weight is set to $w = 0.5$. All results are for the swap mutation method as this gave the lowest fitness.

	sec35mid				**davidsonmidwest**				**sre1314**			
	3		**20**		**3**		**20**		**3**		**20**	
	First	**Best**	**First**	**Best**	**First**	**Best**	**First**	**Best**	**First**	**Best**	**First**	**Best**
Fitness	0.2457	0.0341	0.2381	0.0308	0.1598	0.425	0.1889	0.0213	0.2492	0.0679	0.1523	0.0538
Jumps	0.4667	0.0435	0.4637	0.0493	0.3139	0.0793	0.3703	0.0370	0.4932	0.1284	0.3018	0.1047
Strat	0.0248	0.0248	0.0124	0.0124	0.0055	0.0055	0.0055	0.0055	0.0052	0.0075	0.0029	0.0028

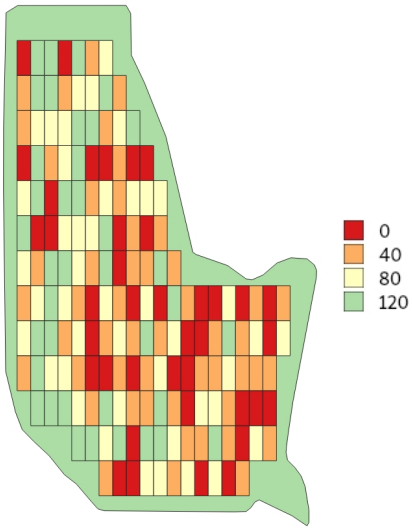


Fig. 7: Optimized prescription map of the field davidsonmidwest with equal sample discretization and $w = 0.75$. The legend indicates pounds of nitrogen per acre.

navigational constraints for the spreader/sprayer. We will also consider application of herbicides. The next step is to deploy this process for actual farmer use. This then becomes part of the workflow (Figure 2) and makes it possible to gather more data, while further decreasing cost to the farmers by lessening strain on their equipment. A key goal in this work is a multi-objective one whereby we maximize profit while also minimizing environmental impact. Therefore, based on this workflow, we will also continue to develop strategies for predicting crop production and net return.

Acknowledgment

We thank the OFPE-project for providing us with site-specific data. This project was funded, in part, by MREDI grant 51040-MUSR12015-02 and NSF grant 1658971.

References

[1] Konstantinos P Ferentinos and Theodore A Tsiligiridis. Adaptive design optimization of wireless sensor networks using genetic algorithms. *Computer Networks*, 51(4):1031–1051, 2007.

[2] Robin Gebbers and Viacheslav I. Adamchuk. Precision agriculture and food security. *Science*, 327(5967):828–831, 2010.

[3] Randy L. Haupt and Sue Ellen Haupt. *Practical Genetic Algorithms*. Wiley Interscience, 2nd edition, 2004.

[4] Charles R. Hicks and Kenneth V. Turner. *Fundamental Concepts in the Design of Experiments*. Oxford University Press, 1999.

[5] Brad Koch, R Khosla, WM Frasier, DG Westfall, and D Inman. Economic feasibility of variable-rate nitrogen application utilizing site-specific management zones. *Agronomy Journal*, 96(6):1572–1580, 2004.

[6] Patrick G Lawrence, Lisa J Rew, and Bruce D Maxwell. A probabilistic bayesian framework for progressively updating site-specific recommendations. *Precision Agriculture*, 16(3):275–296, 2015.

[7] Naisen Liu, Weixing Cao, Yan Zhu, Jingchao Zhang, Fangrong Pang, and Jun Ni. Node deployment with k-connectivity in sensor networks for crop information full coverage monitoring. *Sensors*, 16(12):2096, 2016.

[8] Bruce Maxwell, Paul Hegedus, Philip Davis, Anton Bekkerman, Robert Payn, John Sheppard, Nicholas Silverman, and Clemente Izurieta. Can optimization associated with on-farm experimentation using on-farm experimentation using site-specific technologies improve producer management decisions. In *International Conference on Precision Agriculture (ICPA)*, 2018.

[9] Mina Mirhosseini, Fatemeh Barani, and Hossein Nezamabadi-pour. Design optimization of wireless sensor networks in precision agriculture using improved bqigsa. *Sustainable Computing: Informatics and Systems*, 16:38–47, 2017.

[10] Koushik Nagasubramanian, Sarah Jones, Soumik Sarkar, Asheesh K. Singh, and Arti S. Ganapathysubramanian. Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods*, 14, 2018.

[11] Rajathi Natarajan, Jayashree Subramanian, and Elpiniki I. Papageorgiou. Hybrid learning of fuzzy cognitive maps for sugarcane yield classification. *Computers and Electronics in Agriculture*, 127:147–157, 2016.

[12] Joo Camargo Neto, George E. Meyer, and David D. Jones. Individual leaf extractions from young canopy images using gustafsonkessel clustering and a genetic algorithm. *Computers and Electronics in Agriculture*, 51(1):66–85, 2006.

[13] N. Noguchi, J.F. Reid, E.R. Benson, and T.S. Stombaugh. Vision intelligence for an agricultural mobile robot using a neural network. *IFAC Proceedings Volumes*, 31(5):139–144, 1998.

[14] Yakov Pachepsky and Basil Acock. Stochastic imaging of soil parameters to assess variability and uncertainty of crop yield estimates. *Geoderma*, 85(2):213–229, 1998.

[15] Amy Peerlinck, John Sheppard, and Bruce Maxwell. Using deep learning in yield and protein prediction of winter wheat based on fertilization prescriptions in precision agriculture. In *International Conference on Precision Agriculture (ICPA)*, 2018.

[16] Francis J. Pierce and Peter Nowak. Aspects of precision agriculture. volume 67 of *Advances in Agronomy*, pages 1–85. Academic Press, 1999.

[17] William R Raun, John B Solie, Gordon V Johnson, Marvin L Stone, Robert W Mullen, Kyle W Freeman, Wade E Thomason, and Erna V Lukina. Improving nitrogen use efficiency in cereal grain production with optical sensing and variable rate application. *Agronomy Journal*, 94(4):815–820, 2002.

[18] Lie Tang, L Tian, and Brian L Steward. Color image segmentation with genetic algorithm for in-field weed sensing. *Transactions of the ASAE*, 43(4):1019, 2000.

[19] Dennis Timlin, Yakov Pachepsky, Charles Walthall, and Sara Loechel. The use of a water budget model and yield maps to characterize water availability in a landscape. *Soil Tillage Research*, 58(3):219–231, 2001.

[20] Neil Walton, John Sheppard, and Joseph Shaw. Genetic wavelength selection and filter specification for multi-spectral image classification. In *Proceedings of the ACM Genetic and Evolutionary Computation Conference (GECCO)*, 2019.

[21] Naiqian Zhang, Maohua Wang, and Ning Wang. Precision agriculture a worldwide overview. *Computers and electronics in agriculture*, 36(2-3):113–132, 2002.