



# Cornhusker Economics

## The Impact of Different Data Processing Methods on Site-specific Management Recommendation

Market Report	Year Ago	4 Wks Ago	4-10-19
<b>Livestock and Products.</b>			
<b>Weekly Average</b>			
Nebraska Slaughter Steers, 35-65% Choice, Live Weight. . . . .	115.00	*	*
Nebraska Feeder Steers, Med. & Large Frame, 550-600 lb. . . . .	183.16	179.33	182.93
Nebraska Feeder Steers, Med. & Large Frame 750-800 lb. . . . .	146.31	151.36	153.59
Choice Boxed Beef, 600-750 lb. Carcass. . . . .	217.41	224.94	226.48
Western Corn Belt Base Hog Price Carcass, Negotiated . . . . .	46.26	48.42	74.38
Pork Carcass Cutout, 185 lb. Carcass 51-52% Lean. . . . .	67.76	62.87	80.45
Slaughter Lambs, woolled and shorn, 135-165 lb. National. . . . .	145.73	141.07	145.06
National Carcass Lamb Cutout FOB. . . . .	373.73	372.16	382.04
<b>Crops.</b>			
<b>Daily Spot Prices</b>			
Wheat, No. 1, H.W. Imperial, bu. . . . .	4.57	3.83	4.07
Corn, No. 2, Yellow Columbus, bu. . . . .	3.65	3.41	3.44
Soybeans, No. 1, Yellow Columbus, bu. . . . .	9.65	7.84	7.97
Grain Sorghum, No.2, Yellow Dorchester, cwt. . . . .	5.91	5.29	5.45
Oats, No. 2, Heavy Minneapolis, Mn, bu. . . . .	2.74	3.11	3.21
<b>Feed</b>			
Alfalfa, Large Square Bales, Good to Premium, RFV 160-185 Northeast Nebraska, ton. . . . .	*	175.00	*
Alfalfa, Large Rounds, Good Platte Valley, ton. . . . .	97.50	105.00	112.50
Grass Hay, Large Rounds, Good Nebraska, ton. . . . .	*	92.50	*
Dried Distillers Grains, 10% Moisture Nebraska Average. . . . .	160.00	145.50	153.50
Wet Distillers Grains, 65-70% Moisture Nebraska Average. . . . .	51.50	50.00	47.50
* No Market			

### Introduction

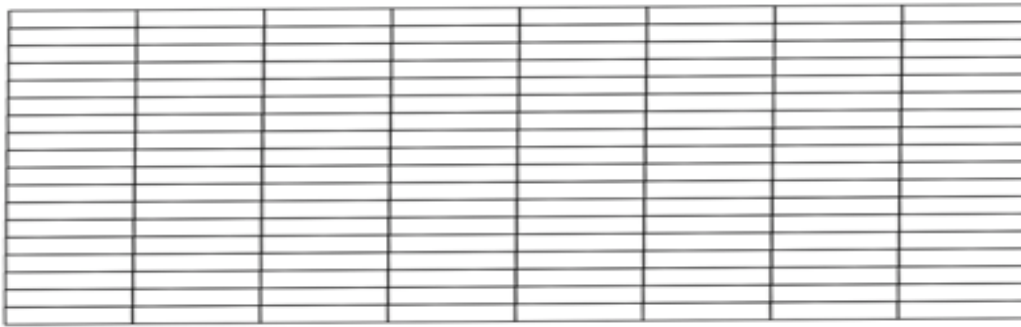
Precision agriculture has the potential to enhance farming profitability substantially via site-specific management of fields. One of the promising ways of generating such profitability-enhancing input is to use recommendation map is on-farm randomized trials. The process of generating an input (say nitrogen) using a recommendation map typically involves the following steps:

1. Design and implement randomized input use trial
2. Collect yield data along with other field characteristics (Slope, Electrical Conductivity, and Organic Matter)
3. Process the data for statistical analysis
4. Conduct regression analysis to estimate production function (how the input affect crop yield)
5. For each of the management units, find the input rate that maximizes profit for that unit

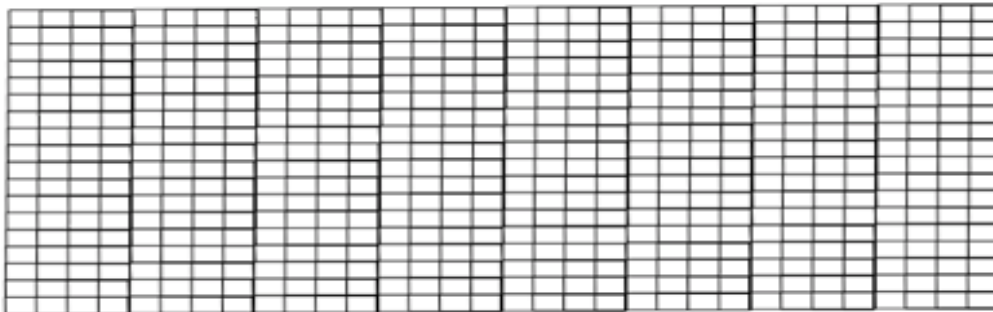
The major focus of this blog post is on step 3: we will examine the sensitivity of regression analysis (step 4) and the resulting recommendation map (step 5) to show the way experimental data is processed. Specifically, we will examine how the way you define analysis units affects steps 4 and 5:

- Method 1: use experimental trial units as regression analysis units
- Method 2: divide each of the experimental trial units into sub-units, and use the sub-units as regression analysis units
- Method 3: use yield monitor yield data points as regression analysis units

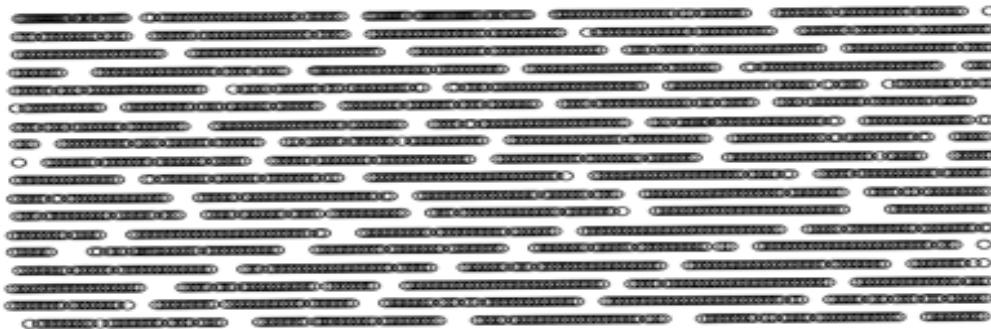
Figures below illustrates the three different types of data aggregation approaches.



Plot



Sub-plot of 4



Point

In academic research studies, all three data processing methods are used. However, to the authors' knowledge, the consequences of using different data processing methods is not well understood in the context of on-farm field trials. Indeed, there is no consensus among practitioners and researchers about the best way to define the analysis unit.

It is well known that yield data from a yield monitor have measurement errors, and the errors tend to be averaged out more when more yield data points are used to find a mean. Thus, processing Method 1 produces yield analysis points that have the least measurement errors, and Method 3 has the highest measurement errors as it uses yield monitor data points as the unit of analysis without any averaging. However, data aggregation masks important information at the same time. Suppose you suspect that electrical conductivity is an important soil characteristic indicator that affects economically optimal nitrogen rates. Electrical conductivity can

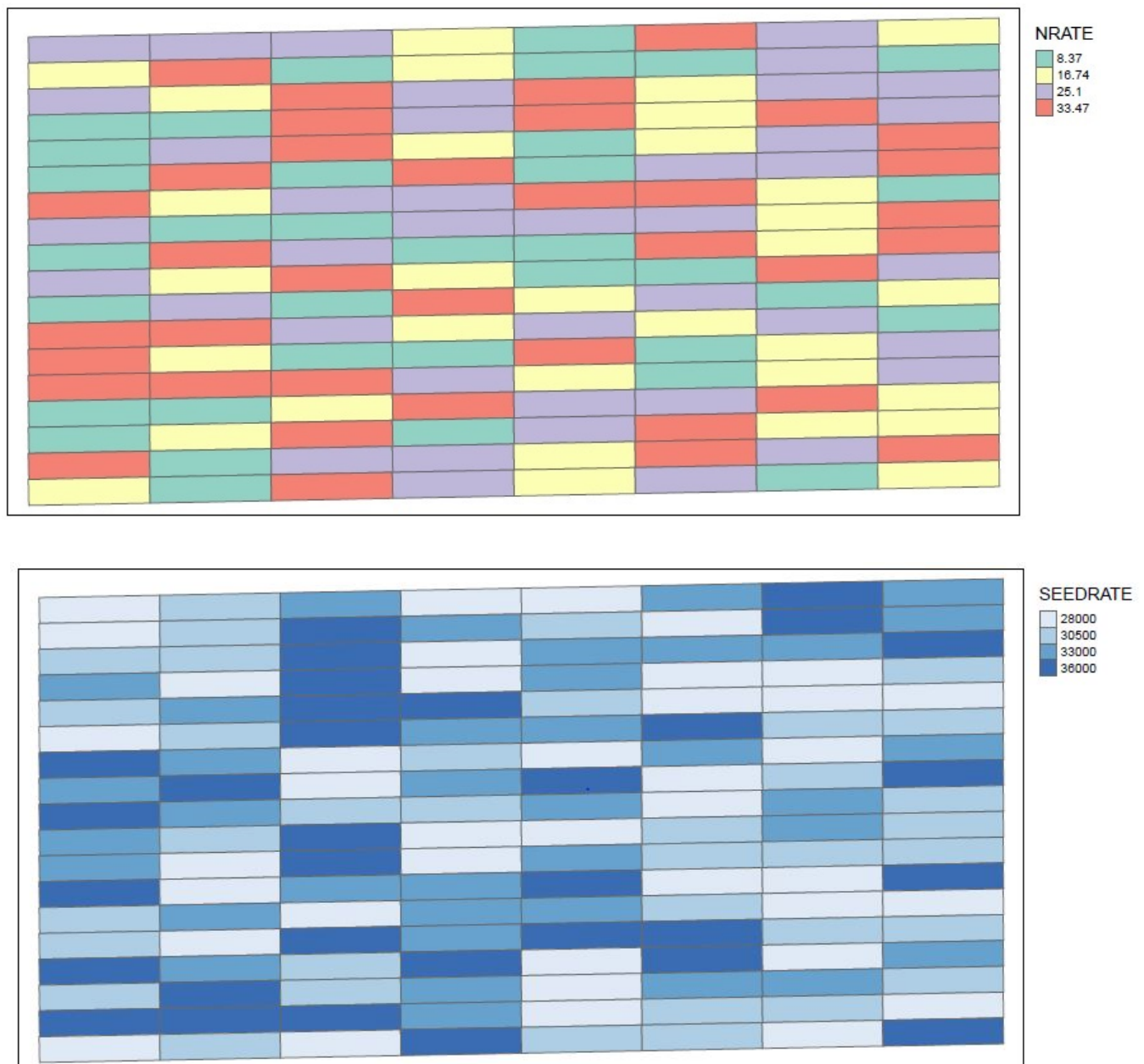
vary quite a lot within an experimental unit. Method 1 requires that all the electrical conductivity values within an experimental unit to be averaged, masking the potentially heterogeneous impact of nitrogen on crop yield depending on the level of EC. On the other hand, Method 2 allows researchers to elicit more granulated interactive impacts of nitrogen and EC because EC values are allowed to take different values within an experimental plot due to aggregation at a finer spatial resolution. In Method 3, each of the yield data points is matched with nearby EC values (there are different ways of matching). Therefore, Method 3 discards the least amount of information to statistically identify the interactive impacts of nitrogen and EC. Given these aforementioned trade-offs, it is an important empirical question as to how the data aggregation method affects the final outcome.

## Method:

For each type of data sets created using the data aggregation methods mentioned above, we will run Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis, which is a statistical method that allows one to identify factors that do not contribute to explaining yield variations. Factors included as explanatory variables are seed rate (seed), Nitrogen rate (NH3), soil electrical conductivity (EC), seed rate square (seedsq), Nitrogen rate squared (NH3sq), EC interacted with Nitrogen rate (ECNH3), and EC interacted with seed rate (ECseed). The variables of particular interest are EC interacted with Nitrogen rate (ECNH3) and EC interacted with seed rate (ECseed). This is because if they are statistically significant (LASSO chose to keep those factors in the model) that means site-specific nitrogen or seed rates should be adjusted based on the value of EC. On the other hand, if they are left out of the model, that would mean that no site-specific nitrogen or seed rates application is necessary.

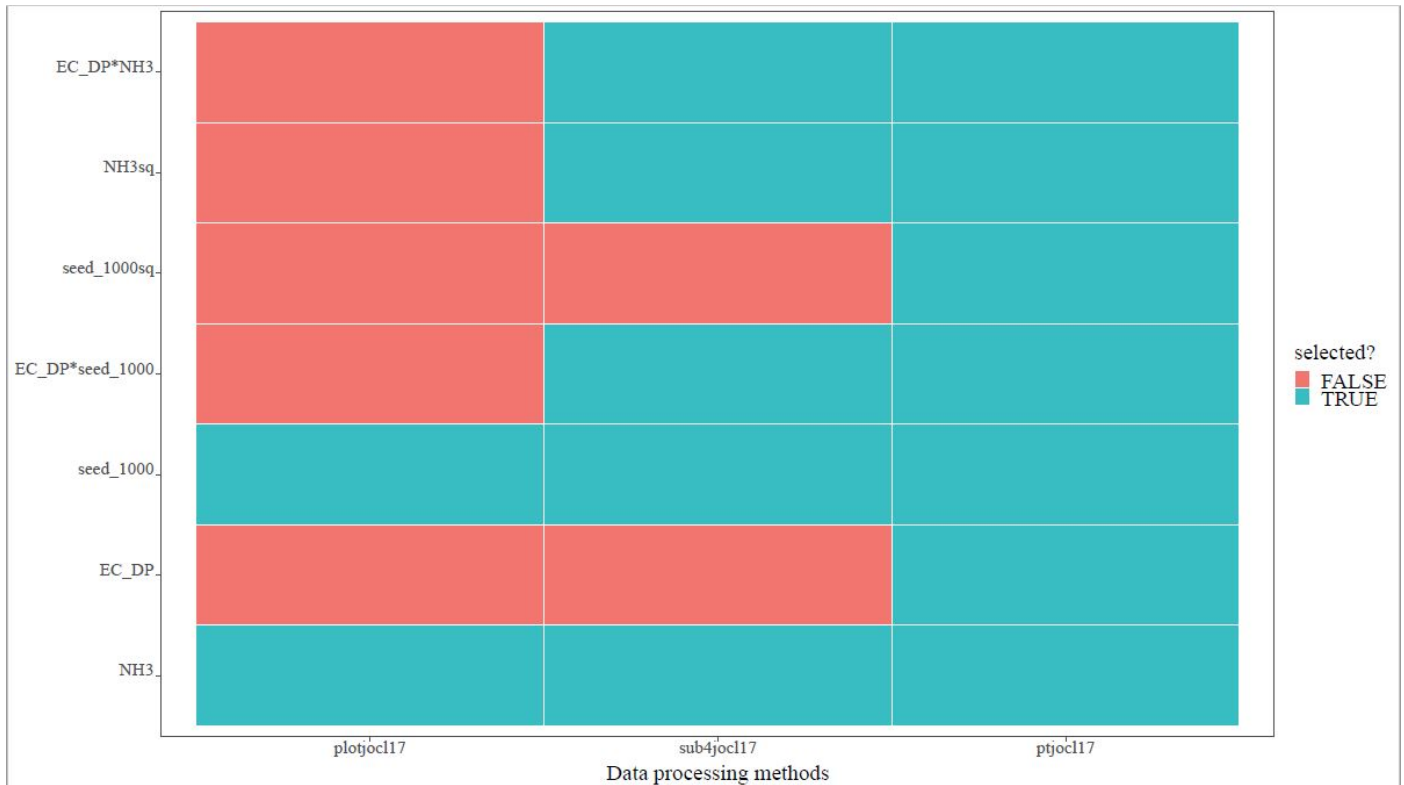
## Data:

Here, we use data obtained from nitrogen and seed experiments for corn production run in 2017 on a 70-acre field in Hamilton county, Nebraska. Figure 2 below shows the experimental design with each plot spanning 280 feet  $\times$  60 feet. The target Nitrogen rates were 8.37, 16.74, 25.10, and 33.47 gallons per acre. The target seed rates were 28000, 30500, 33000, and 36000 seeds per acre. For this field, data on yield, as-applied nitrogen and seed rates, soil electrical conductivity were collected.



## Results

Figure 3 shows the results of statistical analysis on what factors matter in explaining yield variation depending on the way data is processed and analyzed.



Red indicates that the factor was excluded from the statistical model because it is considered irrelevant, while blue indicates it is important to keep it in the model. For example, none of the variables relating to EC deep (EC\_DP) are kept in the model when the plot level data is used to estimate the yield function. As can be seen in the figure, the results from 2017 suggested different data processing methods had strong effects on variable selection results and functional forms. One of the most important findings here is the difference in whether EC deep should be kept in the model or not varies, depending on the way data is processed and analyzed. Point-wise data suggests that one should simply do uniform nitrogen and seed rates application ignoring EC deep, while point-wise data suggests that one should consider site-specific nitrogen and seed rates where the rates are varied based on EC deep. This illustrates how sensitive the final recommendation about nitrogen and seed application rates is to the way we process and analyze data. Unfortunately, we are far from understanding which data processing methods works the best. More research needs to be done on this front. This is an important topic for practitioners (farmers and consultants) because they may have the wrong conclusions about how they should be managing their input use. It is entirely possible that uniform rates are wrongly considered better compared to site-specific rates, and vice-versa.

## Summary

Both practitioners and researchers do not have a consensus on how to define the statistical analysis unit after on-farm experiments are conducted. We need to be aware of the sensitivity of final input use recommendation as demonstrated here. More research is needed to understand what data processing methods work better than the others. Answers to the question can vary context by context.

Taro Mieno  
 Assistant Professor  
 Department of Agricultural Economics  
 402-472-4134  
 tmieno2@unl.edu

Joe Luck  
 Associate Professor  
 Biological Systems Engineering  
 402-472-1488  
 jluck2@unl.edu

Zhengzheng Gao  
 Graduate Student  
 Department of Agricultural Economics