The Roles of Participating Farmers in the Data-Intensive Farm Management Project: Basic Methods and Basic Concepts*

Using GPS-guided precision agriculture technology, researchers and farmers working together on the Data-Intensive Farm Management project (DIFM) are conducting large-scale, on-farm "checkerboard" field trials, beginning the generation of vast amounts of data on how crop yields respond to input application rates, field characteristics, and weather. The goal of DIFM is to revolutionize farm management, working with farmers to implement scientific experiments on their own farms, enabling them to increase profits by making data-driven management decisions.

This document's purpose is to present to farmers an overview of DIFM and the role they play in it. So as not to waste anyone's time, in the first section we discuss the project's basic methods, focusing on the practical aspects of farmer participation in large-scale, on-farm field trials. We also explain how DIFM field trials differ from, and are radically more informative than previous field trial methods. In the second section we present a more conceptual discussion of DIFM (for those interested), examining how data from field trials can be used to increase farm profits.

DIFM's Basic Methodology: A Radical New Way to Conduct Field Trials

Farmers participating in the DIFM project use precision agriculture technology to conduct large scale, on-farm "checkerboard" field trials. Figure 1 is a depiction of a checkerboard trial run on an 80-acre central Illinois field in 2016. The trial is generating data for the examination of how

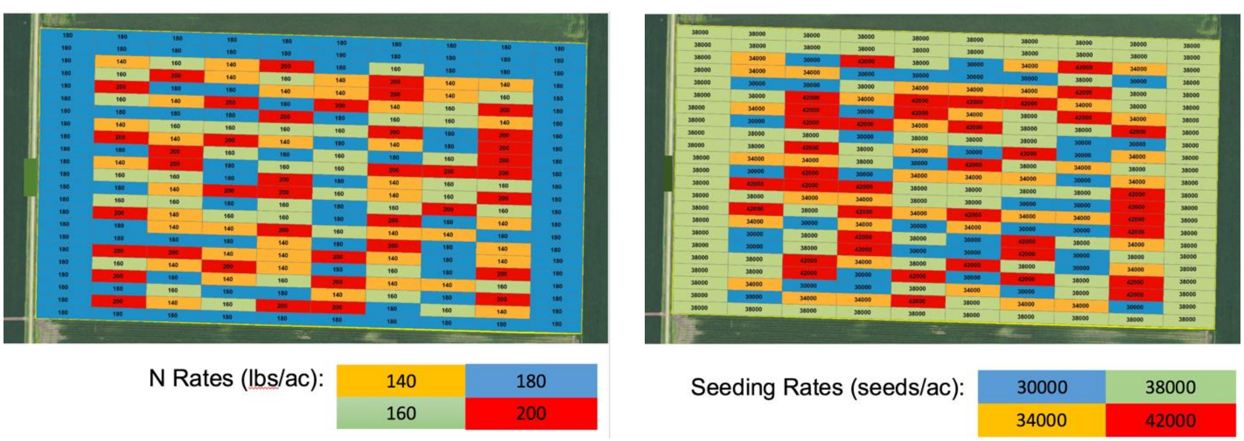

Figure 1. A 2016 Illinois "checkerboard" field trial

* A four-year research project funded by USDA NIFA-AFRI Food Security Program grant 2016-68004-24769. Participating Universities: University of Illinois, University of Nebraska-Lincoln, University of Kentucky, University of Massachusetts-Amherst, University of Maryland, Illinois State University, and University of Puerto Rico-Mayagüez

Figure 2. A farmer putting a seeding rate trial "in the ground"

corn yields respond to nitrogen rates, seed rates, field characteristics, and weather. The field was divided into a grid of 160 experimental plots. One of four seeding rates and one of four N fertilizer rates were randomly assigned to each plot. Researchers use DIFM specialized software to program "instructions" to a variable rate planter and fertilizer applicator to apply those assigned rates in the proper plots as the participating farmer *simply drove through the field in the usual manner*. (See figure 2.) At harvest, the DIFM software will work with a GPS-linked yield monitor to record the yield on each plot.

Coordinating the Design of Multiple Trials and Measuring Field Characteristics to Combine Data from Many Fields over Many Years

The DIFM field trial depicted in figure 1 is part of the beginning of a four-year effort that will design in coordination one hundred large-scale checkerboard trials, in Illinois, Nebraska, Kentucky, Argentina, and Uruguay, over four years. DIFM researchers consider these one hundred trials as only the beginning of a much larger, much longer-term project, and they believe that over this long run, the research could radically increase the amount of information available to inform farm management decisions. That the experiments are coordinated is crucial; for statistical analysis, more data is almost always good. As will be discussed in the following subsection, earlier field trial research has been severely limited by the inability of researchers to combine data from different studies. DIFM's designing many studies in many places over many years in a coordinated fashion will allow the data to be used together as a whole, and this will greatly increase the statistical power of the experiments' outcomes.

Figure 3. Small plot trials

Advantages of DIFM Methods over Traditional Field Trial Methods

Checkerboard field trials like the one depicted in figure 1, especially if repeated a number of years at many locations, will generate far more useful data than can be created by using either of the two field trial methods traditionally employed: small-plot field trials and larger-scale "strip trials." Crop scientists have run small-plot trials for more than a century. These

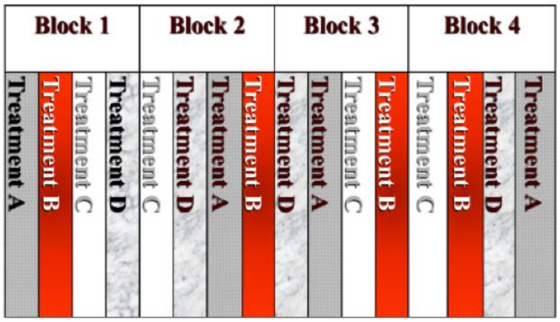

Figure 4. Strip trial

have always had to be implemented using extremely labor-intensive techniques, for example by researchers marking off plots of land using measuring tapes and orange flags, applying inputs by hand or specialized machinery at varying rates on different plots, and harvesting without the benefit of large-scale farm machinery. (See figure 3.) This labor intensity meant that it was only financially feasible to run trials on very small areas of land, often less than an acre for an entire experiment, and usually only for a few years. In addition, because DIFM trials are conducted on commercial farm fields, they directly generate specific information from those fields about those whole fields. That information is immediately pertinent to future management of those fields, whereas information from small plots is less likely to pertain to other fields in other places.

More recently, researchers have begun running larger-scale "strip trials," in which different application rates are assigned to field-length strips. (See figure 4). Such trials provide far fewer observations than do DIFM's checkerboard trials, and in addition, strip trial analysis assumes that field characteristics are uniform throughout a strip—an assumption that, given the length of most farm fields, is often strongly violated.

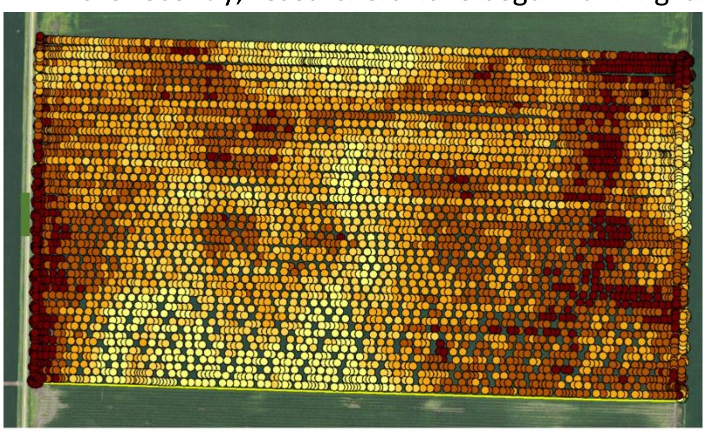

Figure 5. Electro-conductivity map on an Illinois field

Field Characterization and On-field Weather Measurement

To best use the data from multiple experiments all together, the DIFM project will *characterize* all experimental fields, and conduct statistical research to discover how field characteristics are related to optimal input management. One example of a field characterization method that DIFM will use is the measurement of a field's apparent electro-conductivity (EC). Figure 5 presents one DIFM field's EC map. Researchers have long thought that measurements of EC might be useful for input management—though *what* EC tells us about optimal management remains largely unknown. By gathering EC data on many fields and also running field trials on those fields, DIFM will be able to tell whether, and how, information from an EC map

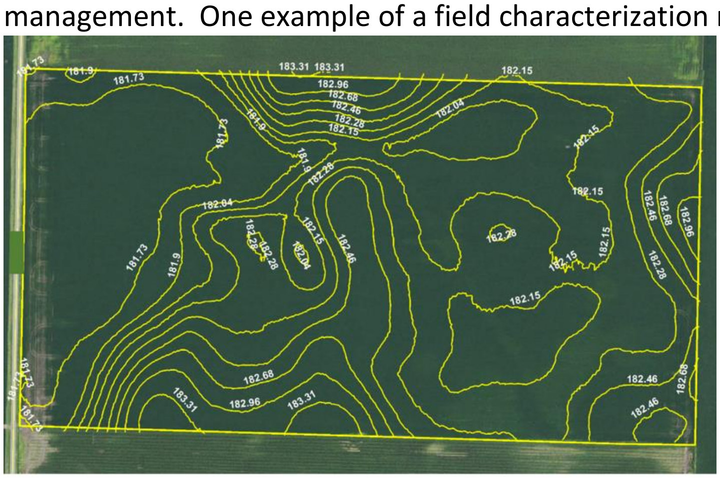

Figure 6. Digital elevation map of an Illinois field

might be used to help improve farm management. Another important field characteristic is elevation; for example, during a dry year, yields may respond very differently to fertilizer at the top of a field's hills than at the bottom of its valleys. DIFM will create a digital elevation map for every field in the experiments; figure 6 depicts one such map.

In addition to measuring field characteristics, DIFM will measure weather, on each experimental field by installing equipment that measures and records precipitation and temperature many times per day. Naturally, changing weather can radically change how yields respond to management, and accounting for weather in the statistical analysis has to be of highest priority to the DIFM project.

The Role and Responsibilities of DIFM's Participating Farmers

As mentioned above, DIFM field trials are highly computerized and automated, specialized software "instructs" variable rate equipment to work with GPS technologies to implement the experiment while the farmer simply drives through the field in the usual manner. Figure 2 shows a farmer actually putting a seeding rate checkerboard trial "into the ground." The beauty of the DIFM method is that it can generate huge amounts of pertinent field trial data on a farmer's actual fields, but at minimum bother to the farmer. Still, of course, participating farmers play an active role in the research, and of course take on certain responsibilities while participating. DIFM researches request that in February of their first year of participation, farmers attend an organizational meeting in which their roles in the project are presented and discussed. The basic contents of that discussion are that,

<u>Prior to planting</u>, the farmer

- dedicates a field of at least 80 acres to the project;
- provides a shape file showing the field's boundary;
- provides the latest soil fertility information available for the field (best as a shape file showing actual sample points and laboratory analysis.)
- provides an A-B line for planting, fertilization, and harvest;
- provides a yield map from at least one year (and preferably two) of the crop species to be planted.
- provides the width and number of rows, where applicable, of all equipment; and
- provides information about the fertilization and seeding rates that would be used on that field if the farmer were not participating in the experiments.

<u>After planting</u>, the farmer provides an as-applied map (as a shape file) for both seeding and fertilization. This tells us what was actually applied to the field. This confirms that the experiment was correctly implemented.

<u>After harvest</u>, the farmer provides a yield map as a shape field on a one-second interval for the entire field.

Any questions about the responsibilities of participating farmers can be directed to University of Illinois crop scientist, Don Bullock (dbullock@illinois.edu).

Compensation to Farmers for Any Losses due to Participation

The participation agreement between a participating farmer and the DIFM project will involve a formal contract signed by both parties. DIFM will compensate farmers, to ensure that they experience no income loss from participating in the project. Since a number of randomly-placed plots in each checkerboard trial will be managed in the same way that they would have been managed had the farmer not participated, we can use those to estimate what farm income on the whole field would have been had the farmer not participated. Comparing that income to the actual income generated in the field trials, an accurate compensation estimate can be calculated. In addition, DIFM will provide each participating farmer an annual $500 "thank you" payment, just to help defray the costs of the farmer's time.

Concepts Underlying the DIFM Project

In this section, we explain how the data generated in DIFM field trials can be used to improve farm management. We will explain that the data are used to statistically estimate yield response functions, which can then be used to estimate various kinds of economically optimal input management strategies.

What's a Yield Response Function?

A *yield response function* is a mathematical description of how crop yields respond to the *things that crop yields respond to*. Simplifying just a little, we can place all factors that affect yield into three categories: *managed inputs* (such as N fertilizer application rate, planting date, seed type, and tillage practice), *field site characteristics* (such as ground slope, soil texture, clay content), and *weather* (such as rainfall in the first week of June, and average temperature in July). To shorten how we describe these complicated relationships, below we show equation (1) (which is the only equation in this document!). Essentially, equation (1) is just a sentence describing a general concept: "Yield depends in this way on managed inputs, site characteristics, and weather." The natural question that comes out of the conceptual statement in equation (1) is, "What way is *this* way?" That is, to use the concept in equation (1) to make better farming decisions, we need to move from concepts to numerical realities: *How much* do yields respond to particular changes in N application rates? *How much* does this response depend on rainfall at different times of the growing season? *How much* do these relationships change when terrain, soil organic matter, or soil clay content change? These kinds of questions are summed up by the idea of *knowing the response function f*. If we knew what *f* was, we could come up with excellent farm management advice. *The major thrust of the DIFM research project is to learn more about f, so that researchers and extension personnel can work with farmers to come up with better input management plans.*

$$(1)\ \underset{\text{Yield}}{y} = \overset{\text{depends in this way on}}{f}(\underset{\text{managed inputs}}{\mathbf{x}}, \overset{\text{site characteristics}}{\mathbf{c}}, \underset{\text{weather}}{\mathbf{z}})$$

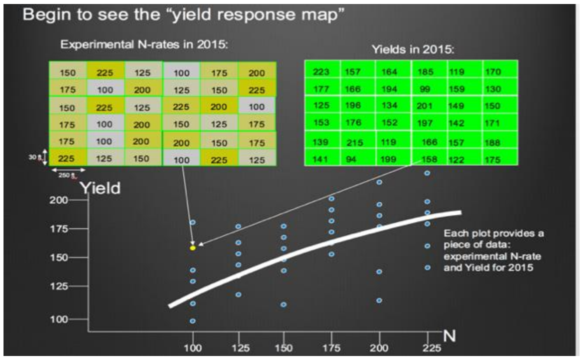

Figure 7. What the data “looks like”

How Can We Learn More about Yield Response Functions?

We can learn more about yield response functions by (a) using precision technology to generate varied input application data from large-scale, on-farm field trials, (b) combining that data with data about field characteristics and weather, (c) conducting statistical estimations of yield response functions using that data, and then (d) conducting economic analysis using those yield response functions and information about the probabilities of weather events. Figure 7 illustrates some of the basic concepts of the statistical estimation of yield response function. The figure shows a hypothetical field trial and its results. The trial depicted has 36 plots. (As seen in figure 1, In reality farm fields of the size used in the DIFM project have far more plots.) We have randomly assigned to each plot one of six experimental N application rates: 100, 125, 150, 175, 200, and 225 pounds of fertilizer per acre. At harvest, a yield monitor is used to estimate the yield on every plot. Those estimated yields are depicted in the right-hand panel of figure 1 for the hypothetical experiment. We show the 36 data “observations” in the bottom panel figure 7, where each of the 36 points represents a (N application rate, yield) result. For example, in the plot in the bottom row and fourth column form the left, an N rate of 100 lbs/acre was applied, and a yield of 158 bu/acre was recorded. This observation, (100, 158) is shown by the yellow point of the figure, along with the 35 other (NRate, Yield) observations. Using the “cloud” of data observations, it is possible to statistically “fit” a curve through those data points, which provides some information about how yield responds to N rates on the field.

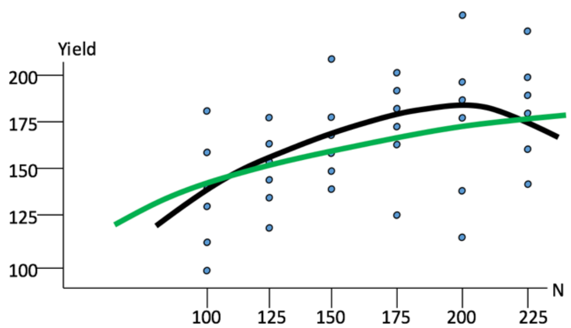

Figure 8. Because the data cloud is “spread out,” it is hard to say whether the black curve or the green curve fits the data better.

There is an important difficulty with the data analysis as so far explained however. That difficulty is that the data points in figure 7 are very “spread out.” So, statistically speaking, one cannot have much confidence in the white curve fit to the data. Given this data, it would be hard to tell, for example, whether the green curve in figure 8 is a more accurate depiction of how yields respond to the N rate “on average” on this field than is the black curve. In fact, as illustrated in figure 9, it may be that neither curve in figure 8 provides information that would

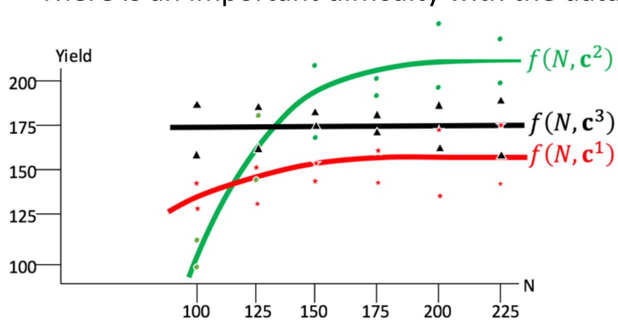

Figure 9. Taking account of the influence of site characteristics may be important for accurate estimation of yield response to N.

much improve the farmer's fertilizer management. Figure 9 shows a case in which a field has three "management zones." Here we define a management zone as an area or group of areas in a field that have the same field characteristic measurements. For example, on every location in a management zone, field characteristics such as ground slope, soil texture, soil organic matter, etc. are the same. We denote the three management zones' characteristics values as $\mathbf{c}^1$, $\mathbf{c}^2$, and $\mathbf{c}^3$. (For simplicity, the weather variables $\mathbf{z}$ are omitted from the illustration.) Figure 9 illustrates that each management zone has its own yield response curve, which are denoted $f(N, \mathbf{c}^1)$, $f(N, \mathbf{c}^2)$, and $f(N, \mathbf{c}^3)$. Note that neither of the curves in figure 8 represent well any of the three management zones' yield response curves. Thinking a minute about how a farmer might want to vary N application rates within the three management zones illustrated in figure 9 reveals three facts about optimal N management in the illustration's field: (a) In zone one, a farmer would never want to apply N fertilizer at a rate greater than 175 lbs/ac. This is because the response curve $f(\mathrm{N}, \mathbf{c}^1)$ "flattens out," with a yield of 150 bu/ac at all N rates above 175. Since fertilizer costs money, but more than 175 lbs/ac does produce more yield than does 175 lbs/ac, then the farmer would make a lower profit applying more than 175 lbs/ac than applying 175 lbs/ac. (This does not imply that the optimal rate is 175. But it does make clear that the optimal rate is not above 175. The value of the optimal rate would depend on prices, which for simplicity we do not attempt to illustrate in the figure.) Similarly, in management zone 2, the farmer would not want to apply more than 200 lbs/ac, and in management zone 3, it would be a mistake for the farmer to apply more than 100 lbs/ac.

DIFM Farmers Own Their Data

Farmers who participate in the DIFM project will own the data generated by the field trials run own their farms. DIFM researchers reserve the right to use that data in perpetuity. But DIFM personnel cannot sell the data to other parties, nor give the data to other parties without express written consent of the farmer to whom that data belongs.

For more information, contact

David Bullock
Professor and Principal Investigator of the DIFM Project
University of Illinois Dept. of Agricultural and Consumer Economics
dsbulloc@illinois.edu

Maria Boerngen
Assistant Professor and DIFM Farmer Liaison
Illinois State University Department of Agriculture
maboern@ilstu.edu

Krystal Montesdeoca
DIFM Project Coordinator
University of Illinois Dept. of Agricultural and Consumer Economics
kmontes2@illinois.edu