Harper Adams University

Proceedings of the 5th Symposium on Agri-Tech Economics for Sustainable Futures

19 – 20th September 2022, Harper Adams University,
Newport, United Kingdom.

Global Institute for Agri-Tech Economics,

Food, Land and Agribusiness Management Department,

Harper Adams University

Proceedings of the 5th Symposium on Agri-Tech Economics for Sustainable Futures

Global Institute for Agri-Tech Economics (GIATE)
Food, Land & Agribusiness Management Department
Harper Adams University
Newport, Shropshire, United Kingdom TF10 8NB

Harper Adams University

Website: https://www.harper-adams.ac.uk/research/giate/

Symposium Website: https://www.agritechecon.co.uk/

ISBN: 978-1-7398183-3-3

Edited by D. Paparas and K. Behrendt

Published by **HAU Publications (Ebooks)**

Cover Image: Hands Free Hectare, Harper Adams University

Citation: [Authors, 2022. Title.] In: D. Paparas and K. Behrendt (eds.) *Proceedings of the 5th Symposium on Agri-Tech Economics for Sustainable Futures*. Global Institute for Agri-Tech Economics, Food, Land & Agribusiness Management Department, Harper Adams University. HAU Publications, Newport, United Kingdom, 19-20 September 2022, [pp].

All Full Papers published in this volume have been subject to single-blind peer-review administered by the proceeding's editors. Reviews have been conducted by expert referees, who have been requested to provide unbiased and constructive comments aimed, whenever possible, at improving the work. The proceeding's editors have taken all reasonable steps to ensure the quality of the materials they publish and their decision to accept or reject a paper for publication has been based only on the merits of the work and its relevance to the symposium.

The Economic Performances of Different Trial Designs in On-Farm Precision Experimentation: A Monte Carlo Evaluation

Xiaofei Li[A*], Taro Mieno[B] and David S. Bullock[C]

[A] *Mississippi State University, Starkville, Mississippi, USA*

[B] *University of Nebraska-Lincoln, Lincoln, Nebraska, USA*

[C] *University of Illinois Urbana-Champaign, Urbana, Illinois, USA*

Abstract

On-farm precision experimentation (OFPE) has expanded rapidly over the past years. While the importance of efficient trial designs in OFPE has been recognized, the design efficiency has not been assessed from the economic perspective. This study reports how to use Monte Carlo simulations of corn-to-nitrogen (N) response OFPEs to compare economic performances of thirteen different OFPE trial designs. The economic performance is measured by the profit from implementing the N "prescription" (i.e., estimated site-specific economically optimal N rates) provided by analysing the OFPE data generated by a trial design. Results showed that the choice of trial design affects the final economic performance of OFPE. Overall, the best design was the Latin square design with a special pattern of limited N rate "jump" (LJ), which had the highest average profit and lowest profit variation in almost all simulation scenarios. The economic performance of the high efficiency fixed-block strip design (SF1) was only slightly lower than that of LJ, and could be a good alternative when only strip designs are available. In contrast, designs with gradual trial rate changes over space were less profitable in most situations, and should be avoided. Those results were robust to various nitrogen-to-corn price ratios, yield response estimation models, and field sizes used in the simulations. It was also found that the statistical efficiency measures of trial designs roughly explained the designs' economic performances, though there are still much part remaining unexplained.

Keywords

On-farm precision experimentation, field trial design, Monte Carlo simulation, economic performance, economically optimal nitrogen application rate.

Presenter Profile

Xiaofei Li is an assistant professor in the department of agricultural economics at the Mississippi State University, USA.

* Corresponding Author: Xiaofei Li, Mississippi State University, Starkville, Mississippi, USA 39759, email: xiaofei.li@msstate.edu

Introduction

Over the past few years agricultural scientists have been increasingly designing and implementing a revolutionary kind of agronomic field trial, generally identified as *on-farm precision experimentation* (OFPE). In OFPE, researchers and farmers collaborate to run agronomic experiments, using variable rate input technology with GPS to change input rates over multi-hectare farm fields, and using yield monitors to gather geo-spatial yield data at harvest. Because OFPE implementation is largely automated—the machine operator basically "just drives"—trial costs are dramatically lower than in traditional "small-plot" agronomic field trials (Panten, et al. 2010; Piepho et al. 2011; Bullock et al. 2019; Alesso et al. 2020; Lacoste, et al. 2022). Figure 1 presents three maps to illustrate an OFPE conducted in 2020 on a 58.8-ha DeKalb County, Illinois cornfield by the Data-Intensive Farm Management Project (Bullock, et al. 2019). The left-hand panel shows the OFPE's design, which randomized nitrogen fertilizer application rates over space. The trial design was comprised of 287 rectangular plots, each of which was 24.4 m wide (the width of the urea spreader) and between 56.9 and 73.2 m long. Each plot was assigned one of the seven experimental application rates: 83, 91, 99, 111, 119, 127, and 139 kg ha 1. The middle panel of Figure 1 shows that the "as-applied" N application rates accurately followed the design. The right-hand panel of Figure 1 shows a map of the field's measured yield values.

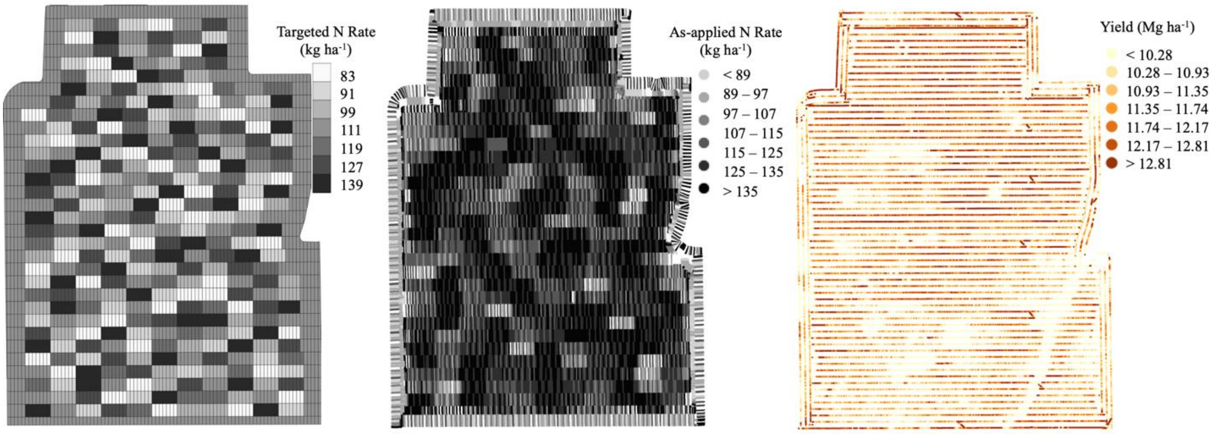


Figure 1. Trial design, as-applied map, and yield map, from an N-rate-on-corn OFPE conducted on a 58.5-ha field in DeKalb County, Illinois, 2020.

A principal aim of generating OFPE data is to empirically estimate site-specific yield-to-input-rate response functions. Knowledge of site-specific yield response functions can allow the generation of profit-increasing site-specific input application rate recommendations. Agricultural scientists have been working to understand yield response to inputs using small-plot agronomic field trial data for more than a century, and have always been concerned about optimal field trial design (e.g., Smith 1907; Spillman 1923; Eden and Fisher 1929). In fact, R.A. Fisher invented fundamental aspects of modern statistical analysis to analyse data from small-plot field trials (Fisher 1926; Box 1976, 1978). But OFPE data is different from small-plot field trial data in important ways and, much in the way that the generation of small-plot field trial data necessitated Fisher's work on efficient small-plot field trial design and analysis of the data, incoming OFPE data necessitates increased research on efficient OFPE design and statistical analysis.

Increasing Prevalence of On-Farm Precision Experiments

OFPE began near the close of the 20th Century with pioneering independent research by Cook and Bramley (1998) in Australia, by Donald Bullock and Ronald Milby in the USA (Bullock, et al., 2002; Rund, 2003; Bullock, 2021), by Doerge and Gardner (1999) in the USA, and by Lowenberg-DeBoer and Aghib (1999) in the USA. A handful of OFPEs were reported conducted in the early 2000s (Pringle et al., 2004; Panten et al., 2010; Whelan et al., 2012), but OFPE has expanded rapidly over the past five years or so. Recognizing this expansion, the International Society of Precision Agriculture organized an October 2021 conference in Montpelier, France with the theme of on-farm experimentation (International Society of Precision Agriculture, 2021). The USDA's Natural Resource Conservation Service's Conservation Innovation Grant program awarded $25 million to on-farm trials research, including a $4 million grant to the Data-Intensive Farm Management Project (Bullock, et al., 2019) which is funding researchers from fourteen US universities to run 360 OFPEs from 2021 through 2023 in thirteen US states, and to develop cyber-infrastructure to be used by commercial crop consultants and their farmer-clients to run future OFPEs and use the data for input application management (USDA-NRCS 2020). The USDA's National Institute for Food and Agriculture's National Information Management & Support System has begun funding a Multistate Research Project titled "Frontiers in On-farm Experimentation," which brings a multidisciplinary group of US-based scholars to conduct OFPEs and research about OFPEs (National Information Management and Support System, 2021).

Field Trial Design Efficiency

Research into the statistical efficiency of agronomic field trial designs has a long and prestigious history; indeed, R.E. Fisher developed much of the framework of modern statistical theory and applied experimental practices in his work in the 1920s and 1930s with agronomic small-plot field trial data from the Rothamsted Research Station (Box 1980). But the bourgeoning of whole-field OFPE is bringing new questions about trial design efficiency to the fore. While long-established concepts about how the geometric properties of field trial designs, such as "spatial balance" and "evenness" can also be used to understand OFPE design efficiency, differences between small-plot and OFPE trials in plot geometry and the spatial heterogeneity of field characteristics call for re-examination of some of the conclusions reached in the historical literature on the statistical efficiency of agronomic field trial design.

Many previous studies of the efficiency of on-farm field trials (e.g. Alesso et al, 2019, 2000) have examined the effects of trial design on the statistical accuracy (in terms of RMSE, Type I error, etc.) of yield response parameter estimators. In the present report, we instead employ an economic measure of trial design efficiency. The idea is that, a better design should generate higher quality trial data to support more accurate yield response estimations, and consequently result in economically superior input management recommendations. Using economic measures of field trial design efficiency allows us to discuss our research results in dollars and cents, terms easily understood by statisticians and non-statisticians alike.

Materials, Data and Methods

Simulated Experimental Field

Field Layout

We conducted Monte Carlo simulations of OFPEs to examine trial design efficiency. The simulations generated data on site-specific corn yield response to N fertilizer application rates on a simulated field, illustrated in Figure 2. The field was 864 meters long and 432 meters wide, covering an area of approximately 37.3 hectares, which is a typical size for row crop production in the U.S. Corn and Soy Belt. The field was assumed to be farmed in the direction of its long side, and was partitioned into a 144 × 72 grid of 6m × 6m "cells", where the field "characteristics" values were assumed spatially uniform within each cell but spatially stochastic among cells. Every trial design featured six targeted N rates. The width of the N applicator was assumed to be 18 m. As is the case currently in OFPEs, it was assumed that yield monitor technology could not accurately record large changes in yield over short distances, but rather required time and space to adjust its measurements accurately. It was assumed that when the harvester passed between plots assigned differing N rates, the yield monitor had to pass through a 12m "transition zone" before accurately measuring yield, but that thereafter could accurately record yields within 18m × 12m "subplots" made up of a 3 × 2 grid of cells. The N applicator was also assumed to require time and space to adjust the applied N rates. N trial rates were assigned to 72 m (12-cell) long "plots." The N applicator was assumed to be able to accurately apply N subplot-specifically (but not cell-specifically). Data from transition zones were not used in the statistical analyses, but each of the field's 288 N plots provided data from five subplots, meaning that useful data was generated on 1,440 subplots after excluding the transition areas. Each subplot contained six cells. Transition areas included 1728 cells, so the field contained 1,440×6 + 1728 = 10,368 cells in total.

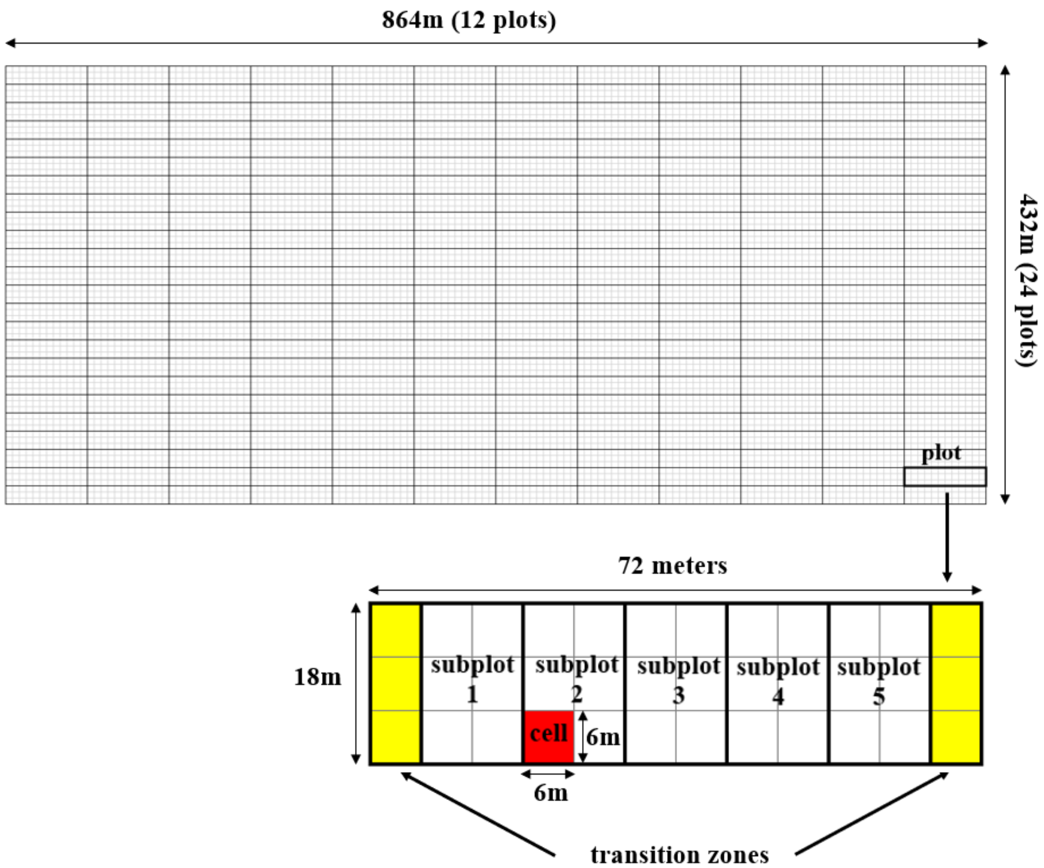


Figure 2. Experimental field layout and definition of spatial units

True Yield Response Function

We assumed that the true underlying corn yield response to nitrogen (N) fertilizer followed a quadratic-plateau functional form with an additive disturbance term:

$$f(N;\beta_0,\beta_1,\beta_2)=\begin{cases}\beta_0+\beta_1 N+\beta_2 N^2\ +\varepsilon, & N<K\\ \text{plateau}+\varepsilon, & N\geq K\end{cases} \qquad (2.1)$$

where $K=-\beta_1/(2\beta_2)$ is the critical *N* application rate above which the yield maintains a $\text{plateau}=\beta_0+\beta_1 K+\beta_2 K^2$. The quadratic-plateau form is widely used by agronomists to model corn yield response to nitrogen (e.g., Cerrato and Blackmer 1990; Bullock and Bullock 1994; Holman, et al. 2019).

Spatial Distributions of True Response Parameters ("Field Characteristics")

In each simulation, the model's "true" triplet of response parameters (β_0, β_1, β_2) varied by cell. Their spatial distributions were derived using a Gaussian random field with a variogram range of 600 m (i.e., there was no spatial correlation between parameters of cells more than 600 m apart). Figure 3 shows the maps from one simulation's response parameters. The spatial distributions of response parameters may be thought of and treated as representing the underlying spatial variability of field characteristics variables (such as soil clay content or topographical slope) that may influence yield directly or through interaction with N.

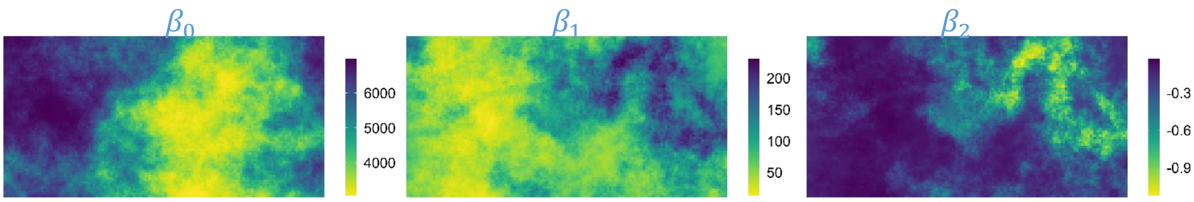


Figure 3. Spatial distribution of true yield response parameters (or "field characteristics") in one simulation

Trial Designs

Each simulated OFPE included six targeted trial N rates (N_1, N_2, N_3, N_4, N_5, N_6), which were set at the 0%, 20%, 40%, 60%, 80%, and 100% percentiles of the experimental field's true cell-level critical N rate, *K*, to ensure a range of trial rates adequate to cover most locations' yield response plateau points. Sets of targeted N rates differed only slightly across simulations, and averaged approximately 80, 128, 154, 190, 224, and 269 kg ha^{-1}.

Figure 4 displays the thirteen types of trial designs considered. Selections of types were based on two considerations. First, we included OFPE designs currently in frequent use, which are randomized strip, grid, and Latin square designs. Second, beyond the randomized designs, we constructed high-efficiency and low-efficiency fixed pattern strip, grid, and Latin square designs using four of the statistical measures developed in the agronomic literature to measure the "efficiency" of the spatial pattern of a design's trial rates distribution. Those measures are: (1) evenness of spatial distribution (Piepho et al., 2018), (2) spatial balance (van Es et al., 2007), (3) Moran's I, and (4) gradation (a measure created by the authors of this paper). We describe these measures in further detail in Appendix.

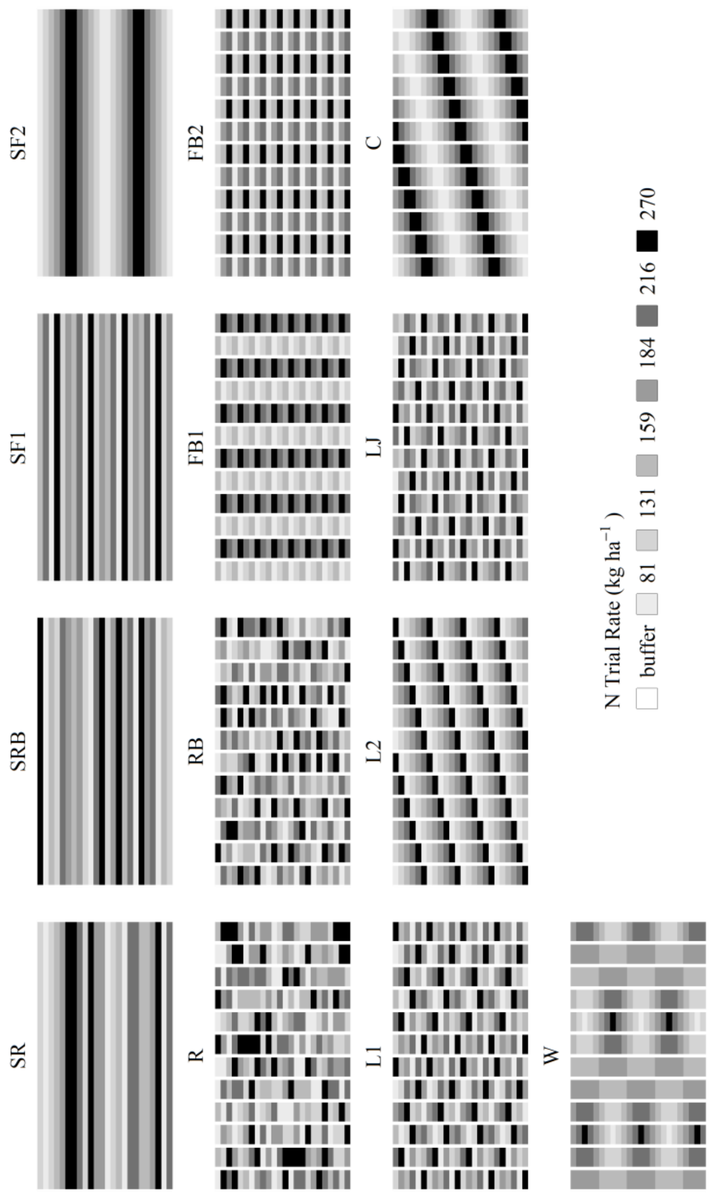


Figure 4. Trial design types

Strip Designs

Strip trials have been frequently run in on-farm research over recent decades and are commonly deployed in current research (e.g., Hicks, et al. 1997; Kyveryga, et al. 2018). Typical strip trial designs allocated targeted application rates among but not within field-length, applicator-width strips of the field. Advantages of strip trial design are that they are simple and can be implemented wihout variable-rate application equipment. In our simulations the field contained 24 strips, each 18m (three cells) wide and 864m (144 cells) long. We examined four kinds of strip design.

(1) In *completely randomized strip* designs ("**SR**") each of the six N target rates was randomly assigned without replacement to four of the field's 24 strips.
(2) In *randomized complete block strip* designs ("**SRB**") the field was partioned into four "blocks," each containing six contiguous strips, to each of which was randomly assigned a targeted N rate without replacement. Blocking is a classical design scheme in agricultural field trials, and our procedures allowed us to estimate its economic benefits.

(3) In the *high efficiency fixed block strip* design **("SF1")** the targeted N rates were spatially patternized to be (N_3, N_5, N_1, N_6, N_2, N_4) within each six-strip block, as illustrated in the upper panels of Figure 5. That patternized block was replicated for the remaining of the field (which is why we named it as "fixed block"). In total there are 720 possible blocked strip trial patterns, and the **SF1** chosen here has the highest average of the field's four statistical "efficiency" measures.

(4) In the *low efficiency fixed block strip* design **("SF2")** the strips' targeted N rates followed the patterns of (N_1, N_2, N_3, N_4, N_5, N_6) and (N_6, N_5, N_4, N_3, N_2, N_1) in alternating blocks, as illustrated in the bottom panels Figure 5. Among the 720 possible blocked strip trial patterns, **SF2** has the lowest average of the field's four statistical "efficiency" measures.

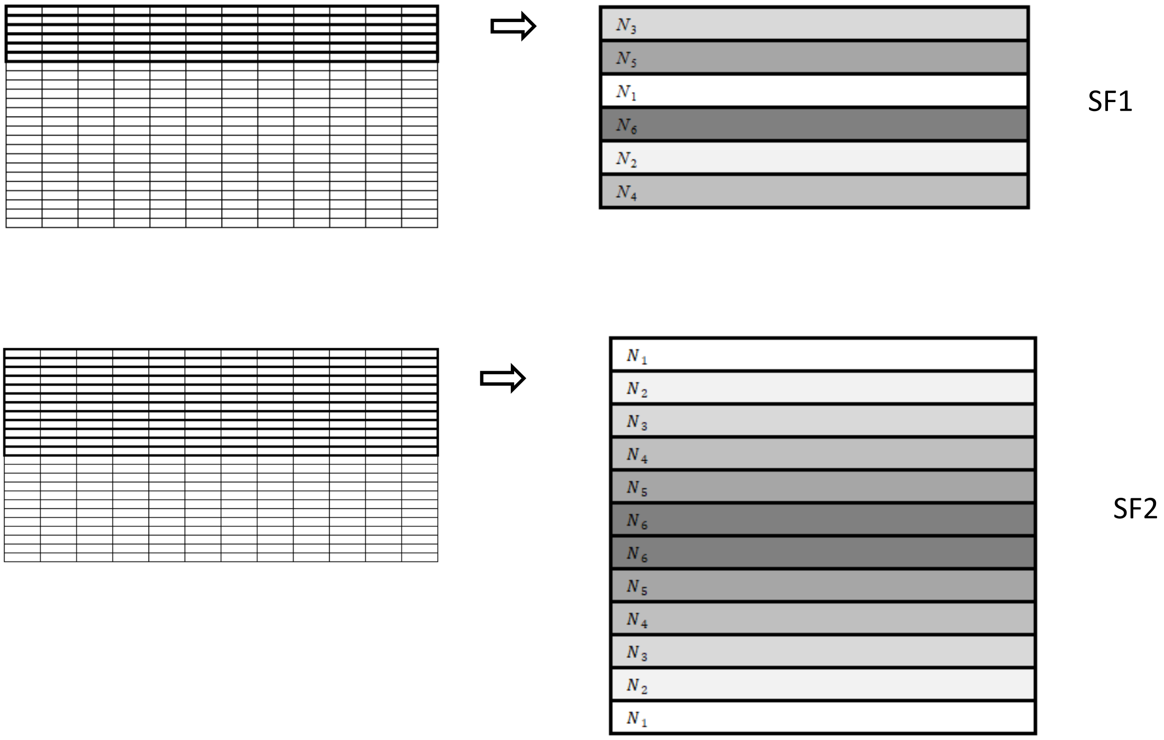


Figure 5. The SF1 and SF2 block patterns

Gridded Designs

In gridded trial simulations, targeted N rates were varied among the 72m-long, 18m-wide plots. Gridded trials can gain statistical advantage over strip trials by increasing the spatial variance of application rates. We examined two types of gridded designs, which we call non-Latin-square designs and Latin square designs.

Non-Latin-square Gridded Designs

(5) In *completely randomized gridded* designs ("**R**"), each of the six N target rates was randomly assigned to 48 of the 288 plots. Agronomists rarely employ completely randomized designs in agricultural field trials, but we used it as a benchmark against which to measure the economic benefit of blocking.

(6) In *randomized complete block gridded* designs ("**RB**") blocks comprised six plots organized in a 3-row and 2-column layout. The six N trial rates were randomly assigned without replacement to the six plots within each block. **RB**s are widely used small-plot agricultural field trials (e.g., van Es et al. 2007; Ahmad et al. 2018; Adhikari et al. 2021).

(7) In *high efficiency fixed block gridded designs*(“**FB1**”) blocks comprised six plots organized in a 3-row and 2-column layout, as in the RB design, but in each block the six N trial rates were assigned in a fixed pattern, with N_3, N_4 in the first row, N_1, N_6 in the second, and N_2, N_5 in the third, as illustrated in the left-hand panel of Figure 6. Out of the 720 possible patterns of the six N rates allocation within a block, **FB1** generates the highest average of the four statistical “efficiency” measures for the whole field’s N rate spatial layout.

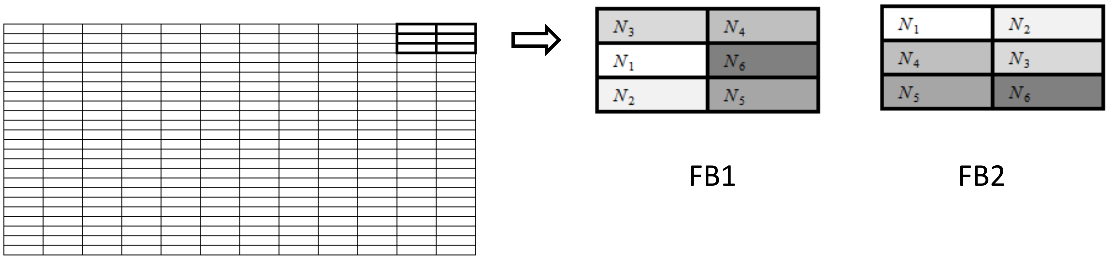


Figure 6. The FB1 and FB2 block patterns

(8) In *low efficiency fixed block gridded designs*(“**FB2**”) the within-block pattern was with N_1, N_2 in the first row, N_4, N_3 in the second, and N_5, N_6 in the third, as illustrated in the right-hand panel of Figure 6. This pattern generates the lowest average of the four statistical “efficiency” measures for the whole field’s N rate spatial layout out of the 720 possible patterns of the six N rates allocation within a block.

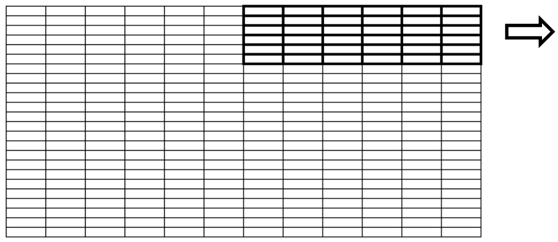
Latin Square Designs

A Latin square design with n targeted rates is defined as an array of $n \times n$ plots in which each rate is assigned exactly once in each row and each column. In our simulations, blocks were 6×6 arrays of 36 plots, the field was partioned into a 4×2 array of blocks, and the spatially arrangements of the six N rates were identical (“fixed”) among the eight blocks. In the trial design statistics literature Latin square designs have long been believed to be efficient, but are still not widely used in agronomic experiments (Fisher 1926; Box 1980; Preece 1990). We considered three specific Latin square designs.

(9) The *high-efficiency Latin square design* (“**L1**”) had the highest average efficiency ranking among all Latin square designs of order 6. Figure 7 (“L1”) displays the N rate pattern in one block of an **L1** trial.

(10) The *low-efficiency Latin square design* (“**L2**”) had the lowest average efficiency ranking among all Latin square designs of order 6. Figure 7 (“L2”) displays the N rate pattern in one block of an **L2** trial.

(11) *The rate jump constrained Latin square design* (“**LJ**”) restricted the size of changes in targeted N rates between adjacent plots within swaths. **LJ** was included in the analysis to examine the costs caused by the common limitation of variable rate input applicators being unable to make large changes in application rates over relatively short distances. We conducted the similar average “efficiency” measure ranking on all Latin squares that satisfy this “rate-jump” restriction, and used the one with the highest average efficiency ranking,. Figure 7 (“LJ”) shows the block pattern of **LJ**.



N_2	N_5	N_3	N_4	N_1	N_6
N_4	N_6	N_1	N_5	N_2	N_3
N_3	N_1	N_2	N_6	N_5	N_4
N_5	N_4	N_6	N_2	N_3	N_1
N_1	N_2	N_4	N_3	N_6	N_5
N_6	N_3	N_5	N_1	N_4	N_2

L1

N_1	N_2	N_3	N_4	N_5	N_6
N_2	N_3	N_4	N_5	N_6	N_1
N_3	N_4	N_5	N_6	N_1	N_2
N_4	N_5	N_6	N_1	N_2	N_3
N_5	N_6	N_1	N_2	N_3	N_4
N_6	N_1	N_2	N_3	N_4	N_5

L2

N_1	N_3	N_5	N_6	N_4	N_2
N_5	N_6	N_4	N_2	N_1	N_3
N_2	N_1	N_3	N_5	N_6	N_4
N_6	N_4	N_2	N_1	N_3	N_5
N_3	N_5	N_6	N_4	N_2	N_1
N_4	N_2	N_1	N_3	N_5	N_6

LJ

Figure 7. Block pattern of the Latin square designs

Other designs

We also considered two designs uncommonly used in reported research. The designs feature very gradual changes in trial rates over space. We included them for purposes of comparison with the other eleven designs.

(12) In the cascade plot design ("**C**"), the N rates changed smoothly from N_1 to N_6, and then back from N_6 to N_1, in both the row and column directions.

(13) In the wave design ("**W**"). It is even more extreme than the Cascade design, such that the N rates changed gradually row, column, and diagonal directions, This design was mentioned in Bramley et al (1999).

Yield Data Simulation Process

While N target rates were assigned by plot, as-applied N rates differed among cells within a plot because the N rate in each cell equalled the target rate plus a disturbance term. The distribution from which these disturbance terms were drawn was estimated from DIFM data. Each cell's yield value was generated in each simulation round by using the value of the cell's β parameters, the cell's assigned N application rate and a value of the spatially autocorrelated yield disturbance term ε in the yield response function defined in (2.1). ε was simulated using the Gaussian random process. The sizes of the yield errors were also calibrated to match the DIFM empirical yield disturbances. Each subplot's simulated cell-level yields were then averaged to obtain the observational unit of yield used in the analysis.

Data Analysis and Economic Evaluation

In each Monte Carlo round the subplot-level averaged simulated yield and trial N rates data were used to estimate the site-specific yield response functions. Three estimation models were used to examine the robustness of the results with respect to estimation methods: (1) the geographically weighted regression model ("GWR"), (2) the boosted regression forest model ("BRF"), and (3) the multi arm causal forest model ("MACF").. The functional form of yield response in the local regressions in the GWR models was assumed to be quadratic. GWR with nonlinear regressions is currently under development (e.g., Lambert and Cho (2022) has developed a linear-plateau GWR model), but the quadratic-plateau GWR is not yet available.

The coefficient estimation of the quadratic term in the GWR model was highly sensitive to sample errors. To alleviate this problem, the quadratic coefficient was held constant in the GWR simulations following Trevisan et al. (2021). However, GWR model only utilizes the minimum information (yield, N rates, and location coordinates) to estimate site-specific response functions, and is not guaranteed to be the most accurate modelling technique. While the better site-specific yield response models are still under development, in this study two machine learning models (BRF and MACF) with perfect field characteristics information (represented by the true response parameters (β_0, β_1, β_2)[7]) were used to mimic the more ideal modelling techniques that can possibly be achieved in the future.

Estimated Subplot-specific Yield Response Functions

For cell *i* ∈ {1, 2, …, 10368}, let $\left(\beta_0^i, \beta_1^i, \beta_2^i\right)$ denote the true value of the field characteristics vector and let $f^i(N) \equiv f\left(N; \beta_0^i, \beta_1^i, \beta_2^i\right)$ of equation (2.1) denote the cell-specific yield response function. In the OFPE practice the N trial rate and yield data are only available at subplot level. Therefore, the estimated yield response functions are subplot-specific, denoted $\hat{f}^j(N)$ for a generic subplot *j*, *j* ∈ {1, 2, …, 1440}. The GWR model generated an estimated quadratic response function for each subplot *j* as $\hat{f}^j(N) = \hat{\beta}_0^j + \hat{\beta}_1^j N + \hat{\beta}_2^j N^2$ where $\left(\hat{\beta}_0^j, \hat{\beta}_1^j, \hat{\beta}_2^j\right)$ denoting the estimated parameter values at subplot *j*. On the other hand, the machine learning models, BRF and MACF, do not require the assumption that the researcher knew the true form of the yield response function. The predicted yield and N rate relationship for each subplot *j*, $\hat{f}^j(N)$, is in a numerical manner by decision trees.

Estimated Subplot-specific Economically Optimal N Rates

In each subplot *j*, the estimated subplot-specific economically optimal nitrogen rate (EONR) was defined as,

$$\widehat{\mathrm{EONR}}^{\mathrm{j}} = \underset{\mathrm{N}}{\operatorname{argmax}} \left[\mathrm{p}\hat{\mathrm{f}}^{\mathrm{j}}(\mathrm{N}) - \mathrm{wN}\right], \qquad (2.2)$$

where *p* was the corn price and *w* was the nitrogen fertilizer price, and the derivation of $\hat{f}^j(N)$ depended on the estimation methods described above. Let R_x denote $\left\{\widehat{EONR}^1, \dots, \widehat{EONR}^{1440}\right\}$, the set of estimated subplot-specific economically optimal N application rates (the "prescription") provided by the estimation methodology used to analyse the data from the on-farm experiment.

Profits from Following the Prescription Provided by an On-farm Precision Experiment

For a generic cell *i* in subplot *j*, the true yield generated from following the R_x was calculated by substituting each subplot estimated $\widehat{\mathrm{EONR}}^{\mathrm{j}}$ into its "true" cell-specific yield response function, $\mathrm{f}^{\mathrm{i}}(\mathrm{N})$ of equation (2.1). Since EONRs were estimated subplot-specifically, each cell *i* in subplot *j* has the same estimated value of $\widehat{\mathrm{EONR}}^{\mathrm{j}}$. For notational purpose we denote the true response function $\mathrm{f}^{\mathrm{i}}(\mathrm{N})$ of cell *i* in subplot *j* as $\mathrm{f}^{\mathrm{j,i}}(\mathrm{N})$. The resulting actual per-hectare profit from applying the R_x was therefore:

[7] That is like to mimic a situation that we know the field characteristics variables that can perfectly predict the yield response parameters, and can also collect those variables data in perfect accuracy.

$$\Pi^{Rx} = \frac{1}{8640}\sum_{j=1}^{1440}\sum_{i=1}^{6}\left[pf^{j,i}(\widehat{EONR}^{j}) - w\widehat{EONR}^{j}\right], \qquad (2.3)$$

where 1440 × 6 = 8640 was the total number of cells in the subplots (which did not include cells in the transition zones) of the experimental field.

Π^{Rx} defined above is the actual profit from implementing the on-farm trial's R_x. It comes from applying the estimated subplot-specific EONRs to the field, not from applying the true subplot-specific EONRs, defined as

$$EONR^{j} = \underset{N}{argmax} \sum_{i=1}^{6}\left[pf^{j,i}(N) - wN\right], j = 1, 2, \dots, 1440. \qquad (2.4)$$

Let Π^{true} denote the profits that could be earned from the field if the producer knew every cell-specific yield response function and had the technological capability to apply N subplot-specifically:

$$\Pi^{true} = \frac{1}{8640}\sum_{j=1}^{1440}\sum_{i=1}^{6}\left[pf^{j,i}(EONR^{j}) - wEONR^{j}\right]. \qquad (2.5)$$

Let $\Delta\Pi$ denote the difference in between Π^{Rx} and the true maximum profit:

$$\Delta\Pi = \Pi^{Rx} - \Pi^{true}. \qquad (2.6)$$

Note that $\Delta\Pi$ is always negative, but when Π^{Rx} is closer to Π^{true} profits from the information garnered from the OFPE data are higher.

Three price ratios (nitrogen fertilizer price divided by corn price, both in \$/kg) were used in the simulations: 4.16 (low), 6.56 (medium), and 10.35 (high). They were obtained by taking the values at the fifth, fiftieth, and ninety-fifth percentiles of historical monthly price ratios from 1990 to 2022 (National Agricultural Statistics Service; DTN Retail Fertilizer Trends). The April 2022 corn price of \$0.28/kg was used in the simulation results, and nitrogen was assigned prices of approximately 4.16×0.28 = \$1.16/kg, 6.56×0.28 = \$1.84/kg, and 10.35×0.28 = \$2.90/kg. The discussion below is based on the simulation results when assuming the \$1.84/kg price of nitrogen fertilizer. That the N price has never actually been as high as \$2.90/kg price did not affect the EONR estimations since they were determined by the relative nitrogen-corn price ratio rather than absolute the prices. The absolute profit values Π^{Rx} and Π^{true}could be over- or under-estimated by extreme N prices, but those over- or under-estimations were linear scale-ups or scale-downs of the normal profit values and did not affect the economic performance rankings of trial designs.

Results and Discussion

Comparisons of Designs and Key Questions Addressed

Figure 8 shows boxplots of the simulated *ΔΠ* of the thirteen experimental designs from one thousand rounds of simulation. The diagram was based on a medium price ratio (N price divided by corn price, both in \$/kg) of 6.56. Profits were calculated based on three site-specific yield response models (GWR, BRF, and MACF). The values above each boxplot denote simulations' mean *ΔΠ*.

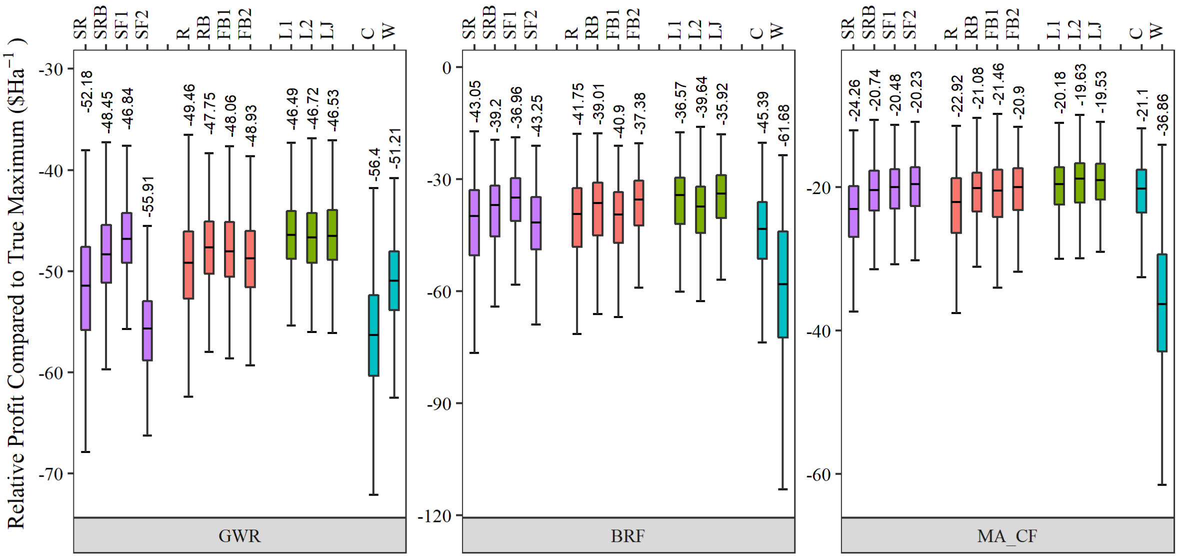


Figure 8. Boxplots of simulated profits from one thousand rounds of simulation for the thirteen experimental designs

Simulation results provide quantitative insight into several key questions related to the economic performances of trial designs. Answers are based on comparisons of profits generated from one thousand rounds of simulations of a 31.3 ha field and a price ratio of 6.56. Simulations were also run under other field size and price scenarios, and instances in which any scenario significantly affected the conclusions stated below are noted.

Is blocking economically beneficial? Yes.

Blocking designs have been commonly regarded as statistically superior ("more efficient") to completely randomized designs. Our simulation results demonstrated that blocking designs also have higher economic returns.

For strip designs, profits from blocking (that is, from using **SRB** instead of **SR**) were approximately \$3.5 to \$4/ha under all models. For the gridded designs, profits from blocking (that is, from using **RB** instead of **R**) were approximately \$1.7 to \$2.7/ha under different models. Blocking also lowered the standard deviations of the $\Delta\Pi$ estimations under all models.

(2) *Is patternizing within-block targeted application rates economically beneficial?* Sometimes, but it depends on the pattern and estimation model used.

For strip designs, the high efficiency fixed-block strip design (**SF1**) was \$0.3/ha to \$2.2/ha more profitable than the randomized block strip design (**SRB**), depending on the estimation model used. The standard deviations of profits of **SF1** were also smaller than **SRB**. But low efficiency fixed-block strip design (**SF2**) was less profitable than **SRB** under GWR (-\$7.5/ha) and BRF (-\$4/ha) models, while slightly more profitable under the MACF (\$0.5/ha) model.

For the gridded designs, the low efficiency fixed-block design (**FB2**) was slightly more profitable (\$0.2 to \$1.6/ha) than the randomized block design (**RB**) under BRF and MACF models, but the high efficiency fixed-block design (**FB1**) was slightly less profitable than **RB** under all models.

(3) *Does increasing the statistical "efficiency" of a design's spatial properties raise profits?* Sometimes, but it depends on the design type and estimation model.

The highly efficiency fixed-block strip design (**SF1**) was significantly more profitable than the low efficiency strip design (**SF2**) by as much as $9/ha under GWR model and $6/ha under BRF model, but not under the MACF model. For the gridded designs, the high efficiency fixed-block design (**FB1**) was slightly more profitable than the low efficiency design (**FB2**) under the GWR ($0.9/ha) model, but less profitable under the BRF (-$3.5/ha) and MACF (-$0.5/ha) models. For Latin square designs, the high efficiency Latin square design (**L1**) was more profitable than the low efficiency Latin square design (**L2**) under the GWR ($0.2/ha) and BRF ($3/ha) models, but less profitable under the MACF (-$0.7/ha) model.

Roughly speaking, the high efficiency designs were more profitable than their respective counterparts under the GWR and BRF models (except for **FB1** vs. **FB2** under BRF), while less profitable under the MACF model though the magnitudes of the profit difference were small. The statistical "efficiency" measures could have been effectively used as general guidelines for trial design, though caution should be exercised on cases of exceptions.

(4) *Are gridded designs better than strip designs?* Not necessarily.

Perhaps surprisingly, the high efficiency fixed-block strip design **SF1** provided similar or even slightly higher profits than the six-rate-block gridded designs (**RB**, **FB1**, or **FB2**). **SF1** profit was $0.90/ha greater than **RB** profit under GWR, $0.06/ha greater under **GWR2**GWR2, and $0.40/ha greater than **FB2** under BRF. Under MACF, **SF2** provided the highest profit, which was about $0.70/ha higher than from **FB2**. The standard deviation of profits from **SF1** were also smaller than or very close to those from **RB**, **FB1**, and **FB2**.

For strip vs Latin square gridded designs, however, the Latin square designs can be slightly more profitable. The high efficiency Latin square design (**L1**) was more profitable than the high efficiency strip design (**SF1**) for all estimation models. But the profit difference was not sizable.

(5) *Are the 6-by-6 Latin square designs better than the 6-rate blocking designs?* Yes.

The high efficiency Latin square design (**L1**) was more profitable than the six-rate-block design (**RB**, **FB1**, or **FB2**) under all models. But the profit difference was not sizable only about $7/ha under GWR, and was below $0.80/ha under the other models.

(6) *Does the inability of the machinery to change rates abruptly significantly lower the value of information from the experiments*? No.

Simulation results showed there was essentially no profit penalty from restricting the N rate "jumps" between adjacent plots in the Latin square designs. In fact, the Latin square design with constrained "jump" (**LJ**) was even slightly more profitable, and also more stable, than the best Latin square design (**L1**). The profit difference was quite small, though, at $0.04/ha from GWR, $0.35/ha from BRF, and $0.10/ha from MACF.

(7) *How did Cascade and Wave designs perform?* Poorly.

Profits from the Cascade and Wave designs were almost always the lowest among all designs. Based on GWR, Cascade design (C) profit was about $10/ha lower than the best-performing **LJ** design. Based on BRF and MACF, Wave design (W) profit was $26/ha and $17/ha lower than **LJ** design.

Overall ranking of designs and sensitivity analysis

Sensitivity to price ratios

Results above were generally robust with respect to the price ratio. Boxplot figures for profits under low and high price ratios are shown in Figures A.1 and A.2 of the Appendix. Of course, the absolute size of profit levels varied substantially with the price ratios, but the relative performances of the designs changed little.

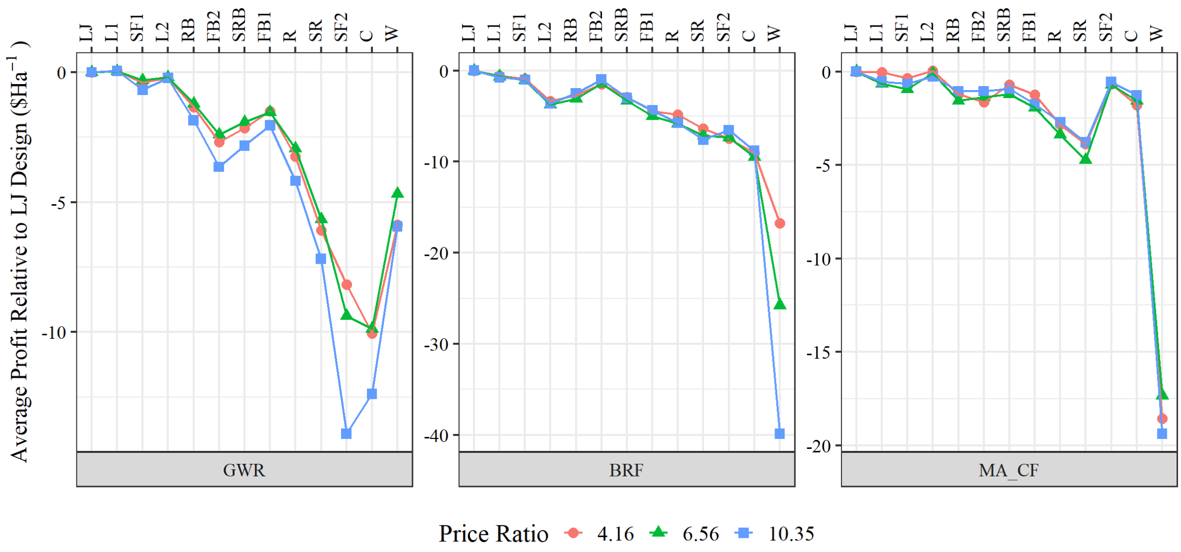


Figure 9. Average simulated profits from 1,000 rounds of simulation for the thirteen experimental designs, based on price ratios and estimation models.

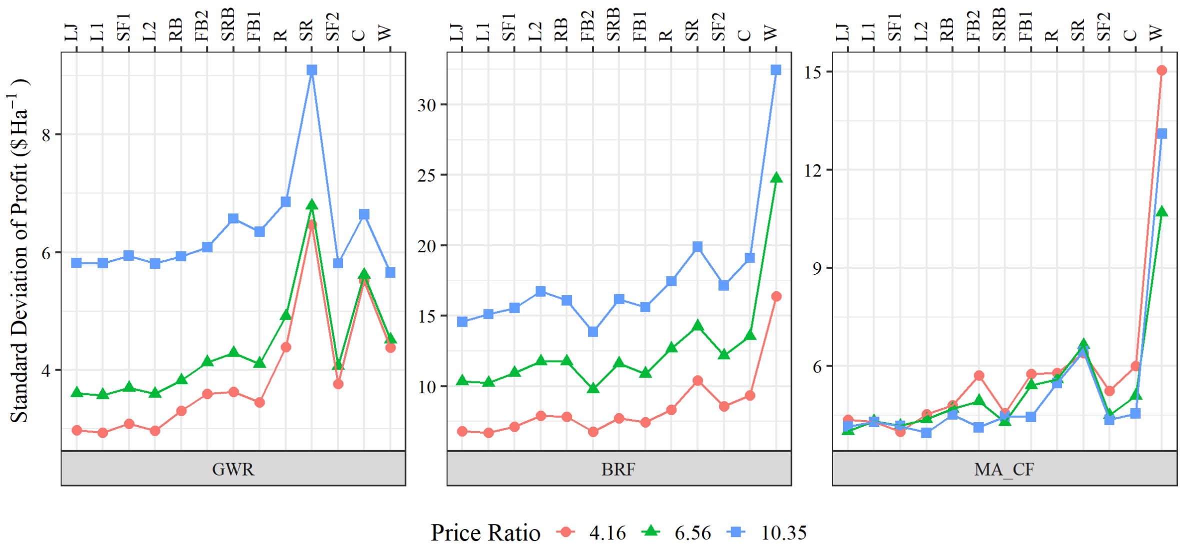


Figure 10. Standard deviation of simulated profits from 1,000 rounds of simulation for the thirteen experimental designs, based on price ratios and estimation models.

Design rankings

Figure 9 plots average $\Delta\Pi$ of each design for all estimation models and price ratios (the average profit values were extracted from Figures 8, A.1, and A.2), and uses the profit of **LJ** design as the benchmark. The values in Figure 9 shows the relative economic performance of each design compared to design **LJ**. The horizontal ranking of designs was based on the pooled average $\Delta\Pi$ over all estimation models and price ratios, which is {**LJ**, **L1**, **SF1**, **L2**, **RB**, **FB2**, **SRB**, **FB1**, **R**, **SR**, **SF2**, **C**, **W**}. Similarly, Figure 10 plots the standard deviations of $\Delta\Pi$ following the same horizontal ranking of designs.

LJ was the best design (highest average profit, and lowest standard deviation of profit) for all price ratio and model scenarios. **L1** or **L2** was in some cases a close second to **LJ**. But when taking into account the benefit from **LJ** of avoiding machinery problems by restricting abrupt N rate changes between plots, **LJ** is almost always a desirable design.

SF1 was also worth of considering given its close economic performances to **LJ** (less than $1/ha lower) but essentially no N rate changes between plots along the application direction, and therefore much lower requirements for experimental equipment.

The overall rankings of the designs were highly consistent across price ratios. The robustness of design ranking to price scenarios is especially useful as it avoids the selecting of optimal design conditional on harvesting time price that is usually difficult to predict at the time of implementing trials.

Sensitivity to estimation models

It should be noticed that the rankings of the designs varied with the estimation model used. For example, **SF2** was among the poorest performing designs under GWR, but was among the top performing designs under MACF. **RB** was more profitable than **FB2** under GWR, but less profitable under BRF, and almost identical under MACF. Other designs, such as **RB**, **FB1**, **FB2**, and **RSB**, also slightly differed in rankings across estimation models. In addition, the magnitudes of profit differences between designs were much smaller under MACF compared with other models, meaning the selection of trial design may matter less when using MACF model to derive Rx. Nonetheless, the general trends in design performance rankings were roughly similar across estimation models. Especially, the rankings of **LJ** (the top-ranked design) and **SF1** (very close to the top) were very robust to estimation models.

Sensitivity to field sizes

The design performance rankings were also highly robust to different field sizes. Figures A.3 and A.4 showed the simulation results for a 18.7 ha field, half the size of the baseline field.

The overall rankings of the design performances were similar to the baseline field results, with some slight changes. **LJ** design was still the best choice, and **SF1**’s overall ranking was even slightly better than **L1** (though the differences were quite marginal). The shapes of the trend lines of average and standard deviation of profits along the designs were still similar to the baseline field results, and therefore most of the previous conclusions hold.

Profit differences between designs were slightly larger on the smaller field for the bottom-ranked designs. For example, under the medium price ratio (6.56) the profit difference between **LJ** and **C** designs was $9.50/ha under GWR for the baseline field, and $15/ha for the smaller field. But the effects of field size on profit differences were less significant for the top-ranked designs. The standard deviations of profit were significantly larger on the smaller field than on the baseline field. Those findings may suggest the economic penalty of selecting “bad” designs increases for smaller sized fields.

Relationship between statistical efficiency measures and economic performances

We calculated the four statistical "efficiency" measures mentioned earlier (evenness of distribution, spatial balance, Moran's I, and gradation) for all thirteen designs. In addition, we tried two extra efficiency measures of design based on as-applied N rates (instead of the target N rates), which we named as "local N rate variation" and "local accidental correlation between N rate and yield error". Details of the two extra measures are described in Appendix Text A.1 "Statistical Measures of Designs".

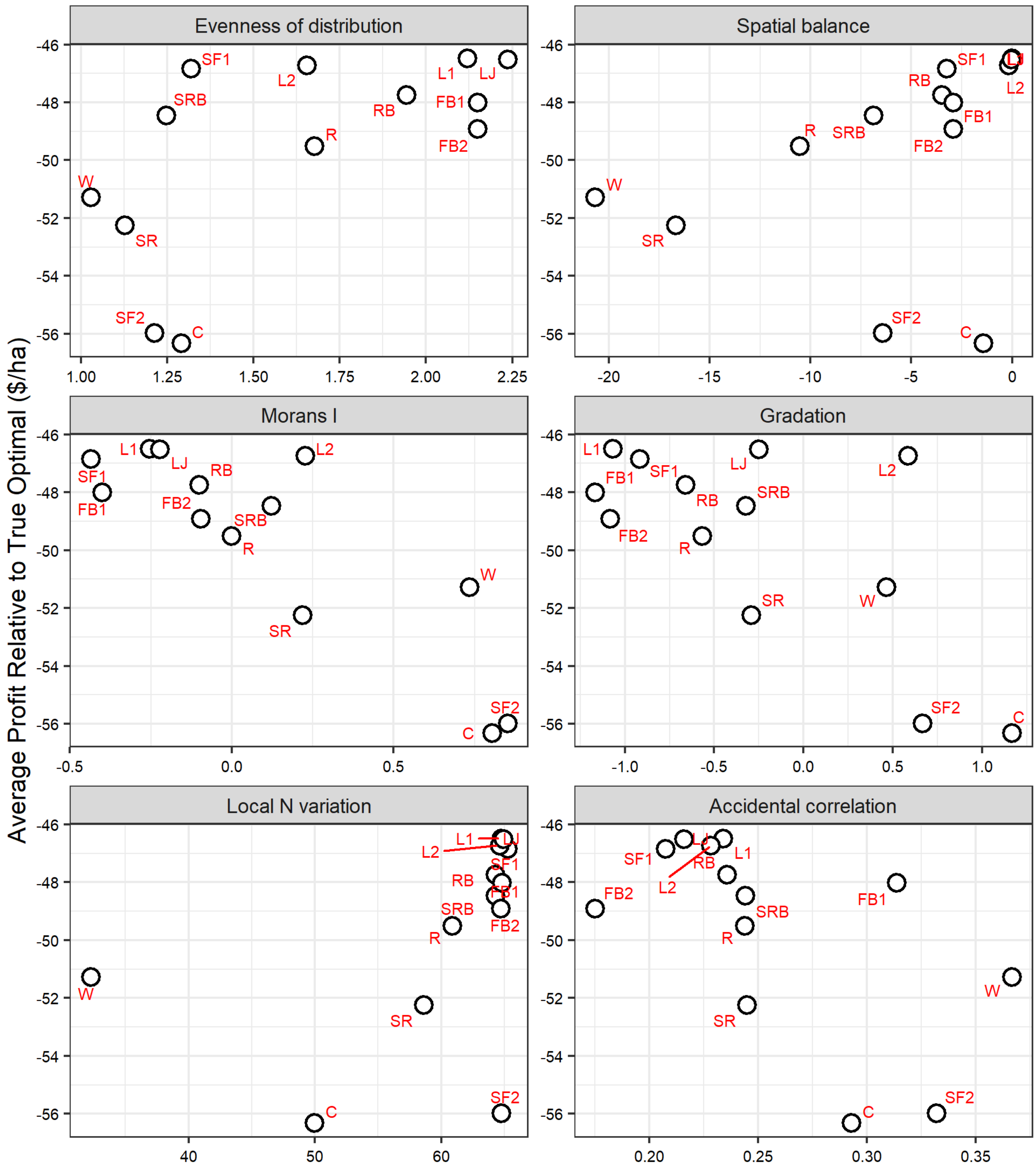


Figure 11. Scatter plots of experimental designs' statistical "efficiency" measures and average simulated profits, from 1,000 rounds of simulation, based on a 31.3 ha field, 6.56 price ratio, and GWR estimation model.

Figure 11 presents for all thirteen designs' scatter plots showing the relationships between a trial design's profits and measures of its statistical "efficiency" measures. The relationships between the statistical efficiency measures and economic performances of trial designs were roughly consistent with expectations from the literature. In general, designs with more even distributions, better spatial balance, less spatial autocorrelation (Moran's I), less regular gradation, larger local N variation, and smaller local accidental correlation tended to have higher average profits (as well as smaller standard deviations of profits, which are not shown in Figure 11). Figure 11 was based on GWR estimation model. The relationships between statistical measures and designs' economic performances were similar for BRF and MACF models as well. Details are illustrated in Appendix Figures A.5 and A.6.

But the statistical efficiency measures were only loosely related to the economic performances of trial designs, and much about these relationships remains unexplained. **LJ** was top-ranked in most measures, but ranked only in the middle for gradation. Cascade (**C**) had very good spatial balance but low profits. **SF1** had a very uneven spatial distribution but high profits. No one measure of statistical efficiency measures by itself fully explained the economic performances of the trial designs. Different measures were also conflicting with each other. The statistical efficiency measures provide some helpful insights to guide the trial design selections, but they are not sufficient enough to fully explain the designs' economic performances.

Conclusions

The first take-away from the reported research is that the choice of trial design affects the final economic performance of OFPE. Overall, the best design was the Latin square design with a special pattern to limit N rate "jump" (**LJ**). It had the highest average profit and lowest profit variation in almost all simulation scenarios. The sizes of the economic advantages of **LJ** varied. In addition, **LJ** may limit the damage to variable application equipment that can come from abrupt, large changes in application rates.

The economic costs of using strip designs instead of gridded designs may be low in some cases. The economic performance of the high efficiency fixed-block strip design (**SF1**) was comparable to that of **LJ** in various scenarios, and could be a good alternative if only strip designs are available.

Blocking raises profits. Furthermore, the fixed block designs, by properly patternizing the spatial distribution of application rates within blocks and avoiding "clumping", may work better than randomization within blocks, particularly for strip designs.

Designs with gradual trial rate changes in every direction (**L2**, **SF2**, **C**, and **W**) were less profitable in most situations. Especially, the Cascade (**C**) and Wave (**W**) designs should be avoided.

Relative design performance depended little on prices. While design profitability varied considerably across estimation models, the profitability of the **LJ** and **SF1** was consistently high across all estimation models.

Statistical efficiency measures of trial designs roughly explained the designs' economic performances. In general, more profitable designs exhibited spatially even and balanced distributions of N rates, and "fluctuated" N rate changes were more profitable than gradual N rate changes.

The conclusions above are subject to limitations that should be addressed in future research. The thirteen trial designs examined do not exhaust the list of trial designs. Only three of the many available estimation methods were examined. No attempt was made to analyse trial design performance over multiple years involving changes in weather. Only one functional form of yield response was taken into account. Finally, the field used in the simulations typified a "flat and black" central Illinois field. It is well known that spatial heterogeneity of field characteristics increases the potential profitability of site-specific input management. Future research should examine trial design profitability on fields with more spatially heterogeneous characteristic values.

References

Alesso, C. A. *et al.* (2019) 'Experimental designs and estimation methods for on-farm research: A simulation study of corn yields at field scale', *Agronomy Journal*, 111(6), pp. 2724–2735. doi: 10.2134/agronj2019.03.0142.

Alesso, C. A. *et al.* (2021) 'Design of on-farm precision experiments to estimate site-specific crop responses', *Agronomy Journal*, 113(2), pp. 1366–1380. doi: 10.1002/agj2.20572.

Bullock, D. S. *et al.* (2019) 'The data-intensive farm management project: Changing agronomic research through on-farm precision experimentation', *Agronomy Journal*, 111(6), pp. 2736–2746. doi: 10.2134/agronj2019.03.0165.

Bullock, D. S., Lowenberg-DeBoer, J. and Swinton, S. M. (2002) 'Adding value to spatially managed inputs by understanding site-specific yield response', *Agricultural Economics*, 27(3), pp. 233–245. doi: 10.1016/S0169-5150(02)00078-6.

Cerrato, M. E. and Blackmer, A. M. (1990) 'Comparison of Models for Describing; Corn Yield Response to Nitrogen Fertilizer', *Agronomy Journal*, 82(1), pp. 138–143. doi: 10.2134/agronj1990.00021962008200010030x.

Cook, S. E. and Bramley, R. G. V. (1998) 'Precision agriculture - Opportunities, benefits and pitfalls of site-specific crop management in Australia', *Australian Journal of Experimental Agriculture*, 38(7), pp. 753–763. doi: 10.1071/EA97156.

van Es, H. M. *et al.* (2007) 'Spatially-Balanced Complete Block designs for field experiments', *Geoderma*, 140(4), pp. 346–352. doi: 10.1016/j.geoderma.2007.04.017.

Lacoste, M. *et al.* (2022) 'On-Farm Experimentation to transform global agriculture', *Nature Food*, 3(1), pp. 11–18. doi: 10.1038/s43016-021-00424-4.

Lowenberg-DeBoer, J. and Aghib, A. (1999) 'Average Returns and Risk Characteristics for Site Specific P and K Management: Eastern Corn Belt On-Farm Trial Results', *Journal of Production Agriculture*, 12(2), pp. 276–282.

Panten, K. *et al.* (2010) 'Enhancing the value of field experimentation through whole-of-block designs', *Precision Agriculture*, 11(2), pp. 198–213. doi: 10.1007/s11119-009-9128-y.

Piepho, H. P. *et al.* (2011) 'Statistical aspects of on-farm experimentation', *Crop and Pasture Science*, 62(9), pp. 721–735. doi: 10.1071/CP11175.

Piepho, H. P., Michel, V. and Williams, E. (2018) 'Neighbor balance and evenness of distribution of treatment replications in row-column designs', *Biometrical Journal*, 60(6), pp. 1172–1189. doi: 10.1002/bimj.201800013.

Pringle, M. J., Cook, S. E. and McBratney, A. B. (2004) 'Field-Scale Experiments for Site-Specific Crop Management. Part I: Design Considerations', *Precision Agriculture*, 5, pp. 617–624.

Quinn, R. (2022) 'DTN Retail Fertilizer Trends', *DTN- Progressive Farmer*. Available at: https://www.dtnpf.com/agriculture/web/ag/crops/article/2022/07/06/urea-prices-drop-11-fertilizer-slip

Appendix: Statistical "Efficiency" Measures of Designs

The agronomic trial design literature stresses a number of design types with properties that tend to lend to the efficiency of estimates of yield response functions and economically optimal input application rates. We discuss three of these efficiency measures below, and also three additional measures we developed as part of the presented research.

1) Evenness of Distribution

A common opinion in the trial design literature is that a good design should have evenly distributed treatment rates over space. The left-hand panel of figure A.7 illustrates a spatially even trial design, and the right-hand panel illustrates a design in which trial rates are maximally "clumped," which is what it means for a trial design to be spatially uneven.

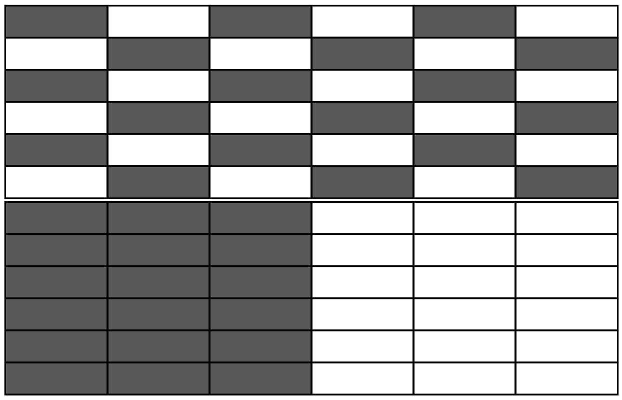


Figure A.7. Illustrations of extreme cases of spatially even and uneven trial designs.

We followed Piepho, et al. (2018) by measuring the evenness of spatial distribution by the minimum spanning tree of the Euclidean distances among plots of the same treatment.

2) Spatial Balance

Another balance measure examines the spatial distances between plots of a treatment pair. Following the definition of van Es et al. (2007), a spatially balanced design should have the distances associated with all treatment pairs as similar as possible. The distance associated with treatment pair (1, 2), for example, was calculated as the mean of distances of all possible lines connecting plots of rate 1 and plots of rate 2. The spatial balance was measured as the standard deviation of the mean distances associated with the fifteen treatment pairs of six trial rates.

3) Moran's I

Moran's I is the widely used statistic to measure data spatial autocorrelation (Moran, 1950). A high Moran's I value implies that similar treatment rates are distributed close to each other over space, which probably suggests poor evenness of distribution.

4) Gradation of N Rate Changes

Then gradation to N rate measure is original to this study. The idea is to measure whether the N rates change gradually or with wide fluctuations over space. We define a gradation index for N plot *i* as:

$$GR_i = (N_i - N_{i-1}) \times (N_{i+1} - N_i),$$

where N_{i-1} and N_{i+1} are the N rates before and after plot *i*. A positive gradation index reflects gradual changes in N rates, and a negative gradation index reflects more fluctuation in N rate changes.

5) Local Variation in N Rates

We speculated that having sufficient local spatial variation of N rates in would improve site-specific yield response estimation. Data with little local variation in N rates could reduce the accuracy or local regression estimations. In our simulated experimental fields, we define "local area" as a moving window of 6 rows and 2 columns of plots. The standard deviation of N rates within each moving window was computed, and the average standard deviation among all windows was used as the measure of local N rate variation for the field.

6) Accidental Correlation

Basic econometric theory shows that correlations between independent variable observations (N rates in this study) and the error term (yield noise in this study) biases the estimated regression coefficient. We assumed yield errors to be spatially dependent. In spatially patternized designs targeted N rates can be correlated with yield errors. The higher is the incidence of this kind of "accidental correlation", the larger will be estimation errors in the local regressions. We constructed moving windows of 6 rows and 2 columns of plots to compute the local correlation between N rates and yield errors. The average of the absolute value of the correlations across all windows was used as the measure of accidental correlation.

Figures 11 and A.5 – A.6 show correlations between the trial design statistical efficiency measures described above and the simulated profits from running the trials, analysing the data and implementing the resultant R_xs based on GWR, BRF and MACF models, respectively.

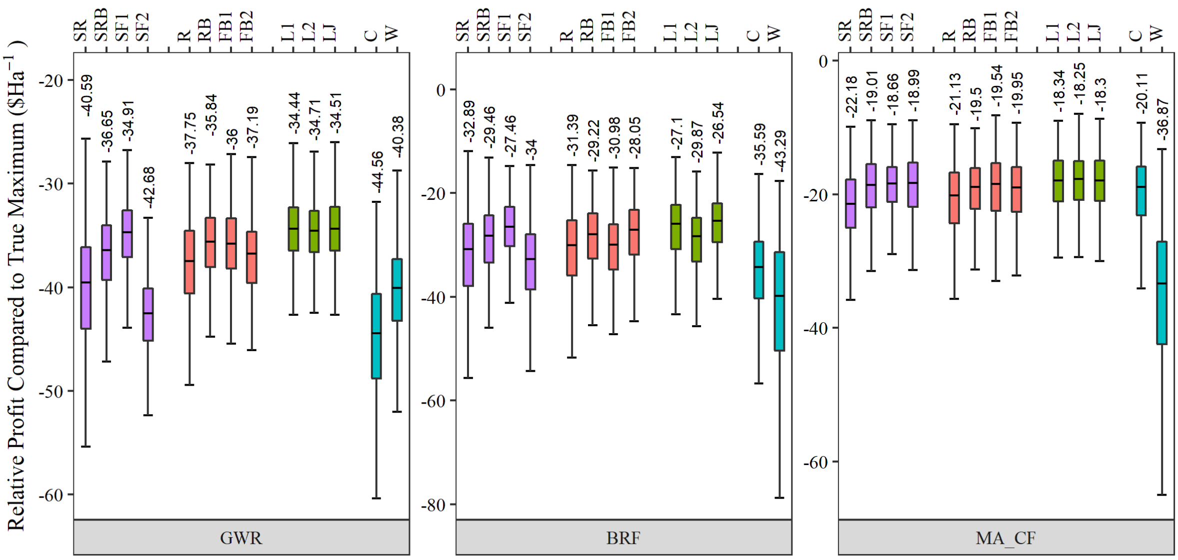


Figure A.1. Boxplots of the difference between a trial designs' profits and true maximum profits, from 1,000 rounds of simulation for each the thirteen experimental designs, based on a 4.16 price ratio.

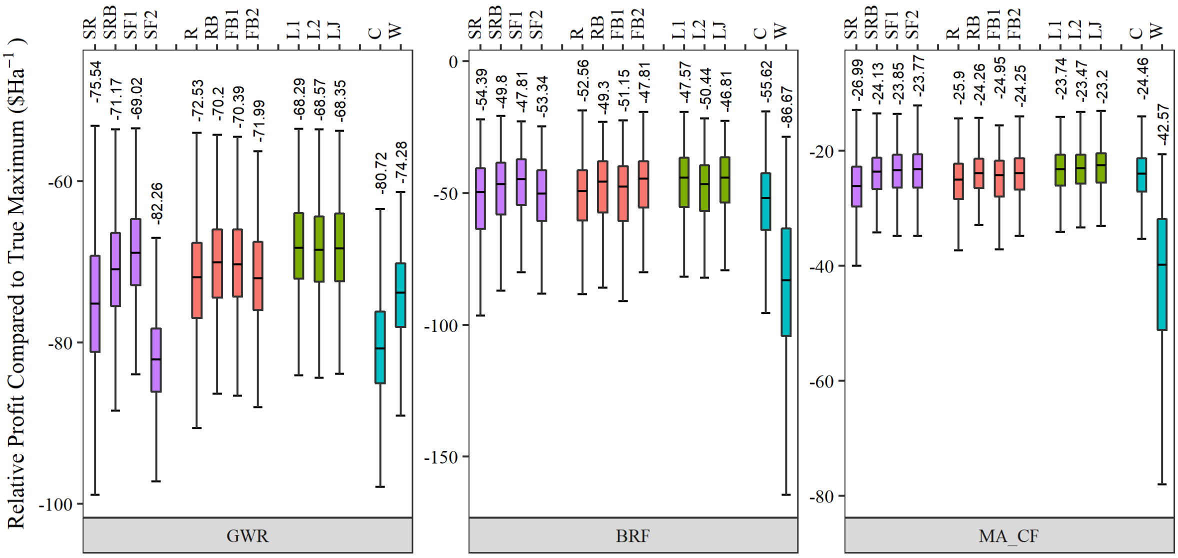


Figure A.2. Boxplots of the difference between a trial designs' profits and true maximum profits, from 1,000 rounds of simulation for each the twelve experimental designs, based on a price ratio of 10.35.

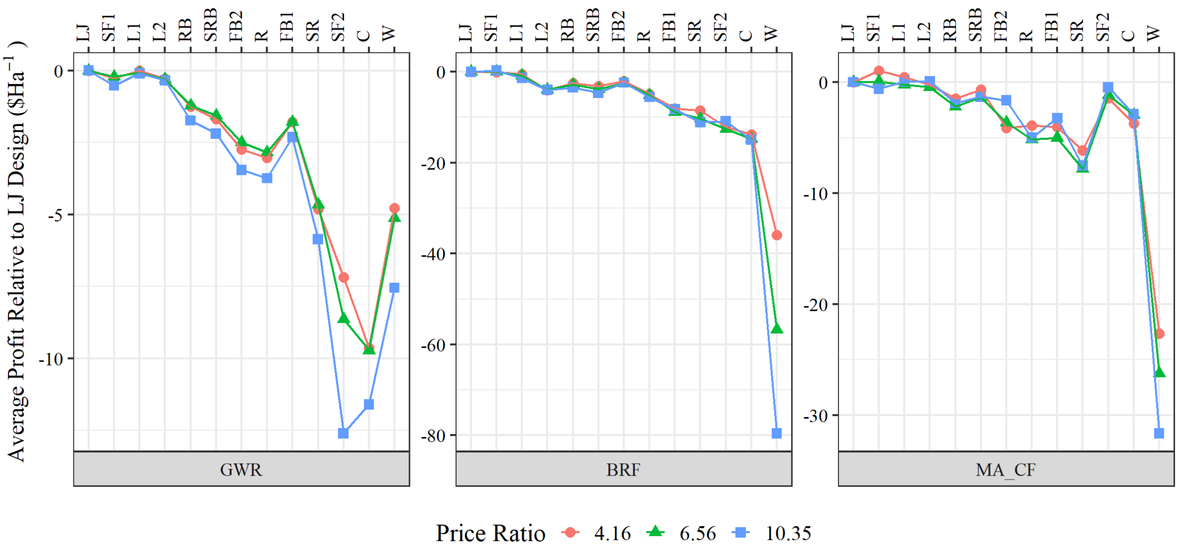


Figure A.3. Average simulated profits on the smaller experimental field, from 1,000 rounds of simulation for the twelve experimental designs, assuming price ratios of 4.16, 6.56, and 10.35. Profit values show the difference between the trail design's profits and the LJ design's profits.

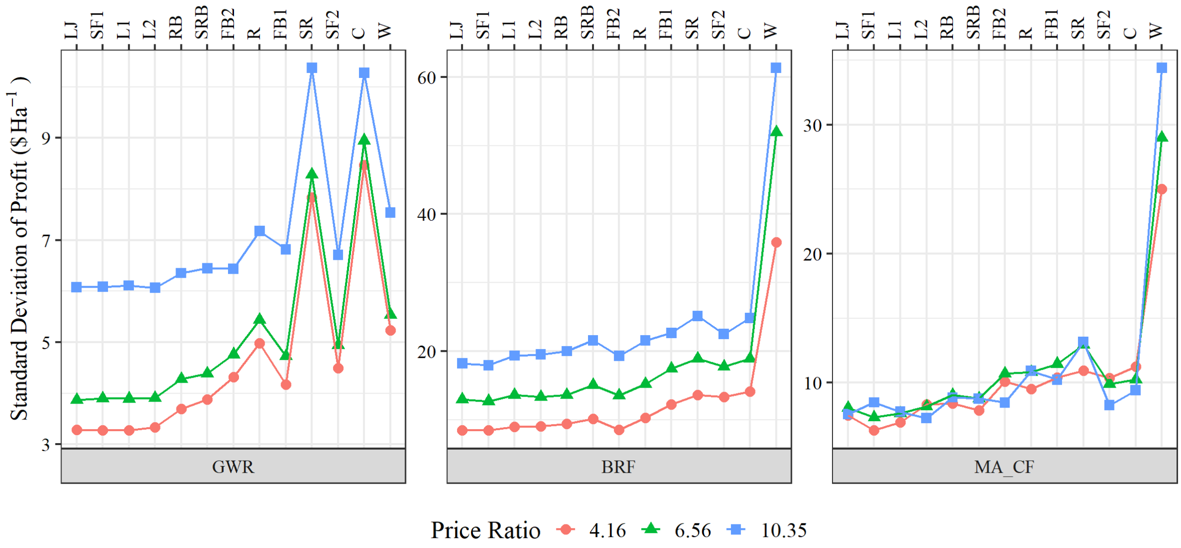


Figure A.4. Standard deviation of simulated on the smaller experimental field, from 1,000 rounds of simulation for the twelve experimental designs, assuming price ratios of 4.16, 6.56, and 10.35. Profit values show the difference between the trail design's profits and the LJ design's profits.

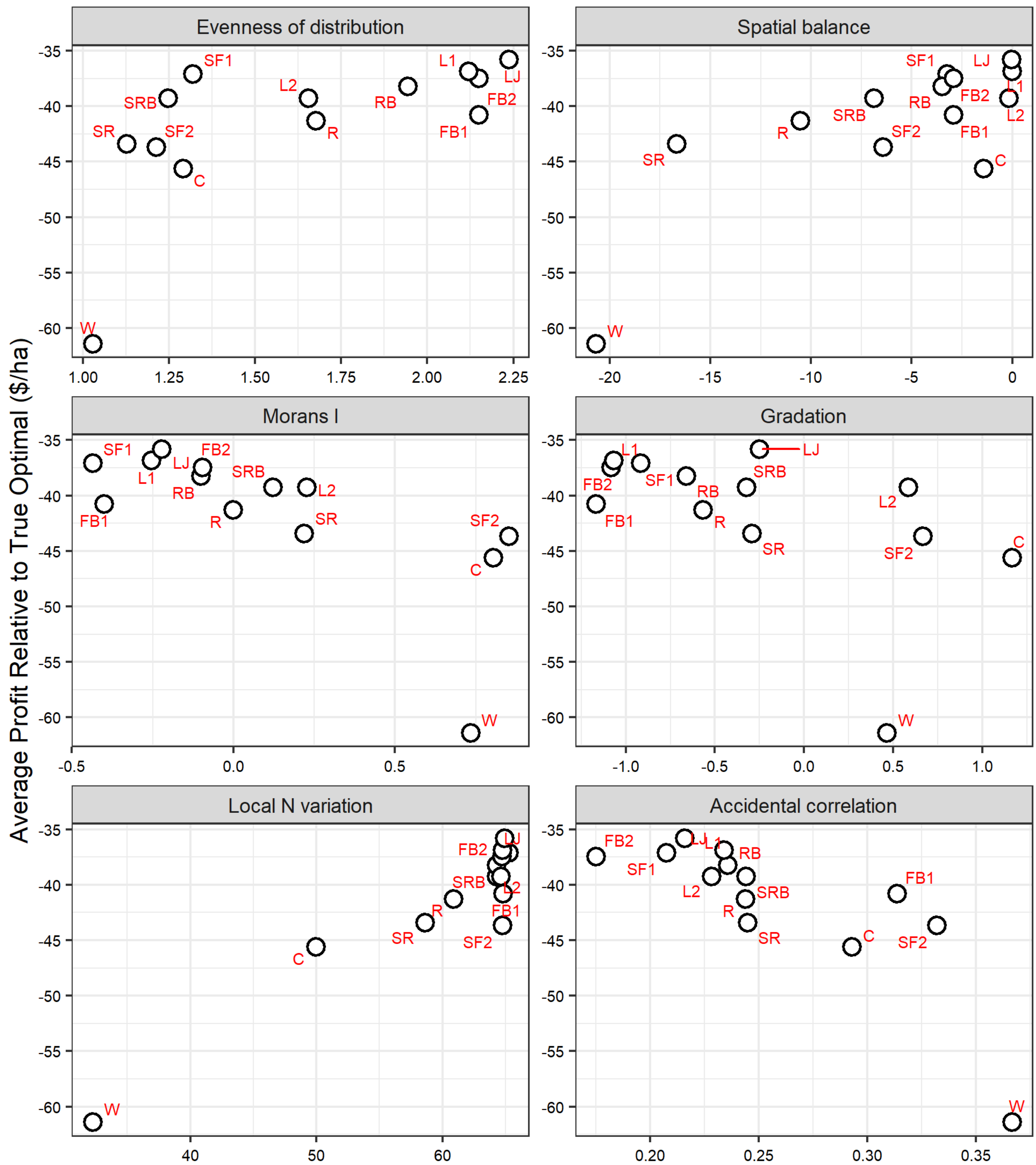


Figure A.5. Scatter plots of experimental designs' statistical "efficiency" measures and average simulated profits, from 1,000 rounds of simulation, based on a 31.3 ha field, 6.56 price ratio, and BRF estimation model.

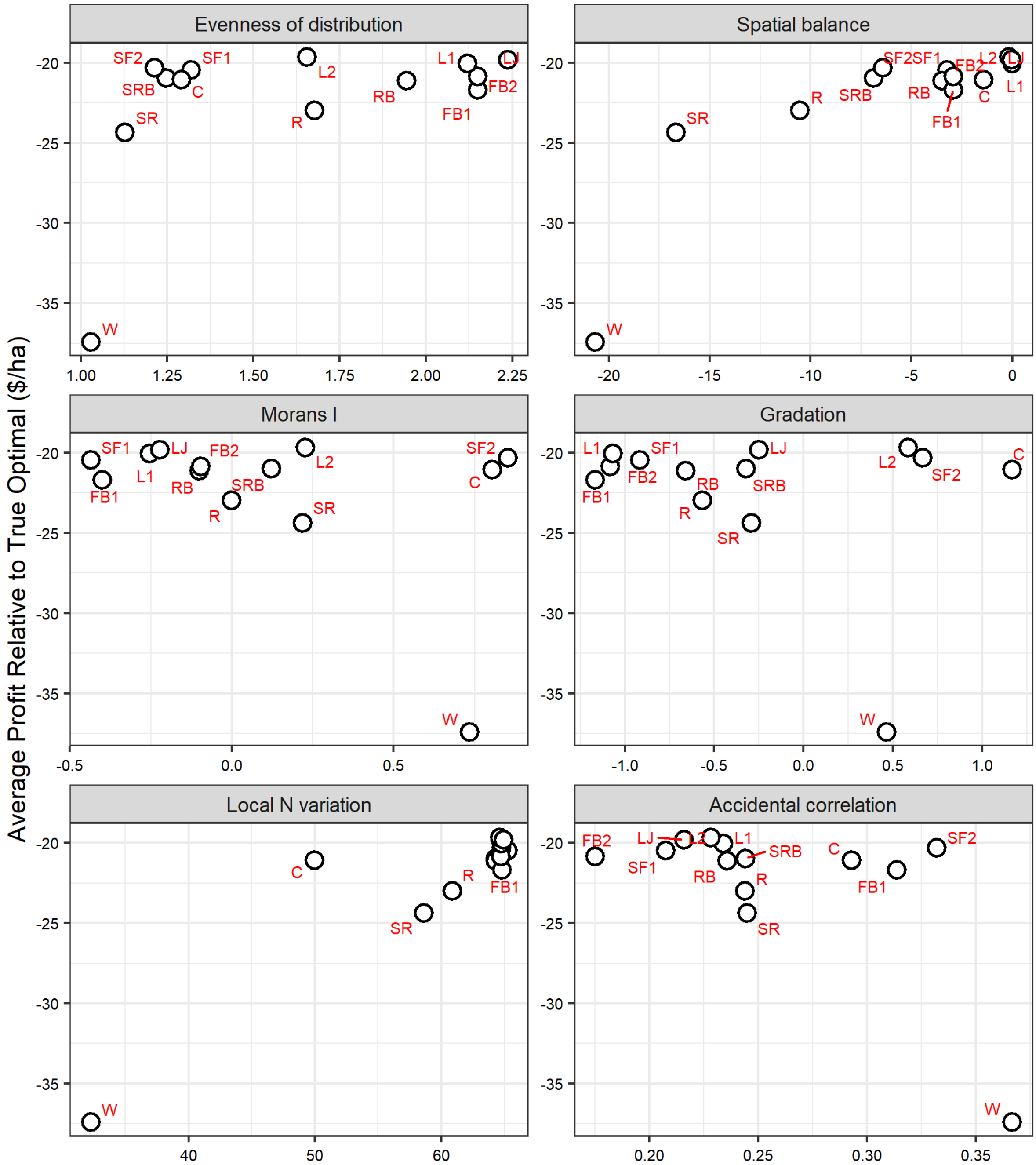


Figure A.6. Scatter plots of experimental designs' statistical "efficiency" measures and average simulated profits, from 1,000 rounds of simulation, based on a 31.3 ha field, 6.56 price ratio, and MACF estimation model.