Causal Forest Approach for Site-specific Input Management via On-farm Precision Experimentation

Shunkei Kakimoto[[1]](#footnote-1), Taro Mieno[[2]](#footnote-2), Takashi S. T. Tanaka[[3]](#footnote-3), and David S. Bullock[[4]](#footnote-4)

Estimating site-specific crop yield response to changes in input (e.g., seed, fertilizer) management is a critical step in making economically optimal site-specific input recommendations. Past studies have attempted to estimate yield response functions using various Machine Learning (ML) methods, including the Random Forest (RF), Boosted Random Forest (BRF), and Convolutional Neural Network (CNN) methods. This study proposes use of the Causal Forest (CF) model, which is one of the emerging ML methods that comprise “Causal Machine Learning.” Unlike previous prediction-oriented ML methods, CF focuses strictly on estimating heterogeneous treatment effects of inputs, rather than predicting yield. We report results of using Monte Carlo simulations assuming various production scenarios to test the effectiveness of CF in estimating site-specific economically optimal nitrogen rates (EONRs), comparing CF with the yield-prediction-oriented ML methods RF, BRF, and CNN. CF’s estimations of site-specific EONRs were superior under all scenarios considered. Furthermore, we also show that the quality of a model’s yield predictions provides little if any information about the quality of its EONR predictions.

Introduction

Despite the potential economic and environmental benefits of site-specific input (e.g., nitrogen (N), seed) management, the adoption rate of variable rate application technology remains relatively low (Lowenberg-DeBoer and Erickson, 2019). This is primarily because of the difficulty in estimating site-specific yield response functions accurately within a field, which is critical for the identification of site-specific economically optimal input rates (EOIRs) (Bullock et al., 1998).

Historically, site-specific yield response functions have been estimated using the data generated from strip trials conducted on only parts of fields (Anselin et al., 2004; Bullock et al., 2002; Hurley et al., 2004; Makowski et al., 2001; Mamo et al., 2003; Miao et al., 2006; Puntel et al., 2019; Ruffo et al., 2006; Scharf et al., 2005, 2002). But recently whole-field on-farm precision experimentation (OFPE) (Bullock et al., 2019) has allowed implementation of improved statistical experimental designs (e.g., Latin Square, Randomized Block) are used for an entire field (Laurent et al., 2019; Licht and Witt, 2019). Data generated from whole-field OFPEs typically have greater local variation in experimental input application rates and soil/field characteristics than do strip trial designs, which has improved the statistical identification of heterogeneous impacts of input management changes on yields. The resultant availability of more spatially detailed experiment data has encouraged a resurgence of interest in estimating site-specific yield response functions. For example, Barbosa et al. (2020), Krause et al. (2020) and Grant Gardner (2021) used OFPE data to estimate site-specific yield response functions and economically optimal site-specific N fertilizer and/or seed rates. These recent studies used machine learning (ML) methods extensively in place of more traditional parametric modeling approaches. Barbosa et al. (2020) applied the Convolutional Neural Network approach, Krause et al. (2020) used Random Forest-based approaches, and Grant Gardner (2021) used the shape-constrained generalized additive model. But this trend of applying ML methods may raise concerns. Historically, agronomy is a field that has paid meticulous attention to identifying the “causal” impacts of input management on yield through clever experimental designs. But the aforementioned recent applications of ML methodologies pay little attention to causal identification of inputs on yield, instead validating their models by yield prediction accuracy. But the ultimate goal of OFPE is the accurate prediction of EOIRs, not yields. These are distinct objectives, and achievement of one need not imply the other.

The lack of and need for causal interpretation of the results of ML research applications have been increasingly recognized, in particular in economics and medicine (Arti et al., 2020; Athey, 2018, 2017; Athey and Imbens, 2019; Moraffah et al., 2020; Schölkopf et al., 2021; Storm et al., 2020). In response, causal machine learning (CML) methods were recently developed. Unlike traditional prediction-oriented ML methods, CML focuses on identifying causal impacts of an event (in our context, an increase or decrease in the input application rate). The growing academic literature applying CML methods includes Richens et al. (2020), who pointed out that the inability of the traditional ML methods to identify cause and effect relationships between disease and patients’ characteristics could lead to misdiagnoses, and showed that CML-based diagnoses could improve the accuracy of diagnostic prediction. In addition, Carbo-Valverde et al. (2020) used a CML to identify characteristics causing bank customers to adopt online banking services, and Bozorgi et al. (2020) applied CML to study the mining processes.

In this paper, we test the accuracy of site-specific economically optimal N fertilization rates (EONRs) estimation based on Causal Forest (CF) approach developed by Athey and Imbens (2016) and Wager and Athey (2018). We compare the EONR prediction performances of our CF-based approach to the widely used ML methods Random Forest (RF), Boosted Random Forest (BRF), and Convolutional Neural Network (CNN). Further, we also test how yield prediction performance is related to EONR prediction performance to examine whether it is appropriate to select the model based on its ability to predict yield when the ultimate goal is to identify EOIR. We use Monte Carlo simulations, where we synthetically generate yield and soil/field characteristics datasets for a stylized “field.” Because real-world “true” EOIRs cannot be measured, real-world data cannot be used to test the accuracy of statistical methods estimating EOIRs. This important difference contrasts starkly with yield prediction, since actual yields can be ground-truthed.

Predicting Yield vs. Predicting Changes in Yield in Response to Changes in Input Management

In the following, we describe the fundamental information necessary to identify EOIR: the site-specific treatment effect of input application on yield. We then discuss the causal forest method in detail, show how it is particularly suitable to obtain this fundamental information, and contrast it to more traditional prediction-oriented ML methods, such as RF, BRF, and CNN.

A Simple Profit-maximization Problem

Let be the price that a producer receives for a unit of a crop, be the per-unit price paid for a unit of an input, represent other costs of production, and describe the relationship between crop yield, the input application rate and a vector of field characteristics variables . A simple model of producer behavior assumes that the producer chooses a level of to maximize profits (revenues minus costs):

Elementary calculus shows that the necessary condition for maximization is, that , which implicitly defines the economically optimal input application rate as a function of prices and soil/field characteristics, and may be arranged as (2):

Identity (2) states that the necessary elements of the information needed for profit maximization are the input price, the output price, and the partial derivative of the yield function with respect to the input. While knowledge of the yield function in Eq. (1)) is sufficient to solve the maximization problem above (since its derivative can be calculated), identity (2) shows that it is not a necessary condition; rather, knowing the function’s partial derivative is necessary. In the following, we discuss a method of using Causal Forest methods that can provide statistical advantages to the estimation of economically optimal input application rates because it is built to estimate approximations of the derivative directly, and not the yield function itself.

Estimating EOIR and Treatment Effects

Recent papers using ML methods to estimate the effectiveness of various agricultural input management strategies have based their analyses on estimates of a yield function, (e.g., Krause et al. (2020); Barbosa et al. (2020); Grant Gardner (2021)). In the following, we present a Causal Forest methodology that does not rely on estimation of how the level of yield depends on the input application rate , but rather relies on a discrete version of Eq. (2), examining the discrete changes in yield that result from discrete *changes* in the input application rate .

Consider an increasing ordering of input application rates . Among these K rates, for a generic site a specific EOIR can be identified if the *treatment effect* function is known. This function is defined as the discrete change in yield that results from a discrete change in the input level from to , and it depends on the characteristics at the site analyzed:

Once treatment effects have been identified for every , economic analysis can be conduced to identify the EOIR for site . The profit change from increasing the input rate from to is

The EOIR is the value that maximizes the profit gain “relative” to profits from . ( may be the EOIR if no other rates surpass the profit of .)

Causal Forest

CF was developed specifically for identifying heterogeneous treatment effects (Athey and Imbens, 2016; Wager and Athey, 2018). Like RF, CF works by building an ensemble of trees, called *causal trees* (CT), by recursively partitioning covariate space into two leaves, where each leaf estimates conditional average treatment effects as the mean differences between the dependent variable (here, yield) of the treated and control observations that belong to the leaf. Figure 1 shows an example of CT where the treatment is an increase in seed rate from seeds/ha (control) to seeds/ha (treated) with apparent electrical conductivity () and topographical as variables to explain the heterogeneity of the impact of the treatment impact. In the first node, all the samples ( observations) belong to the node, and the causal effect of the treatment was calculated as the difference in the mean yield of the treated and control groups, which turned out to be kg/ha in this example. The first data split criterion is whether the value is greater or less than . The observations with greater than or equal to move to the left node, while those with less than move to the right. For each of the subsamples, the treatment effect is identified as the difference in the mean yields of the treated and control groups within the subsample. The treatment effect for the low- observations was kg/ha, and for the high- observations was kg/ha. This means that for the impact of the increase in seed rate the treatment effect is heterogeneous over space, having a larger positive impact on yield in high- areas. Trees are further developed to have more groups of subsamples to allow for more flexible representations of the heterogeneous treatment effects. The estimated CF model does not predict yield level at any site in the field, but rather estimates site-specific EOIRs by estimating treatment effects. Since CF is designed to estimate the causal effect of an input directly, , it may be suspected that CF will outperform other ML methods that identify EOIR indirectly through yield prediction.

While both RF and CF build trees by recursively partitioning samples, they differ critically in their criteria for splitting. In RF, the sample is split to minimize the mean squared error (MSE) of yield prediction in building trees. Unfortunately, the MSE of treatment effect predictions is not useful for building a CT since the true treatment effects are never observable. However, Athey and Imbens (2016) showed that minimizing the expected MSE of the treatment effect is equivalent to maximizing the variance of treatment effects across the resulting two leaves and minimizing the within-leaf variance. Consequently, by splitting samples in a way that maximizes the variance of treatment effects across the resulting two leaves, CF overcomes the fundamental problem created by the unobservability of true treatment effects. CF requires that the tree-building process is “honest” for treatment effects estimation to be unbiased. Honest tree-building first randomly splits in two the training data samples, then uses one of the subsamples to determine how to split the tree, and uses the other to estimate the treatment effects (Athey and Imbens, 2016; Wager and Athey, 2018). This honest subsampling technique avoids overestimation of the heterogeneity of estimated treatment effects. Since CF splits samples to maximize the heterogeneity of estimated treatment effects, using the same sample for both the sample splitting and treatment effect calculations renders itself sensitive to noise or outliers, allowing high noise levels to exaggerate the treatment effects.

CF vs. Prediction-oriented ML methods

Whether the indirect (yield-prediction-first) and direct approach is used can lead to meaningful statistical differences when it comes to EOIR estimation. The conventional prediction-oriented ML methods are designed to predict yield, , well, but not necessarily treatment effects. Since the loss function used in prediction-oriented ML uses (typically squared) residuals in yield prediction, prediction-oriented ML is naturally optimized for predicting the “level” of yield instead of identifying the causal impact of the input. CF, on the other hand, is designed specifically to estimate the causal impact of input. Indeed, Wager and Athey (2018) used Monte Carlo simulations to show that CF provided more accurate estimates of heterogeneous treatment effects than those provided by RF. However, the underlying data generating process of their simulations was extremely simple, with the treatment effect influenced by only a single variable. CF has never been tested for more complicated data generating processes, such as those represented by yield response functions in which the treatment effect of an input can be affected by multiple variables in a non-linear fashion. Moreover, CF has never been compared to more advanced ML methods such as BRF and CNN that are considered to have prediction capabilities greater than RF’s. The study reported here builds on Wager and Athey (2018) to compare CF to these prediction-oriented ML methods (not just RF, but also BRF and CNN) in more realistic agronomic experiment settings.

Materials and Methods

The first step in our simulation was to generate one thousand synthetic fields that “resembled” an actual field producing corn where the input of interest is nitrogen. For each simulated field, we generated a training data set and a validation data set. Field characteristics and yield were simulated based on the Mitscherlich-Baule (MB) yield response function. We used the RF, BRF, and CNN ML methods to estimate yield response functions, and we used the CF method to estimate treatment effects. For each method, we estimated site-specific EONRs using the training data set, and checked the statistical performance of the predicted site-specific EONR using the validation data set. Detailed descriptions of these steps are provided below.

Data Generation

Simulated Experimental Fields

The simulated field was created based on an actual 51-ha Illinois field on which the Data-Intensive Farm Management (DIFM) project (Bullock et al., 2019) has run multiple OFPEs. Figure 2 illustrates that simulated trials consisted of 60-ft 240-ft “plots,” each of which was assigned an N fertilizer application rate. Each plot was made up of four (4-rows 1-column) “subplots,” which were unit of analysis used in subsequent statistical analysis. Each subplot contained of a 6-rows 6-columns grid of thirty-six 10-ft 10-ft “cells.” All field characteristics and yield data were generated at the cell level.

Underlying Yield Response Function

We modeled the yield response function as having the following Mitscherlich-Baule functional form:

The first term on the right-hand side of Eq. (3) is the deterministic yield component, which was generated assuming the commonly-used MB functional form, which is flexible in its shape, allows for the existence of a yield plateau, and has been shown to be a plausible specification for N fertilization models for corn (Frank et al., 1990; Llewelyn and Featherstone, 1997; Paris, 1992). The variable in Eq. (3) represents the nitrogen fertilizer application rate (kg/ha) which the trial designs varied by plot. Also in Eq. (3), , , and are functions of a vector of field characteristics variables , which vary spatially on the field. For example, and , which affect yield independently from the applied N fertilizer rate, might be interpreted as depending on the ingredient soil N content supplied by residual N, soil organic matter, etc. can be interpreted as reflecting how efficiently the crop uses N to create grain mass, and might depend on soil properties that affect the availability of applied N fertilizer to the crop, say due to N immobilization, leaching, denitrification, and volatilization (Alva et al., 2006; Johnson et al., 2005). The error term of the yield represents the composite of all the unobserved factors that affect crop yield beyond , , and .

A fictional “researcher” is modeled as having limited knowledge about which variables make up , or even the number of elements of , and has limited ability to measure the values that many of the variables that constitute take on across the field. The researcher is also assumed to not know the functional forms of the yield response function.

Cell-specific Yield Functions and Economically Optimal N Rates

Let index a simulated field’s subplots and index cells within a subplot. Let be the N fertilization rate on plot , and let be the value of on cell , Eq. (4) shows the yield level on cell in subplot is affected by the N rate applied on subplot :

To simplify notation, denote as , as , and as . Then, a cell-specific yield function can be written as,

Parameters , , , and were modeled as spatially autocorrelated using unconditional Gaussian geostatistical simulation based on the spherical variogram model. The gstat package (version 2.0.6 (Gräler et al., 2016; Pebesma, 2004)) in R was used for this purpose. Table 1 shows the main variogram parameter values used in generating , , , and , which were chosen to generate yields consistent with those observed in experiments conducted by the DIFM project. (Codes used to generate parameter values are available at <https://github.com/tmieno2/ML_VRA.git>.)

Each cell ’s “true” optimal N application rate, denoted , was calculated by solving a profit maximization problem:

where, and were the assumed prices of the corn and N fertilizer. Figure 4 maps an example of the identified “true” cell-specific optimal N rates.

Trial Design

In each simulation round, five field-trial N rates were determined based on the range of the round’s cell-specific EONRs. Let () denote the percent quantile of the round’s cell-specific EONRs. Then, the five trial design rates were

In each round, a Latin square design was followed to assign each plot one of the five N rates so established. Figure 5 shows an example of a trial design. Under this design, the N rate was orthogonal to the error term. This is important because if they were correlated, bias in the estimation of the impact of nitrogen on yield would be introduced for all of the ML methods. Since the treatment effect at each leaf is obtained simply by taking the difference of the average value of the dependent variable between the control and treated group within the leaf, CF requires the exogeneity of the treatment conditional on the independent variables to estimate the treatment effect consistently. Note that the violation of this condition leads to biased estimation for not just CF and the other ML methods, as well.

Yield Data Generation and Aggregation

A cell-level data set for the whole field, was generated. For a generic cell , yield was generated through Eq. (3), applying parameter values , the disturbance term level , and the cell’s assigned experimental N rate, (Figure 6). These were aggregated up to the subplot level by taking the mean of the cell-level values inside each subplot, which were the unit of observation. The subplot data were therefore

Figure 7 shows an example of observed yield data. Note that data generation at the cell level was conducted to reflect the real-world spatial heterogeneity in soil/field characteristics within the subplot level. But data aggregation to the subplot level was conducted because the large sizes of harvesters make it impossible to monitor yields accurately at smaller scales.

Modeling Scenarios

To evaluate the performances of RF, BRF, CNN and CF, we examined four modeling scenarios, each defined by a dataset assumed to be available to the researcher. The first modeling scenario, denoted “aby” assumed that researchers have a dataset of the actual values , , for every cell and use , and as covariates in their ML models. This scenario represents an ideal situation for researchers, in which all the relevant variables are observed, but the true functional form of heterogeneous yield response functions are not known. More realistic scenarios can be compared with this ideal scenario.

In subsequent modeling scenarios, we made the simulated experimental data more realistic by varying the degree to which researchers understood what the needed covariates’ values were those in the data set. In the second scenario, denoted “abytt,” we introduced additional variables and , and the researcher’s dataset contained values of , , , and for all subplots . and had no effect on yield, but their cell-level values were spatially correlated with , and so their subplot-level values were correlated with . This scenario is meant to reflect the lack of scientific consensus about which variables explain heterogeneous yield response to N, making it likely that researchers include irrelevant variables like and in real-world data analysis.

In the third scenario, denoted “aabbyy,” the researcher’s dataset did not contain values of , , or . Rather, it contained subplot-level values of permutations of , , and . Specifically, for each cell , we assigned one of a set of spatially autocorrelated weights to the variable, another set of spatially autocorrelated weights to the variable, and another set of spatially autocorrelated weights to the variable. (There was no spatial correlation between the and data, between and data, or between and .) Six new covariates were created with the weights and the original covariates , and . These were , , , , , and . For each , and , these procedures created two new covariates that were highly spatially autocorrelated with the original covariate, which was meant to reflect that many of the soil/field characteristics used to analyze OFPE data are spatially correlated. The values of , and did not indicate specific soil or field characteristics, but were functions of such. In reality, scientists cannot directly observe variables that can accurately predict yield plateau level, soil N content and N uptake efficiency. Rather, they use multiple observed soil/field characteristics to explain such phenological phenomena. For example, instead of using directly, some some topographical land features and soil properties such as elevation and soil sand content could be used as yield-limiting factors (Jiang and Thelen, 2004; Kravchenko and Bullock, 2000). In reality, scientists may include more than three variables as covariates. This scenario should reduce the accuracy of EONR modeling for all the models compared to the ideal case.

The fourth scenario, denoted “aabbyytt,” had the same components as aabbyy, but also included the and covariates.

OFPE Experiments and Estimating Site-specific Optimal N Rates

RF, BRF, and CNN

The RF, BRF, and CNN methods followed the same conceptual steps to estimate site-specific EONR: 1) estimate the yield production functions and 2) calculate site-specific EONRs based on that estimated function. For the given modeling scenario, RF and BRF predict yield by using all the available covariates as explanatory variables in the estimation process. The grf package (version 1.2.0 (Tibshirani et al., 2018)) in R was used for RF and BRF modeling.

CNN leaves room for researchers to determine its architecture. In this article, a slightly modified version of one of the multi-stream CNN architectures proposed by Barbosa et al. (2020), called “Late Fusion,” was used. Briefly stated, explanatory variables except N rate (e.g. , , etc.) enter the model as a set of element rasters. On the other hand, the input size of the N rate is treated as because it is spatially homogeneous within each subplot. First, each input was connected to an independent convolutional layer with eight filters each with stride one, followed by a max-pooling layer with stride two. Then, a fully-connected rectified linear unit (ReLU) layer with sixteen neurons was added after each max-pooling layer, followed by a single ReLU neuron. Finally, multiple neurons were concatenated and fed to the fully-connected ReLU layer with sixteen neurons, followed by an output with a linear activation function. Barbosa et al. (2020) demonstrated that among several CNN architectures, this architecture modeled crop yield response to N rate management best. CNN was implemented in Python v3.7.6 using Pytorch v1.7.0 (Paszke et al., 2017). The Adam optimizer (Kingma and Ba, 2014) was used with a learning rate of (default value). To avoid over-fitting, early stopping was used to monitor validation loss with a ten epochs of patience.

Let denote a list of subplot-level explanatory variables (soil/field characteristics). Further, let denote an estimated yield response function, estimated by model ( = RF, BRF, or CNN). Then, yield at from model is . The site-specific EONR for each model can then be found by solving the following profit maximization problem for all for each of the models:

where and are the prices of corn and N.

CF

Unlike RF, BRF, and CNN, CF estimates the impact of a binary treatment. In our context, CF estimates changes in yields (i.e., ) caused by changes in N application rates from one experimental rate (control N rate, i.e., ) to another experimental rate (treatment N rate, i.e., ). Since we have five N application rates, four “experiments” (treatments) are identified based on four (, ) combinations. A possible grouping to make such pairwise N application rate combinations, which we call the “CF-base,” used , and . In this approach, the treatment effect is always estimated against the lowest N rate () as [[5]](#footnote-5).

For estimating site-specific EONR, the trained CF-base were used to predict site-specific changes in yields resultant from changes in application rates from to . Let be the the estimated treatment effect of changing subplot ’s application rate from to for site .

Since the CF-base approach always estimates the treatment effects using the lowest treatment rate , the predicted treatment effects already represents the change in yields from the yield at :

With the treatment effects from the base obtained, site-specific EONRs were identified by finding the N rate resulting in the highest profit. Yield levels were not predicted at any point in the CF-based approaches.

RF, BRF, and CF from the grf package have several tunable hyperparameters (e.g., the minimum node size in each tree, the number of covariates used for node splitting, and parameters involving honest tree-building process), and the optimal values for these parameters are selected by cross-validation. In addition to these tuning parameters, we needed to set the number of trees grown in a forest (i.e., *num.trees*). Individual trees in a forest were built with randomly subsampled observations, leading to a different prediction in every forest even when the same dataset was used. To reduce the variance of predictions, in other words, to stabilize the prediction accuracy, we needed to assign *num.trees* a high value, .

Model Evaluation

The accuracy of EONR estimation was judged based RMSE ()of EONR estimation against the true EONR at the subplot level. In addition, using the EONR estimates, we also calculated loss of profit () (i.e., ) relative to the true maximum profit at the subplot level. The maximized profit is the profit under the true yield response functions evaluated at . For RF, BRF, and CNN, the accuracy of yield prediction was judged based on the RMSE () of the yield prediction against the true yields at the subplot level.

Results and Discussions

EONR Estimation

Table 2 shows the mean RMSE of the EONR estimation and over the simulations; it makes immediately clear that CF-base considerably outperformed the other ML methods, and that the profits based on CF-base’s EONR estimates were the closest to the true maximum profits. BRF performed considerably better predicting EONR than did RF. Interestingly, CNN completely failed in its estimations of site-specific EONRs. This happened because the site-specific yield response functions identified by CNN were linear, and all with the same slope. Therefore, CNN failed to capture a declining marginal product of N. The constant marginal product caused the estimated EONR to take on either the lowest or highest experimental N application rate. To examine whether this problem is product solely of the specific choice of CNN architecture, we also tested another CNN model, which was analogous to CNN-ST proposed by Barbosa et al. (2020). This architecture also estimated that yield and N rates were perfectly linearly related, and also estimated site-specific EONR poorly. Note that Barbosa et al. (2020) just compared models through the lens of yield prediction but did not use the estimated models to derive EONRs. In some cases, CNN predicted yield well; but in all cases it predicting EONR poorly.

Figure 8 and Figure 9 show the distributions of the one thousand RMSEs of the EONR estimations and by ML methods and modeling scenarios, respectively. The center of distribution mass of the CF-base’s RMSE values and are to the left of the centers of the distribution masses of the other methods’ RMSE values, showing that CF-base more accurately and consistently estimated EONR across all the modeling scenarios. Interestingly, RF seems particularly vulnerable to the inclusion of more variables, with RMSE () increasing from 43.7 (27.40) in the “aby” scenario to 49.4 (35.55) in the “aabbyytt” scenario. CF-base was more robust to the inclusion of additional variables. Overall, the simulations present CF-base as the clear winner in estimating EONR.

Figure 10 illustrates the underlying cause of CF-base’s superior EONR predictions. It plots the estimated treatment effects against true treatment effects by treatment types and ML methods (RF, BRF, and CF-base) under the “aabbyytt” scenario in one of the simulations rounds. The figure shows that the estimated treatment effects did not exhibit significant bias, with the points clustered around the red 1-to-1 line. However, they differ substantially in the accuracy of their treatment effect estimations. Points are clustered much more tightly about the 1-to-1 line much for CF-base than RF and BRF. That is, CF-base estimates treatment effects much more efficiently than do the RF and BRF methods, which focus on yield prediction rather than treatment effect estimation.

Yield Prediction and EONR Estimation

For RF, BRF and CNN, Figure 11 plots the RMSE values of EONR estimation against the RMSE values of yield prediction, and the red line shows the regression line describing the relationship between them. For all the models the regression line was almost flat, showing little if any correlation between the two measures. In other words, the quality of a model’s yield predictions provides little if any information about the quality of its EONR predictions.

Table 3 shows the mean RMSE of yield prediction over the simulations. Since CF-base does not predict yield, the table only reports the results from RF, BRF and CNN. Compared to the result of EONR estimation, RF performed as well as BRF in the simple modeling scenarios of “aby” and “abytt.” This fact along with the poor performance of RF in predicting EONR show that strong predictive power of yield levels does not imply accurate prediction of EONR. Further, RF shows vulnerability to the inclusion of many correlated variables, as is displayed for EONR prediction as well. While CNN performs worse than RF and BRF in “aby” and “abytt,” it performs better than RF in more complicated modeling scenarios, showing robustness, at least in yield prediction, to inclusion of many correlated variables. Of course, while on average in the “aabbyy” and “aabbyytt” scenarios CNN predicted yields better than RF did, CNN’s EONR predictions were worse than RF’s, which further backs our claim that a good yield prediction capability does not imply good EONR prediction capability.

All previous research has followed a two-step procedure to estimate site-specific EONRs. The first step is to select the model that predicts yield most accurately. The following step is to use that model to predict EONR. Table 4 shows that following this process led to the choice of the best (as measured by profit) EONR prediction model in fewer than 30% of the “aby” and “abytt” simulation rounds, 40% of the “aabbyy” simulations rounds, and 50% of the “aabbyytt” simulation rounds examined here. For example, the table shows that, in the “aby” scenario, RF attained the lowest RMSE for yield prediction in 458 of iterations, and attained the lowest in 24 of those 458 iterations. That is, in 434 out of 458 rounds, the RF models best at yield prediction did not provide EONR estimates that led to high profits. Similarly, in all cases, the CNN method never provided both the best yield prediction and the best EONR estimation simultaneously. In the rounds in which BRF predicted yield best, then the two-step process worked well. For example, table 4 shows that in the 287 “aby” simulation rounds in which BRF predicted yields best, it offered the best EONR recommendations 267 times. But, BRF predicted yields best in fewer than 30% of the rounds, implying that the two-step procedure failed more than 70%.

Virtually all the previous studies have used yield prediction accuracy to rank and select models. This is understandable, since yield can be “ground-truthed” the true causal impact of the treatment cannot be determined from real-world data, making it impossible to cross-validate EONR. However, the simulation results reported here show clearly that using a model’s yield prediction accuracy as a model selection criterion is not well justified if the ultimate objective is accurate estimation of site-specific EONRs.

Conclusion

ML methods are appealing tools that can potentially improve site-specific input rate management by capturing heterogeneous treatment effects introduced by complex non-linear and multidimensional interactions of soil and field characteristics. Many kinds of prediction-oriented ML methods have been used for this purpose. We have introduced the use of CF, a relatively new method, which, unlike prediction-oriented ML methods, focuses on the identification of heterogeneous causal effects of treatment.

We examined the use of CF-base for site-specific input management recommendations. Using Monte Carlo simulations under various modeling scenarios, we compared the CF estimations of site-specific EONR to those of the prediction-oriented ML methods RF, BRF, and CNN. CF-base consistently outperformed the other methods across all the modeling scenarios in terms of EOIR prediction and profit generation. We also showed that the model’s yield prediction accuracy was uncorrelated with its EONR prediction accuracy, and so yield prediction accuracy should not be used as a model selection criterion if the ultimate goal of the experiment is to identify EOIR map.

Despite the relatively simple nature of the simulations, results reported provide intriguing evidence of potential advantages of CF over other ML methods when the purpose of the research is to increase the economic efficiency of input management. Of course, the reported research did not model all situations, and further research is needed. First, a drawback of the CF method is that because it does not estimate a continuous yield response function, it can only compare the economic outcomes of a limited set of experimental N rates included in the field trial that provides the data. This raises important questions about trial design. How many distinct experimental input application rates should be included? All the trial designs examined here were based on five N rates. Having examined models with additional rates would have increased the number of possible N rates from which the EONR were selected, but reduced treatment pair replications and so the estimation accuracy.

Second, the size of the field used in the OFPE can affect the relative performances of the competing methods. CF requires honest sampling for its unbiasedness property. Honest sampling comes at the cost of losing prediction accuracy of treatment effects because fewer samples are used to estimate treatment effects within leaves (Athey and Imbens, 2016). Therefore, if the ML analyses are conducted with OFPE data from a small field, CF may not outperform other ML methods

Third, another interesting subject for future research is the potential of combining the ML methods with post-estimation spatial smoothing of the estimated EONR. It is a well known fact that soil/field characteristics are spatially correlated (F. Dormann et al., 2007; Goovaerts, 1999); this spatial correlation is likely to cause spatial correlation in EONRs, which would suggest that post-estimation spatial smoothing of EONRs may further improve the accuracy of EONR estimation.

Finally, a natural and fruitful extension of the reported research would be to test CF and other ML methods in a dynamic setting where multiple years of experiments are conducted, allowing weather events into the analysis as explanatory variables. Crop simulation models (e.g., APSIM, DSSAT) could be used in place of the Mitscherlich-Baule production function used here to model the underlying yield response function, which could generate more realistic data, using actual soil/field characteristics and weather instead of abstract parameters in the analysis, and so better include the dynamic nature of cropping systems in the study of N management (Archontoulis et al., 2020; Puntel et al., 2018, 2016).

Acknowledgements

This research was supported by a USDA-NIFA-AFRI Food Security Program Coordinated Agricultural Project, titled “Using Precision Technology in On-farm Field Trials to Enable Data-Intensive Fertilizer Management,” (Accession Number 2016-68004-24769), and also by the a USDA-NRCS Conservation Innovation Grant from the On-farm Trials Program, titled “Improving the Economic and Ecological Sustainability of US Crop Production through On-Farm Precision Experimentation” (Award Number NR213A7500013G021).

Tables

Table 1: The Parameters for the Variogram Modeling

| Parameters | Range | Mean | Nugget | Sill |
| --- | --- | --- | --- | --- |
| ymaxi,j | 400 | 1.2⨯104 | 0 | 2.0⨯106 |
| αi,j | 400 | -0.5 | 0 | 2.0⨯10-2 |
| βi,j | 400 | 0 | 0 | 1.0 |
| εi,j | 400 | 0 | 0 | 1.5⨯10-2 |

Table 2: Mean RMSE of EONR Estimation and Profit-deficit by ML Methods and Modeling Scenarios

| Model |  | RF | |  | BRF | |  | CNN | |  | CF-base | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | RMSE | 𝜋̂def |  | RMSE | 𝜋̂loss |  | RMSE | 𝜋̂loss |  | RMSE | 𝜋̂loss |
| aby |  | 43.7 | 27.40 |  | 39.1 | 22.23 |  | 62.8 | 58.99 |  | 25.1 | 6.81 |
| abytt |  | 46.0 | 30.60 |  | 38.4 | 21.60 |  | 62.8 | 58.99 |  | 24.8 | 6.59 |
| aabbyy |  | 48.2 | 33.73 |  | 40.2 | 23.31 |  | 62.8 | 58.99 |  | 25.7 | 7.08 |
| aabbyytt |  | 49.4 | 35.55 |  | 39.5 | 22.75 |  | 62.8 | 58.99 |  | 24.8 | 6.64 |
| NOTE: 𝜋̂loss indicates loss of profit ($/ha) relative to the true maximum profit at the subplot level. The maximized profit is the profit under the true yield response functions evaluated at Nⁱopt. | | | | | | | | | | | | |

Table 3: Mean RMSE of Yield Prediction by ML Methods and Modeling Scenarios

| Model | RF | BRF | CNN | CF-base |
| --- | --- | --- | --- | --- |
| aby | 1313.8 | 1321.1 | 1368.8 | - |
| abytt | 1337.9 | 1351.5 | 1426.2 | - |
| aabbyy | 1441.7 | 1406.1 | 1426.5 | - |
| aabbyytt | 1469.1 | 1415.9 | 1473.3 | - |

Table 4: Relationship between Yield Prediction Performance and EONR Estimation Performance for Prediction-oriented ML Methods

| Model | RF | BRF | CNN | Total |
| --- | --- | --- | --- | --- |
| aby | 24 (458) | 267 (287) | 0 (255) | 291 |
| abytt | 0 (510) | 281 (281) | 0 (209) | 281 |
| aabbyy | 1 (207) | 380 (384) | 0 (409) | 381 |
| aabbyytt | 0 (208) | 472 (474) | 0 (318) | 472 |
| NOTE: #Y denoted the number of simulation rounds in which the model provided the lowest RMSE for yield prediciton. #YE indicates the number of simulation rounds in which the model provided the lowest RMSEs for both yield prediction and EONR prediction. | | | | |

I would make the table like this:

Table 4: Relationship between Yield Prediction Performance and EONR Estimation Performance for Prediction-oriented ML Methods

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | RF | | BRF | | CNN | |  |
|  | #YE Number of rounds model provided best yield prediction | #Y Number of rounds model provided best EONR estimation along with best yield prediction | Number of rounds model provided best yield prediction | Number of rounds model provided best EONR estimation along with best yield prediction | Number of rounds model provided best yield prediction | Number of rounds model provided best EONR estimation along with best yield prediction | Total number of rounds in which models provided best EONR estimation along with best yield prediction |
| aby | 458 | 24 | 287 | 267 | 255 | 0 | 291 |
| abytt | 510 | 0 | 281 | 281 | 209 | 0 | 281 |
| aabbyy | 207 | 1 | 384 | 380 | 409 | 0 | 381 |
| aabbyytt | 208 | 0 | 474 | 472 | 318 | 0 | 472 |

Figures

Figure 1: An Illustrative Example of Causal Tree

Figure 2: Division of Field: plots, subplots and cells

Figure 3: An Illustrative Example of Spatial Distributions of Field Characteristics

Figure 4: An Illustrative Example of a Spatial Distribution of True Optimal N Rate

Figure 5: An Illustrative Example of Experimental Design of N

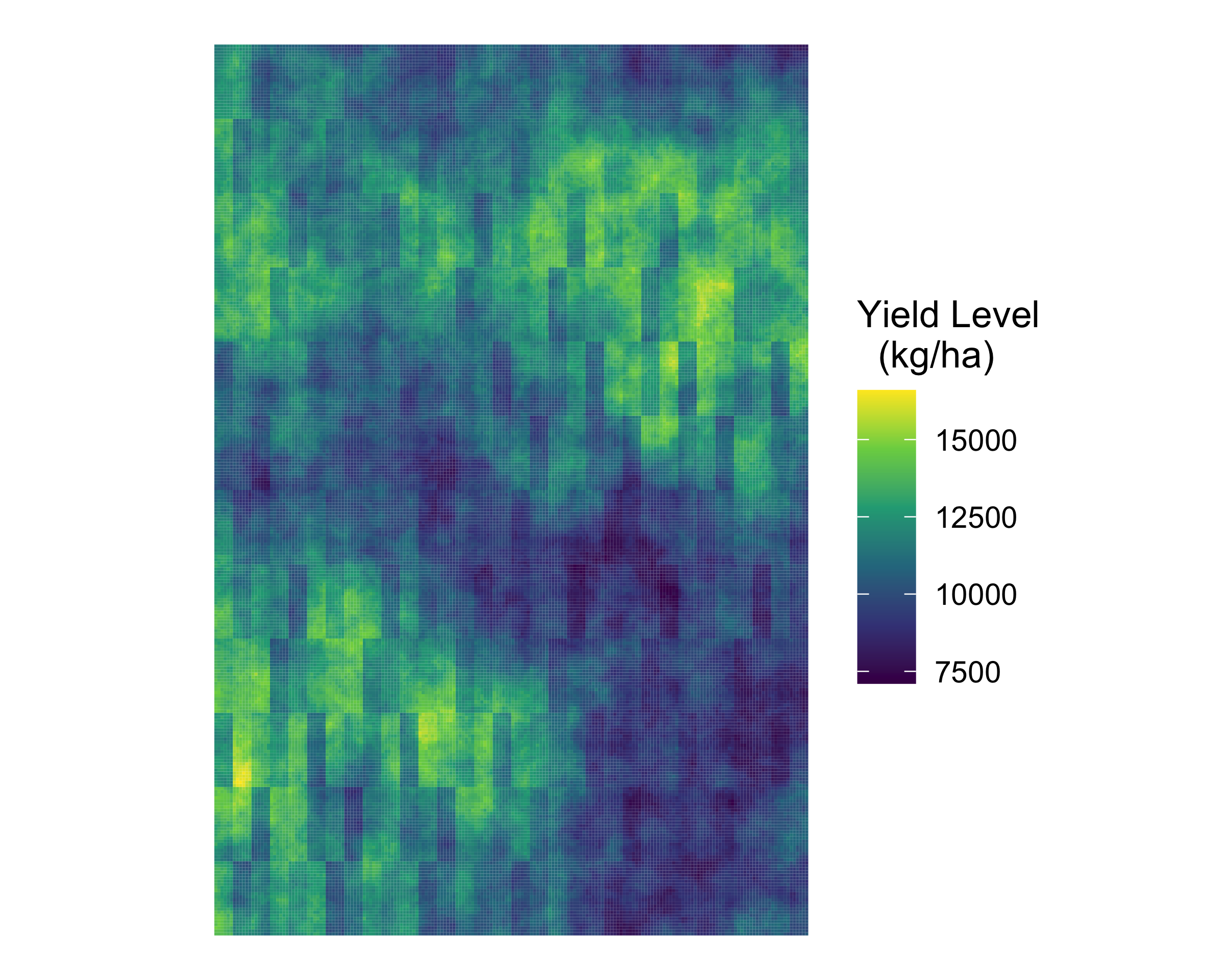


Figure 6: An Illustrative Example of a Yield Distribution Map at the Cell resolution

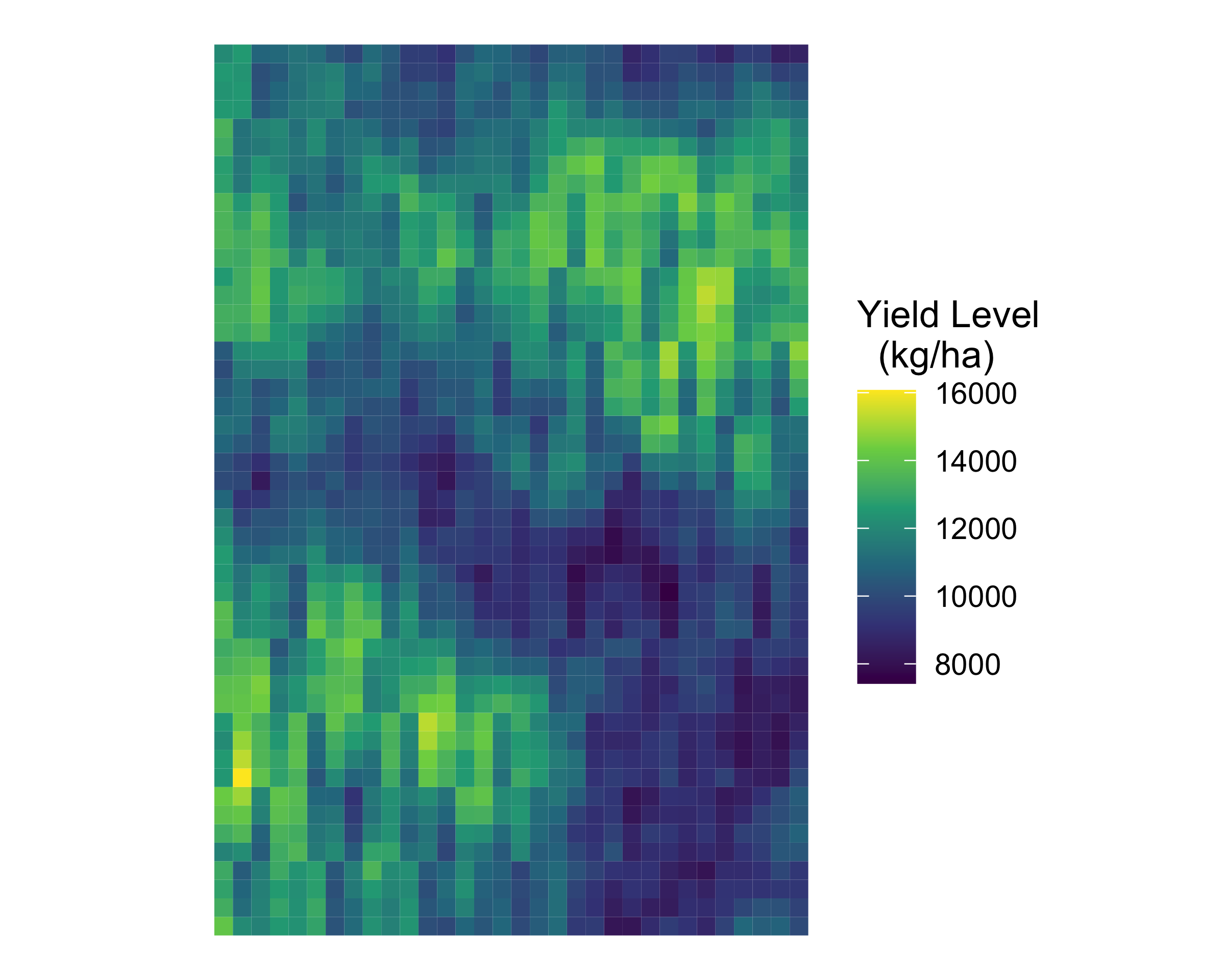


Figure 7: An Illustrative Example of a Yield Distribution Map at the Subplot resolution

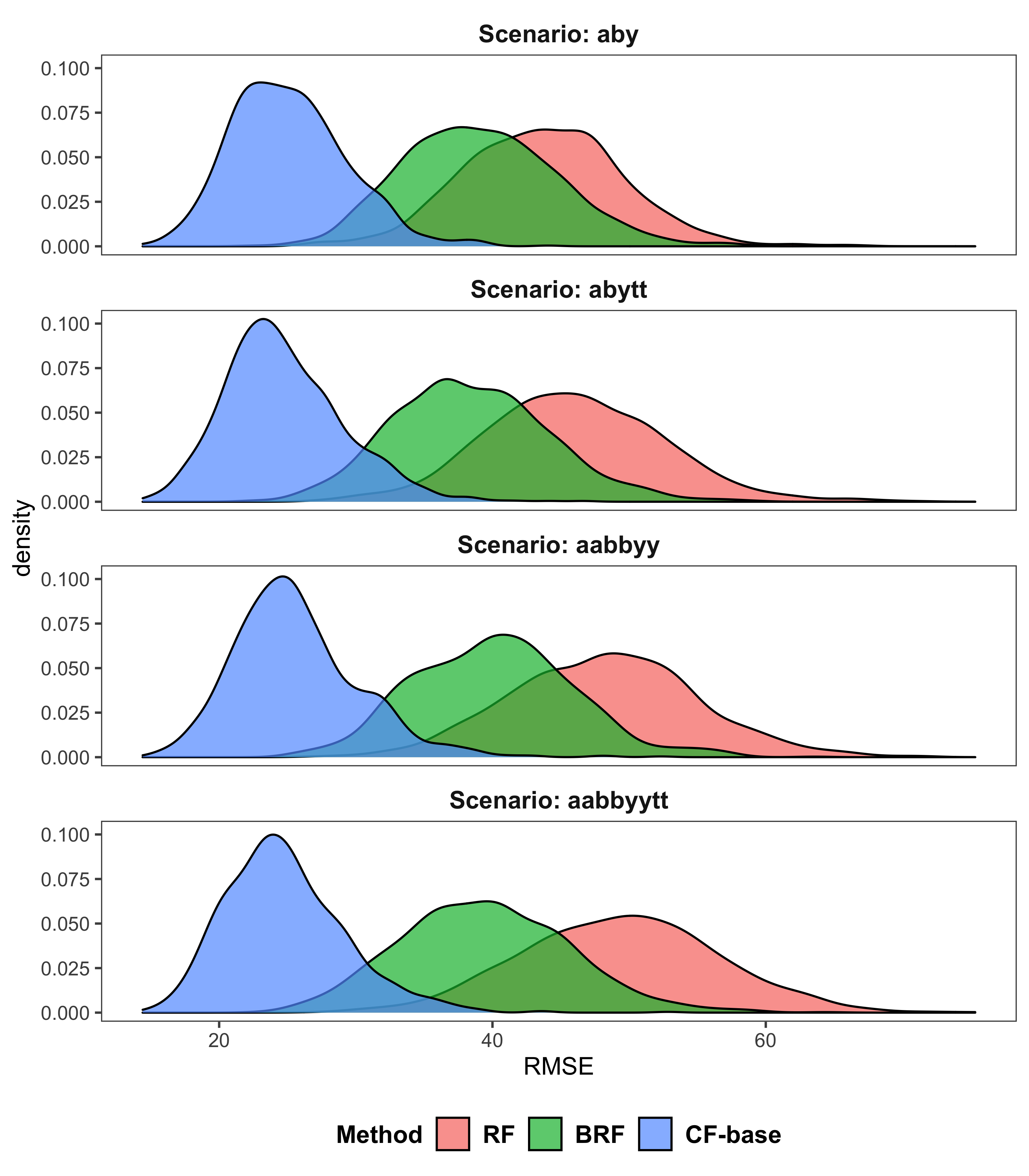


Figure 8: Distributions of RMSE of EONR Estimation over Simulations

Figure 9: Distributions of Profit Deficit over Simulations



Figure 10: True Treatment Effects vs Estimated Treatment Effects (Scenario: aabbyytt)

NOTE: The red line in the figures denotes a 1-to-1 line to show the ideal relationship between true and estimated treatment effects.

Figure 11: The Relationship in the Performance between EONR Estimation and Yield Prediction by ML Methods and Modeling Scenarios

Figure 11: The Relationship in the Performance between EONR Estimation and Yield Prediction by ML Methods and Modeling Scenarios

NOTE: The red line in the figures denotes the best-fit line to explain the relationship between the RMSEs of EONR prediction and yield prediction.

Apepndix

CF-stepwise and CF-base

While CF-base always used as as , the grouping approach of CF-stepwise used , and . Under this set of N application rate combinations, the change in yields from the yield at can be calculated by sequentially adding the estimated treatment effects:

Then, site-specific EONRs were identified in the same manner as CF-base approach.

Table 5 compare the performance of CF-stepwise with that of CF-base in terms of the mean RMSE of the EONR estimation and over the simulations. CF-stepwise showed the comparable performance to CF-base. Figure 12 illustrates the distributions of the mean RMSE of the EONR estimation. Overall, the distributions of CF-base have longer right tail than the distributions of CF-stepwise, meaning that CF-base suffered from some large errors compared to CF-stepwise. Meanwhile, when looking at the distributions of mean (Figure 13), it turns out that CF-base’s loss of profit relative to the maximized profit was smaller than CF-stepwise’s in most of the cases. By nature, RMSE gives a relatively high weight to large errors. However, even though CF-base incurs some such larger prediction errors compared to CF-stepwise, those errors may not have economically significant consequences, and the results suggest that overall the EONR estimates from CF-base are more accurate and consistent than CF-stepwise to derive higher profit. Hence, we conclude that CF-base is the preferred CF-based method to CF-stepwise.

Table

Table 5: Comparison of CF-base and CF-stepwise: Mean RMSE of EONR Estimation and Profit-loss

| Model |  | CF-base | |  | CF-stepwise | |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | RMSE | 𝜋̂loss |  | RMSE | 𝜋̂loss |
| aby |  | 25.1 | 6.81 |  | 25.2 | 7.09 |
| abytt |  | 24.8 | 6.59 |  | 24.3 | 6.57 |
| aabbyy |  | 25.7 | 7.08 |  | 25.3 | 7.10 |
| aabbyytt |  | 24.8 | 6.64 |  | 24.2 | 6.51 |

Figures

Figure 12: Comparison of CF-base and CF-stepwise: Distributions of RMSE of EONR Estimation over Simulations

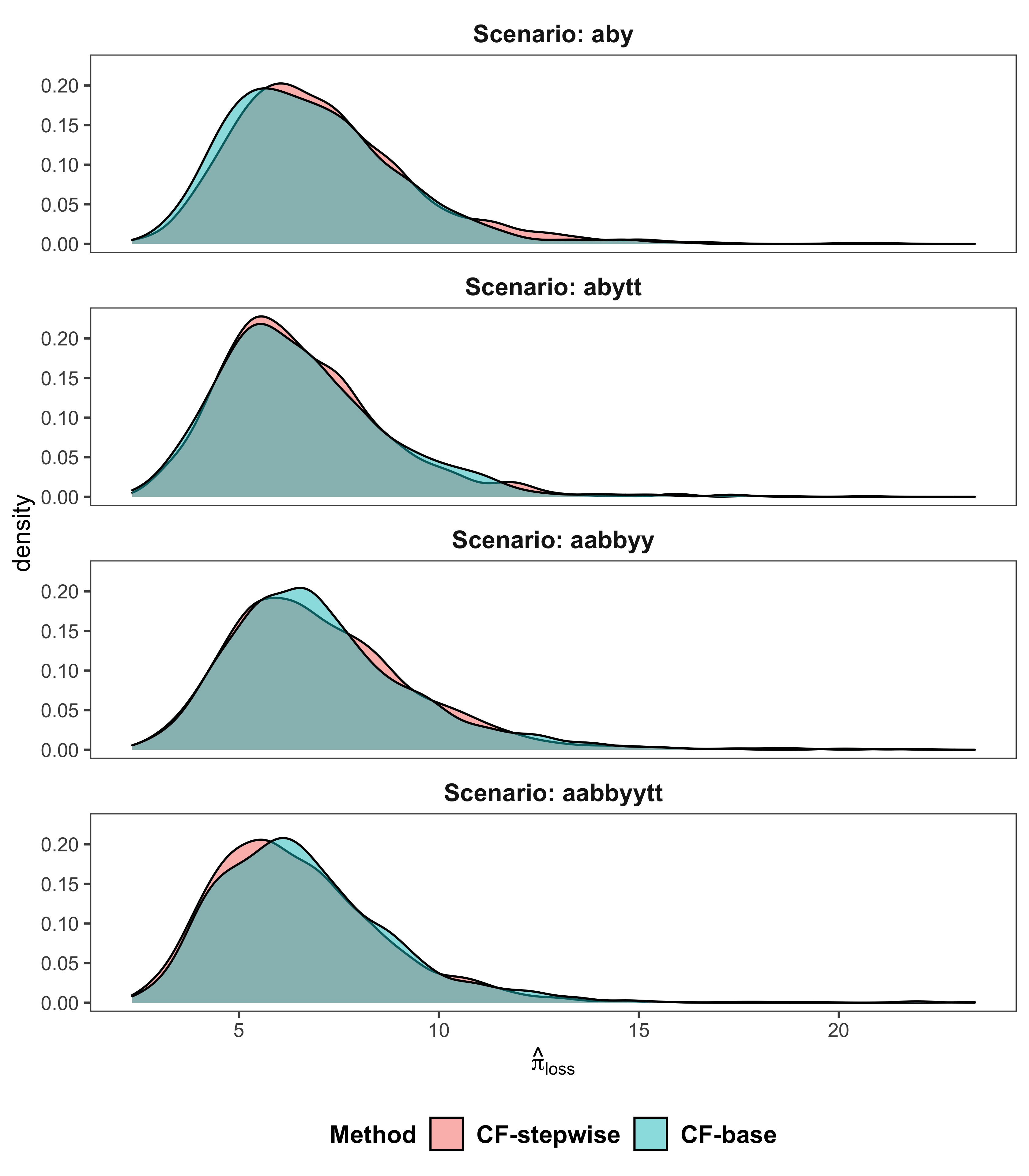


Figure 13: Comparison of CF-base and CF-stepwise: Distributions of Profit Deficit over Simulations

References

Alva, A., Paramasivam, S., Fares, A., Delgado, J., Mattos Jr, D., Sajwan, K., 2006. Nitrogen and irrigation management practices to improve nitrogen uptake efficiency and minimize leaching losses. Journal of Crop Improvement 15, 369–420.

Anselin, L., Bongiovanni, R., Lowenberg-DeBoer, J., 2004. A spatial econometric approach to the economics of site-specific nitrogen management in corn production. American Journal of Agricultural Economics 86, 675–687.

Archontoulis, S.V., Castellano, M.J., Licht, M.A., Nichols, V., Baum, M., Huber, I., Martinez-Feria, R., Puntel, L., Ordóñez, R.A., Iqbal, J., others, 2020. Predicting crop yields and soil-plant nitrogen dynamics in the US corn belt. Crop Science 60, 721–738.

Arti, S., Hidayah, I., Kusumawardani, S.S., 2020. Research trend of causal machine learning method: A literature review. IJID (International Journal on Informatics for Development) 9, 111–118.

Athey, S., 2018. The impact of machine learning on economics, in: The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, pp. 507–547.

Athey, S., 2017. Beyond prediction: Using big data for policy problems. Science 355, 483–485.

Athey, S., Imbens, G., 2016. Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences 113, 7353–7360.

Athey, S., Imbens, G.W., 2019. Machine Learning Methods Economists Should Know About | Stanford Graduate School of Business. Working Paper,.

Barbosa, A., Trevisan, R., Hovakimyan, N., Martin, N.F., 2020. Modeling yield response to crop management using convolutional neural networks. Computers and Electronics in Agriculture 170, 105197.

Bozorgi, Z.D., Teinemaa, I., Dumas, M., La Rosa, M., Polyvyanyy, A., 2020. Process mining meets causal machine learning: Discovering causal rules from event logs, in: 2020 2nd International Conference on Process Mining (ICPM). IEEE, pp. 129–136.

Bullock, D.G., Bullock, D.S., Nafziger, E.D., Doerge, T.A., Paszkiewicz, S.R., Carter, P.R., Peterson, T.A., 1998. Does variable rate seeding of corn pay? Agronomy journal 90, 830–836.

Bullock, D.S., Boerngen, M., Tao, H., Maxwell, B., Luck, J.D., Shiratsuchi, L., Puntel, L., Martin, N.F., 2019. The data-intensive farm management project: Changing agronomic research through on-farm precision experimentation. Agronomy Journal 111, 2736–2746.

Bullock, D.S., Lowenberg-DeBoer, J., Swinton, S.M., 2002. Adding value to spatially managed inputs by understanding site-specific yield response. Agricultural Economics 27, 233–245.

Carbo-Valverde, S., Cuadros-Solas, P., Rodrı́guez-Fernández, F., 2020. A machine learning approach to the digitalization of bank customers: Evidence from random and causal forests. Plos one 15, e0240362.

F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., others, 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. Ecography 30, 609–628.

Frank, M.D., Beattie, B.R., Embleton, M.E., 1990. A comparison of alternative crop response models. American Journal of Agricultural Economics 72, 597–603.

Goovaerts, P., 1999. Geostatistics in soil science: State-of-the-art and perspectives. Geoderma 89, 1–45.

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. The R Journal 8, 204–218.

Grant Gardner, D.S.B., Taro Mieno, 2021. An economic evaluation of site-specific input application r x maps: Evaluation framework and case study. Precision Agriculture.

Hurley, T.M., Malzer, G.L., Kilian, B., 2004. Estimating site-specific nitrogen crop response functions: A conceptual framework and geostatistical model. Agronomy Journal 96, 1331–1343.

Jiang, P., Thelen, K., 2004. Effect of soil and topographic properties on crop yield in a north-central corn–soybean cropping system. Agronomy Journal 96, 252–258.

Johnson, C., Albrecht, G., Ketterings, Q., Beckman, J., Stockin, K., 2005. Nitrogen basics–the nitrogen cycle. Agronomy Fact Sheet Series, Fact Sheet 2.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Krause, M., Crossman, S., DuMond, T., Lott, R., Swede, J., Arliss, S., Robbins, R., Ochs, D., Gore, M.A., 2020. Random forest regression for optimizing variable planting rates for corn and soybean using topographical and soil data.

Kravchenko, A.N., Bullock, D.G., 2000. Correlation of corn and soybean grain yield with topography and soil properties. Agronomy Journal 92, 75–83.

Laurent, A., Kyveryga, P., Makowski, D., Miguez, F., 2019. A framework for visualization and analysis of agronomic field trials from on‐farm research networks. Agronomy Journal 111, 2712–2723. <https://doi.org/10.2134/agronj2019.02.0135>

Licht, M., Witt, M., 2019. Conducting on-farm trials. crops.extension.iastate.edu.

Llewelyn, R.V., Featherstone, A.M., 1997. A comparison of crop production functions using simulated data for irrigated corn in western kansas. Agricultural Systems 54, 521–538.

Lowenberg-DeBoer, J., Erickson, B., 2019. Setting the record straight on precision agriculture adoption. Agronomy Journal 111, 1552–1569.

Makowski, D., Wallach, D., Meynard, J.-M., 2001. Statistical methods for predicting responses to applied nitrogen and calculating optimal nitrogen rates. Agronomy journal 93, 531–539.

Mamo, M., Malzer, G.L., Mulla, D., Huggins, D., Strock, J., 2003. Spatial and temporal variation in economically optimum nitrogen rate for corn. Agronomy Journal 95, 958–964.

Miao, Y., Mulla, D.J., Robert, P.C., Hernandez, J.A., 2006. Within-field variation in corn yield and grain quality responses to nitrogen fertilization and hybrid selection. Agronomy Journal 98, 129–140. <https://doi.org/10.2134/agronj2005.0120>

Moraffah, R., Karami, M., Guo, R., Raglin, A., Liu, H., 2020. Causal interpretability for machine learning-problems, methods and evaluation. ACM SIGKDD Explorations Newsletter 22, 18–33.

Paris, Q., 1992. The von liebig hypothesis. American Journal of Agricultural Economics 74, 1019–1028.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in pytorch.

Pebesma, E.J., 2004. Multivariable geostatistics in S: The gstat package. Computers & Geosciences 30, 683–691.

Puntel, L.A., Pagani, A., Archontoulis, S.V., 2019. Development of a nitrogen recommendation tool for corn considering static and dynamic variables. European Journal of Agronomy 105, 189–199.

Puntel, L.A., Sawyer, J.E., Barker, D.W., Dietzel, R., Poffenbarger, H., Castellano, M.J., Moore, K.J., Thorburn, P., Archontoulis, S.V., 2016. Modeling long-term corn yield response to nitrogen rate and crop rotation. Frontiers in plant science 7, 1630.

Puntel, L.A., Sawyer, J.E., Barker, D.W., Thorburn, P.J., Castellano, M.J., Moore, K.J., VanLoocke, A., Heaton, E.A., Archontoulis, S.V., 2018. A systems modeling approach to forecast corn economic optimum nitrogen rate. Frontiers in plant science 9, 436.

Richens, J.G., Lee, C.M., Johri, S., 2020. Improving the accuracy of medical diagnosis with causal machine learning. Nature communications 11, 1–9.

Ruffo, M.L., Bollero, G.A., Bullock, D.S., Bullock, D.G., 2006. Site-specific production functions for variable rate corn nitrogen fertilization. Precision Agriculture 7, 327–342.

Scharf, P.C., Kitchen, N.R., Sudduth, K.A., Davis, J.G., Hubbard, V.C., Lory, J.A., 2005. Field-scale variability in optimal nitrogen fertilizer rate for corn. Agronomy journal 97, 452–461.

Scharf, P.C., Wiebold, W.J., Lory, J.A., 2002. Corn yield response to nitrogen fertilizer timing and deficiency level. Agronomy Journal 94, 435–441.

Schölkopf, B., Locatello, F., Bauer, S., Ke, N.R., Kalchbrenner, N., Goyal, A., Bengio, Y., 2021. Toward causal representation learning. Proceedings of the IEEE 109, 612–634.

Storm, H., Baylis, K., Heckelei, T., 2020. Machine learning in agricultural and applied economics. European Review of Agricultural Economics 47, 849–892. <https://doi.org/10.1093/erae/jbz033>

Tibshirani, J., Athey, S., Wager, S., Friedberg, R., Miner, L., Wright, M., Tibshirani, M.J., Rcpp, L., DiceKriging, R.I., SystemRequirements, G., 2018. Package ‘grf.’

Wager, S., Athey, S., 2018. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. Journal of the American Statistical Association 113, 1228–1242. <https://doi.org/10.1080/01621459.2017.1319839>

1. Department of Agricultural Economics, University of Nebraska-Lincoln, email; [skakimoto3@hunskers.unl.edu](mailto:skakimoto3@hunskers.unl.edu) [↑](#footnote-ref-1)
2. Department of Agricultural Economics, University of Nebraska-Lincoln, email; [tmieno2@unl.edu](mailto:tmieno2@unl.edu) [↑](#footnote-ref-2)
3. Artificial Intelligence Advanced Research Center, Faculty of Applied Biological Sciences, Gifu University, email; [takashit@gifu-u.ac.jp](mailto:takashit@gifu-u.ac.jp) [↑](#footnote-ref-3)
4. Department of Agricultural and Consumer Economics, University of Illinois, email; [dsbulloc@illinois.edu](mailto:dsbulloc@illinois.edu) [↑](#footnote-ref-4)
5. Another potential grouping approach we considered used , and . We label this approach as the “CF-stepwise.” See Appendix for the detail and the result. [↑](#footnote-ref-5)