



Can machine learning models provide accurate fertilizer recommendations?

Takashi S. T. Tanaka^{1,2,7} · Gerard B. M. Heuvelink^{3,4} · Taro Mieno⁵ · David S. Bullock⁶

Accepted: 8 March 2024 / Published online: 25 March 2024
© The Author(s) 2024

Abstract

Accurate modeling of site-specific crop yield response is key to providing farmers with accurate site-specific economically optimal input rates (EOIRs) recommendations. Many studies have demonstrated that machine learning models can accurately predict yield. These models have also been used to analyze the effect of fertilizer application rates on yield and derive EOIRs. But models with accurate yield prediction can still provide highly inaccurate input application recommendations. This study quantified the uncertainty generated when using machine learning methods to model the effect of fertilizer application on site-specific crop yield response. The study uses real on-farm precision experimental data to evaluate the influence of the choice of machine learning algorithms and covariate selection on yield and EOIR prediction. The crop is winter wheat, and the inputs considered are a slow-release basal fertilizer NPK 25–6–4 and a top-dressed fertilizer NPK 17–0–17. Random forest, XGBoost, support vector regression, and artificial neural network algorithms were trained with 255 sets of covariates derived from combining eight different soil properties. Results indicate that both the predicted EOIRs and associated gained profits are highly sensitive to the choice of machine learning algorithm and covariate selection. The coefficients of variation of EOIRs derived from all possible combinations of covariate selection ranged from 13.3 to 31.5% for basal fertilization and from 14.2 to 30.5% for top-dressing. These findings indicate that while machine learning can be useful for predicting site-specific crop yield levels, it must be used with caution in making fertilizer application rate recommendations.

Keywords Economically optimal input rate · On-farm experimentation · Site-specific management · Variable-rate application · Winter wheat

Introduction

Site-specific crop management aims to use information about within-field variability of soil and topographic properties to increase farming profitability and sustainability. Understanding of site-specific crop yield response facilitates effective site-specific crop management

Extended author information available on the last page of the article

(Bullock et al., 2019). Until recently, site-specific crop management was principally based on farmers' and agronomists' experiences and expectations about crop responses to agro-nomic inputs. The expectations are based largely on inferences obtained from conventional small-plot trials that are presumed to represent what is occurring elsewhere. But these trials are expensive and labor-intensive, and it may be inappropriate to draw inferences from small-plot trials to improve management across many farms (Bullock et al., 2019; Lacoste et al., 2022). In contrast, on-farm experimentations have a potential to provide more actionable and practical insights to farmers as an alternative to small-plot trials.

Based on experimental data, process-based crop simulation models, such as APSIM (Holzworth et al., 2014), DSSAT (Hoogenboom et al., 2019), and WOFOST (Boogaard & de Wit, 2020) have been developed to understand the effects of crop management, soil, and weather on crop growth and final yield. Crop simulation models are basically point-based (Heuvelink et al., 2010). Spatialization of these models is of interest to precision agriculture (PA) as it might contribute to optimizing site-specific crop management (Pasquel et al., 2022). However, spatialization of crop simulation models requires knowledge of site-specific input application rates and model parameters that are difficult to estimate because of data scarcity. Environmental and agricultural models (e.g., crop simulation models) also suffer from error propagation as the uncertainty in model inputs influence the output (Corner et al., 2008; Heuvelink, 1998). Furthermore, crop simulation models can only predict potential, water-limited, or nutrient-limited yield, not the actual yield if other environmental variables not accounted for (e.g., weeds, insect pests, and disease) greatly affect yields (de Wit et al., 2019). Therefore, it is not straightforward to use crop simulation models for the purpose of optimizing site-specific input management.

On-farm precision experimentation (OFPE) is a form of on-farm experimentation that uses PA technology to generate large amounts of crop input application and yield response data. Such data can be used to estimate spatially variable optimal input application rates and thus improve site-specific decision making (Bullock et al., 2019). Combining OFPE and machine learning approaches is expected to present an opportunity to facilitate understanding of site-specific crop yield response (Bullock et al., 2019). Since the early stage in the development of site-specific crop management, a wide range of models (e.g., intuitive, stochastic, and machine learning models) has been proposed to support farmers' decision on the rate and timing of fertilizer application at a given location (Adams et al., 2000). The use of various statistical approaches and machine learning algorithms is nowadays becoming a hot topic, but no consensus has been reached on which model is the best.

Many studies have demonstrated the advantages of using spatial statistical modeling methods, including geographically weighted regression (Evans et al., 2020; Trevisan et al., 2021) and machine learning techniques, such as random forest (RF) (Krause et al., 2020; Paccioretti et al., 2021; Wen et al., 2021) and convolutional neural networks (Barbosa et al., 2020). Although Evans et al. (2020) and Trevisan et al. (2021) attempted to explicitly model the spatially variable crop yield responses, most of previous studies focused only on the accuracy of crop yield prediction. Kakimoto et al. (2022) demonstrated that a machine learning model that accurately predicts site-specific yield levels does not necessarily accurately predict yield response and the associated site-specific economically optimal input rates (EOIRs) of fertilizer. They highlighted the distinction between predicting yield levels at observed input rates and estimating yield response to input. For site-specific input management recommendations, the latter is critical, but not necessarily the former. Estimat-

ing site-specific EOIRs accurately requires that the causal relationship between agronomic inputs and crop yield be discovered accurately.

Covariate selection is an essential process in machine learning modeling. Using machine learning for yield prediction can underestimate the impact of nitrogen fertilizer (N) on crop yields and EOIRs because of the inclusion of redundant or strongly correlated covariates (Kakimoto et al., 2022). Estimation of the impact of input on yield may also be biased when an important covariate is not included. With the increased adoption of PA technologies in commercial farms, there are numerous possibilities for selecting covariates (e.g., elevation data, satellite imagery, on-the-go soil sensor data, and digital soil maps) in establishing a yield prediction model for OFPE. Selecting only influential covariates may be a gold standard in establishing models, but rarely can practitioners identify and quantify the complete set of covariates contributing to yield variability. Previous studies have paid very little attention to the sensitivity of the quality of fertilizer management recommendations to different machine learning approaches.

Many studies have used synthetic data to compare the effects of machine learning algorithms, covariate selection, and experimental design on yield and EOIR prediction accuracies (Alesso et al., 2020; Kakimoto et al., 2022; Saikai et al., 2020). To assess the prediction accuracy of site-specific crop yield response modeling, synthetic data can be generated using crop yield response functions, such as process-based crop simulation models (e.g., APSIM) and mechanistic models (e.g., the Mitscherlich-Baule function). Synthetic data can simulate ‘true’ crop yield response, which enables validating EOIR prediction accuracy. One of the shortcomings of synthetic data is that the spatial distribution of yield and yield-limiting factors are generated based on simple assumptions. Although previous studies have considered random noise (e.g., the nugget effect), real farms have more artifacts, such as wheels, overlaps, missing strips of inputs, and further historical land uses (Roques et al., 2022; Zhou et al., 2022). Therefore, synthetic data cannot fully represent the real-world conditions, and may not be capable of providing fair insights into the model uncertainty in machine learning approaches to the analysis of OFPE data.

The aim of this study was to quantify the uncertainty involved in modeling the inclusion of the application rates of two fertilizers and soil properties as covariates in a machine learning model of site-specific crop yield prediction, and to examine how the model uncertainty quantitatively affects the estimation of site-specific EOIRs and gained profits. An OFPE was conducted in Japan to assess the effects of soil properties and application rates of basal and top-dressed fertilizer on winter wheat yield. Site-specific crop yield response models were established using different combinations of machine learning algorithm and covariate selection. Site-specific EOIRs were derived for each of these combinations. A frequency distribution of the estimated EOIRs and gained profits was further assessed as a measure of model uncertainty.

Materials and methods

Experimental design and data collection

A split-plot or checkerboard OFPE (Fig. 1) was implemented in 2019–2020 Gifu, Japan ($35^{\circ}11'N$, $136^{\circ}39'E$) to measure the effects of changing fertilizer application rates on the

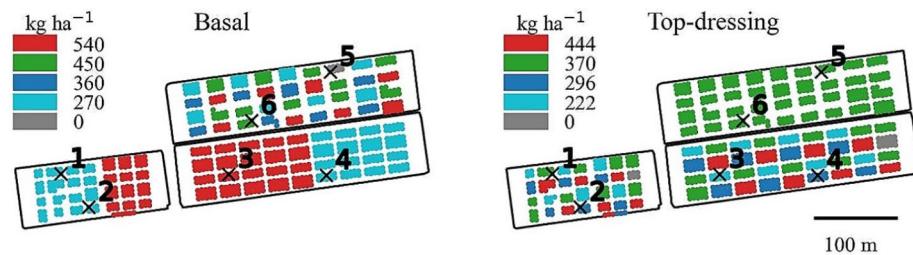


Fig. 1 Experimental design of the on-farm precision experiment. White space represents borders (e.g., transition zones and headlands). The numerals beside the X marks show the six randomly selected locations from which data were obtained to create scatterplots and histograms of the EOIRs of the basal NPK 25–6–4 and top-dressed NPK 17–0–1 fertilizers (Fig. 7)

yield of the ‘Satonosora’ wheat variety. The trial was conducted in cooperation with the Japanese farming company Fukue-eino, which owned the variable-rate application and yield monitoring equipment. In the first OFPE season, a checkerboard design was not implemented across all fields because the farmer was not convinced that the rate transition between plots could be achieved smoothly. Therefore, a split-plot design was implemented for the rest of the fields. Just before seeding (early November), a slow-release basal fertilizer NPK 25–6–4 was applied at rates of 0, 270, 360, 450, and 540 kg ha⁻¹. Before the booting stage (early March, Zadoks 41), NPK 17–0–17 was top-dressed at rates of 0, 222, 296, 370, and 444 kg ha⁻¹. The number of plots receiving no fertilizer was limited due to the risk of yield loss. A variable-rate fertilizer broadcaster with an 18-m working width (Axis 40.2, Kuhn, France) was used for both applications. All other managements (e.g., disease and weed control) were uniform. No serious disease and weed problems were observed. Yield data were collected using a combine harvester with a yield monitor sensor (WRH1200, Kubota, Japan). Although the combine had a 2.6-m header width, after data preprocessing based on the manufacturer’s recommended procedures yield values were averaged to obtain single values for each 5 m × 5 m cell within a grid. Cells in “transition zones” at the beginnings and ends of trial plots, in headlands and/or in buffer zones around the field’s perimeter were excluded from further analysis. The resulting dataset used for analysis contained 970 observations at 5 m × 5 m spatial resolution.

Soil properties were used as covariates to perform site-specific crop yield response assessment. In mid October 2019, prior to the basal fertilizer application, a total of 52 soil samples were collected near the centroids of a 30 m × 30 m grid defined over the field. Within a 1 m² area over the centroid of each soil sampling grid cells, three randomly located partial surface soil samples (0–150 mm) weighing approximately 0.5 kg each were collected and mixed to produce one composite sample. The composite samples were air-dried and sieved through a 2.0-mm mesh before chemical analysis. Soil pH, electrical conductivity (EC), mineralizable N, available phosphorus (P), cation exchange capacity (CEC), exchangeable calcium (Ca), exchangeable magnesium (Mg) and exchangeable potassium (K) were measured. Mineralizable N was determined according to Inokō’s (1986) method. Soils were anaerobically incubated at 30 °C for four weeks, and inorganic N was extracted with a 2 M KCl solution. The concentrations of NH₄⁺ and NO₃⁻ in the extracts were determined using the indophenol method (Keeney & Nelson, 2015) and the Cataldo method (Cataldo et al., 1975). Mineralizable N was calculated by balancing the inorganic N (NH₄⁺

and NO_3^-) before and after anaerobic incubation. Available P was measured by the Truog method (Truog, 1930). Cation exchange capacity was measured by saturating the soil with a neutral 1 mol L⁻¹ ammonium acetate solution, washing with 80% ethanol to remove soluble NH₄⁺, and extracting exchangeable NH₄⁺ with 2 mol L⁻¹ KCl. The concentrations of Ca, Mg, and K were determined by inductively coupled plasma atomic emission spectroscopy (ICP-AES, ULTIMA 2, HORIBA, Japan).

Interpolation of soil sample property values

Because the 5 m x 5 m resolution of the dataset's grid was finer than the 30 m x 30 m resolution of the soil sampling grid, soil property measurements taken from samples pulled near the centroids of the 30 m x 30 m cells had to be spatially interpolated to assign values at the centroids of the 5 m x 5 m cells. Interpolated values were calculated using the 'geoR' package (Ribeiro & Diggle, 2001) of R version 3.6.2 (R Development Core Team, 2019) and applying the empirical best linear unbiased prediction (E-BLUP) method (Lark et al., 2006). Box–Cox transformation (Box & Cox, 1964) was applied prior to geostatistical modeling when the distribution of the observations was highly skewed, and predicted mean values from the E-BLUP were back-transformed. The Matérn covariance function (Webster & Oliver, 2007) and the restricted maximum likelihood estimator were used for estimation of the semi-variogram parameters. The resultant interpolated soil maps are presented in S1. Since covariate values were smoothed using kriging with external drift, they also had interpolation errors (S2). Interpolation errors were evaluated as the coefficients of variance (CV) by dividing the kriging standard deviation by the kriging prediction.

Data analysis

Four machine learning regression models, RF, XGBoost, support vector regression (SVR), and artificial neural network (ANN) were trained with different combinations of covariates to model site-specific yield responses to the fertilizers. RF and SVR were implemented using the Python module 'scikit-learn' (version 1.1.1) (Pedregosa et al., 2011). XGBoost was implemented with xgboost (version 1.5.1) (Chen & Guestrin, 2016). ANN was implemented using the Keras (version 2.9.0) machine learning application programming interface (Chollet, 2015) with the TensorFlow (version 2.9.0) (Abadi et al., 2015) backend.

All 255 possible combinations (i.e., $2^8 - 1$) that can be made from using between one and eight soil properties as covariates were included in the estimations, for each of the four machine learning algorithms, meaning that a total of 1,020 cases were examined. Because the study area was relatively small, spatial differences in weather and other environmental factors were assumed negligible and excluded from the analysis. Of course, the inference space of the experiment should not be assumed to be expandable beyond the field itself. Further research with real-world large-scale experiments is needed to test the robustness of the results reported.

The dataset's 970 observations were randomly split into a 679 observations training dataset and a 291 observations test dataset. Hyperparameters of RF and SVR were determined by grid search with a five-fold cross-validation using the training dataset. This procedure

was repeated three times with different subsets. Then models were retrained with an optimal hyperparameter using the training dataset. For RF, grid searches were performed to optimize the $n_estimators$ (the number of decision trees). For SVR, grid searches were conducted to optimally assign values to parameters C and ε . The ANN architecture involves three hidden layers. The input layer was fed into a rectified linear unit (ReLU) layer with 64 neurons, followed by batch normalization. The batch normalization layer was fed into the ReLU layer with 128 neurons, followed by the ReLU layer with 128 neurons again. Finally, the fully-connected ReLU layer with 128 neurons was fed into an output layer with a linear activation function. According to a preliminary experiment, the model performance of ANN did not largely depend on the architecture, and three to four layers and 32–128 neurons were sufficient to model crop prediction. To avoid over-fitting, early stopping was used to monitor validation loss with a ten epochs of patience. 30% of the training dataset was used to calculate the validation loss for ANN. For both the training and test datasets, model prediction accuracies were evaluated by root mean square error (RMSE) and R^2 .

In this study, model uncertainty refers to the variability in EOIR and gained profit that are predicted from different model algorithms and covariates. To evaluate site-specific uncertainty in decision making for fertilizer application among the selection of algorithms and covariates, site-specific EOIRs were calculated by treating the predicted values from the models as deterministic outcomes. First, site-specific net revenue ($\$/\text{ha}^{-1}$) was defined as,

$$\Delta r_i = p \times y_i - w_{BF} \times BF_i - w_{TF} \times TF_i \quad (1)$$

where $p = \$1.16 \text{ kg}^{-1}$ ($136.8 \text{ JPY kg}^{-1}$) is the price of wheat grain, y_i is the model's predicted wheat grain yield at location i , $w_{BF} = \$1.58 \text{ kg}^{-1}$ ($187.0 \text{ JPY kg}^{-1}$) is the basal fertilizer price, BF_i is the basal fertilizer application rate at location i , $w_{TF} = \$0.60 \text{ kg}^{-1}$ (71.2 JPY kg^{-1}) is the top-dressing fertilizer price, and TF_i is the top-dressing fertilizer application rate at location i . Prices were obtained from the farmer in the corresponding year.

To assess EOIR estimation robustness, fertilizer application rates were optimized by running the model at intervals of 5 kg ha^{-1} (basal fertilizer: $1.25 \text{ kg N ha}^{-1}$, $0.30 \text{ kg P}_2\text{O}_5 \text{ ha}^{-1}$, $0.20 \text{ kg K}_2\text{O ha}^{-1}$; top-dressing fertilizer: $0.85 \text{ kg N ha}^{-1}$, $0.00 \text{ kg P}_2\text{O}_5 \text{ ha}^{-1}$, $0.05 \text{ K}_2\text{O ha}^{-1}$) with other values of soil property covariates unchanged. Application rate ranges were 270 – 540 kg ha^{-1} and 222 – 444 kg ha^{-1} for the basal and top-dressing fertilizers. Rates less than the minimum application rates were not tested because of the limited number of experimental plots receiving no fertilizer. According to the information of a local crop advisory service, N and K were assumed to limit crop yield across the fields because the interpolated values were smaller than the recommended ranges (S1). Therefore, the machine learning models might assess the effect of multiple nutrients, such as N and K on crop yield. To explore the robustness of the site-specific EOIR estimations, mean values and CVs were calculated for each experimental grid. CVs were evaluated by the ratio of the standard deviation to the mean either from all combinations of algorithm and covariate selection ($n=1,020$) or from combinations of covariates for each algorithm ($n=255$) for each experimental grid. Six locations were randomly selected for visualizing the distributions of the basal and top-dressing EOIRs (Fig. 1). Furthermore, gained profits by adopting optimal site-specific fertilization were evaluated by subtracting the net revenue under the uniform conventional rate (i.e., 450 and 370 kg ha^{-1} for the basal and top-dressing fertilizers) from the net revenue under the optimal site-specific fertilization rate.

Results

Results of on-farm precision experiment and yield prediction performance

The relationships between yield and inputs are shown as box plots in Fig. 2. Yield tended to increase with basal fertilizer rates. The median value of yield was extremely low when no top-dressing fertilizer was applied. There were no large differences in yield among the top-dressing application rates from 222 to 444 kg ha⁻¹. This indicates the importance of OFPE for recommending basal fertilizer application rates rather than top-dressing application rates. Furthermore, high variations for each treatment indicate that yield responses can vary substantially, even within a small area.

Both the RMSE and R^2 values of the test dataset indicated that RF had the best yield prediction performance (Fig. 3). Although XGBoost generally showed high prediction accuracies, there were some cases with very high RMSEs ($> 0.5 \text{ t ha}^{-1}$) and low R^2 values (< 0.2). SVR and ANN were not capable of predicting yield values more than approx. 4.0 t ha⁻¹ (Fig. 4). Meanwhile, RF and XGBoost underestimated yield values more than approx. 4.5 t ha⁻¹. This result indicates that the difference in yield underestimation in the ranges of high yield might affect overall yield prediction accuracies. All models failed to predict extremely high yield values ($> 6 \text{ t ha}^{-1}$). This might be due to the lack of important covariates that is related to high yield values. Importantly, the inaccuracies of yield prediction could lead to underestimation of EOIRs in high yield levels.

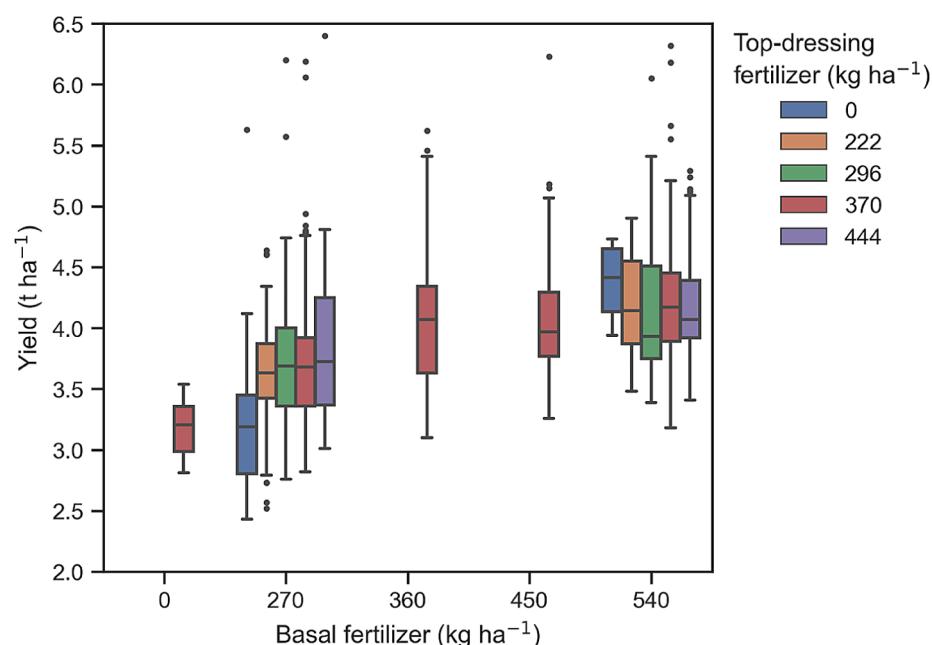


Fig. 2 Box plots of yield for each application rate for basal and top-dressing fertilizers across the fields. Lower and upper box boundaries indicate 25th and 75th percentiles. Lines inside boxes represent medians. The ranges between the lower and upper whiskers are 1.5 times the interquartile range. Filled circles show outliers falling outside 1.5 times the interquartile range

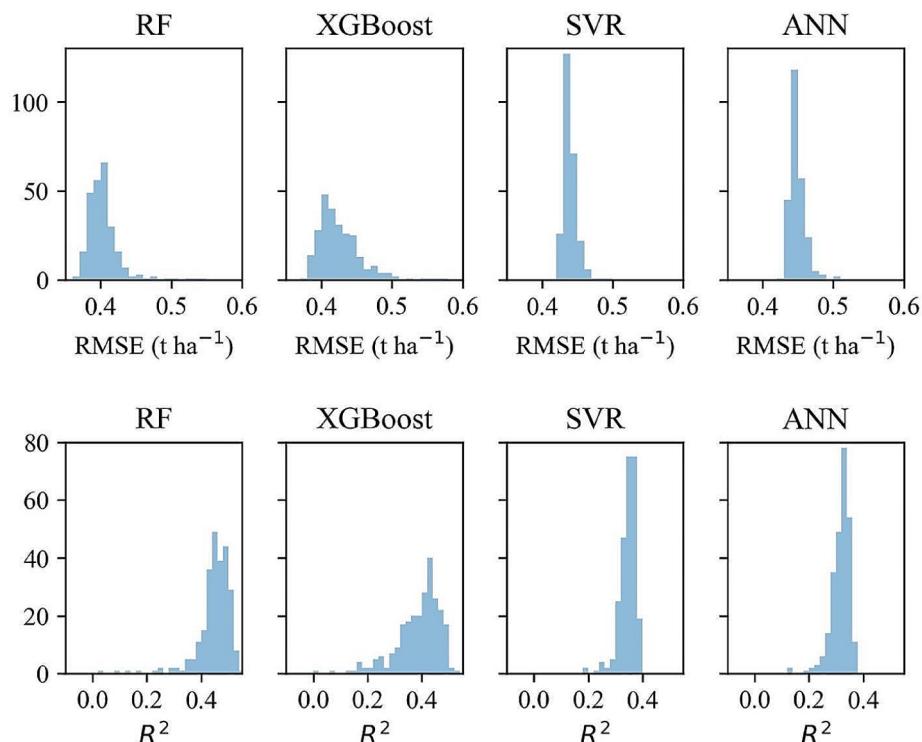


Fig. 3 Histograms of RMSE and R^2 values for all four machine learning models with different combinations of covariates

Variability in EOIRs and gained profits

CV values of the EOIRs were spatially heterogeneous (Fig. 5) for all covariate combinations using all machine learning algorithms, ranging from 13.3 to 31.5% for the basal and from 14.2 to 30.5% for the top-dressing fertilizer. Spatial distributions of EOIR CV values for each machine learning algorithm are shown in Fig. 6. RF was not sensitive to covariate selection, while SVR and ANN were very sensitive to it. However, this study did not attempt to identify which machine learning models were best for generating economic profits. Indeed, doing so is not possible since the true EOIRs, which are needed to validate the model performance, cannot be directly observed. Therefore, it should not be concluded that RF is the best machine learning model for optimizing site-specific input management.

Spatial distributions of mean values of the EOIRs varied greatly among machine learning algorithms (Fig. 7). For both fertilizers, recommended application rates from the tree-based models RF and XGBoost were less spatially heterogeneous and relatively lower than those from SVR and ANN. For example, for basal fertilizer, the estimated EOIRs ranged from 270 to 427 kg ha^{-1} for RF. In contrast, the corresponding range was from 348 to 538 kg ha^{-1} for ANN. This result indicates that crop yield response can vary substantially even within a small area according to the algorithm selection.

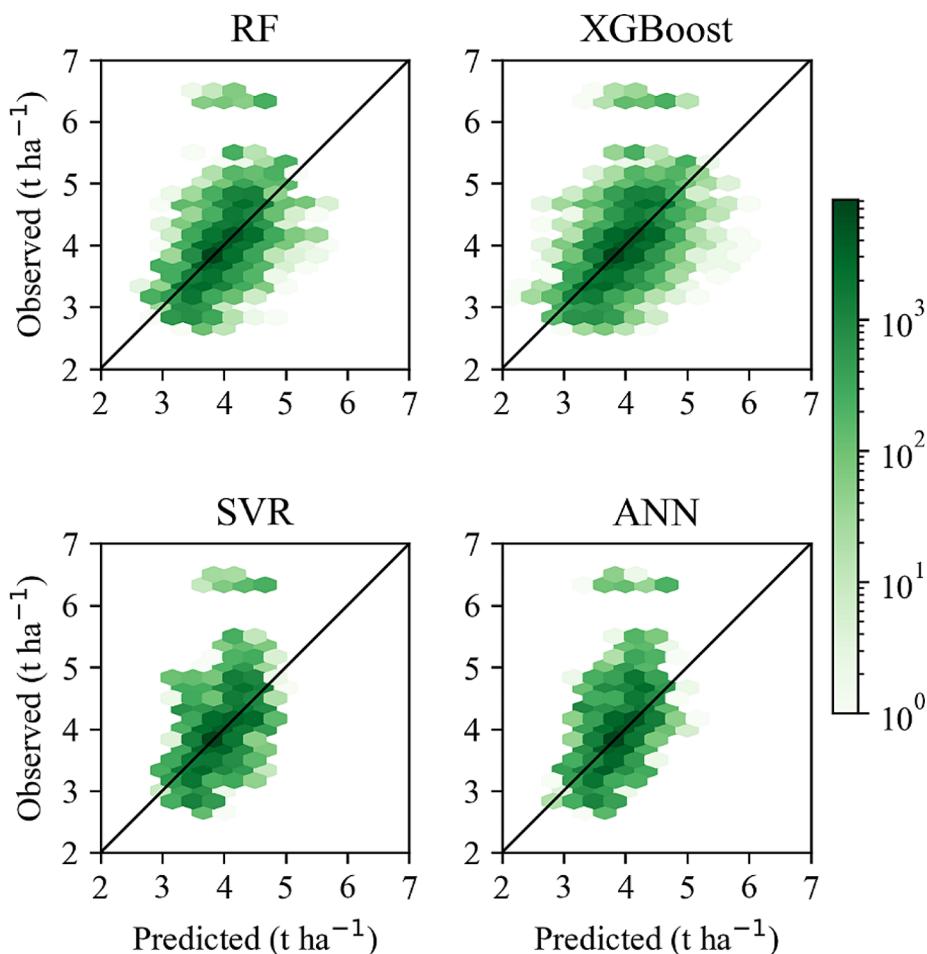


Fig. 4 Density scatter plots of predicted against observed yield for all four machine learning models with different combinations of covariates. The black line indicates the 1:1 reference line

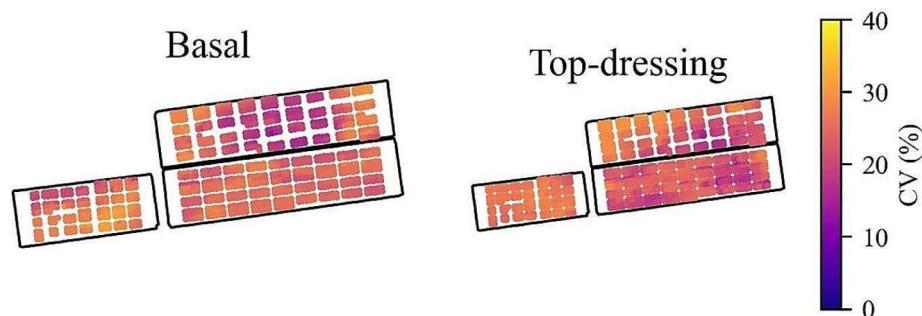


Fig. 5 Spatial distributions EOIR CV values, resultant from the combined effects of algorithm and covariate selection

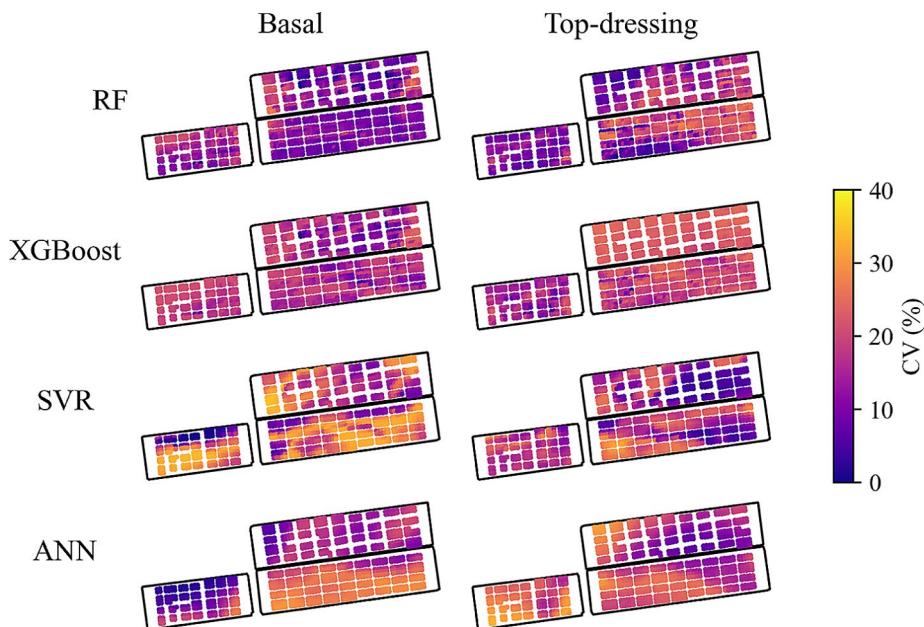


Fig. 6 Spatial distributions of EOIR CV values derived from 255 covariate combination runs per combination of machine learning algorithm and fertilizer type

Figure 8 shows the scatterplots and histograms of estimated EOIRs from all the 1,020 cases at the six locations (See Fig. 1). The 5th- and 95th-percentile borders (indicated by the dashed lines in Fig. 8) show that the EOIRs for both fertilizers had large variations. More specifically, the interval of values containing the central 90% of EOIRs ranged almost from the lowest to highest applied rates for top-dressing fertilizer. Furthermore, a clear unimodal distribution was found only for the basal fertilizer at location (1) In contrast, a bimodal distribution was evident for basal fertilizer at locations 3 and 4 and for top-dressing fertilizer at locations 1 and (2) In these cases, decision makers would be forced to make an extreme choice between the lowest and highest fertilizer application rates, which could ultimately result in a completely different revenue. These results indicate that the EOIR predictions are highly sensitive to algorithm and covariate selection, and that simple averaging methods, such as the ensemble learning approach may not provide reliable recommendations. The high model uncertainty begs the question of whether machine learning approaches can be used effectively for site-specific input management.

The estimated gained profits (relative to the uniform input management) ranged from 150 to 660 \$ ha⁻¹ depending on selected algorithms and covariates (Fig. 9). RF and XGBoost occasionally showed extremely high gained profits. ANN showed lower gained profits and higher uncertainty than other models. SVR showed a lower uncertainty in predicted gained profits. Thus, gained profits predicted by machine learning approach are quite sensitive to algorithm and covariate selection.

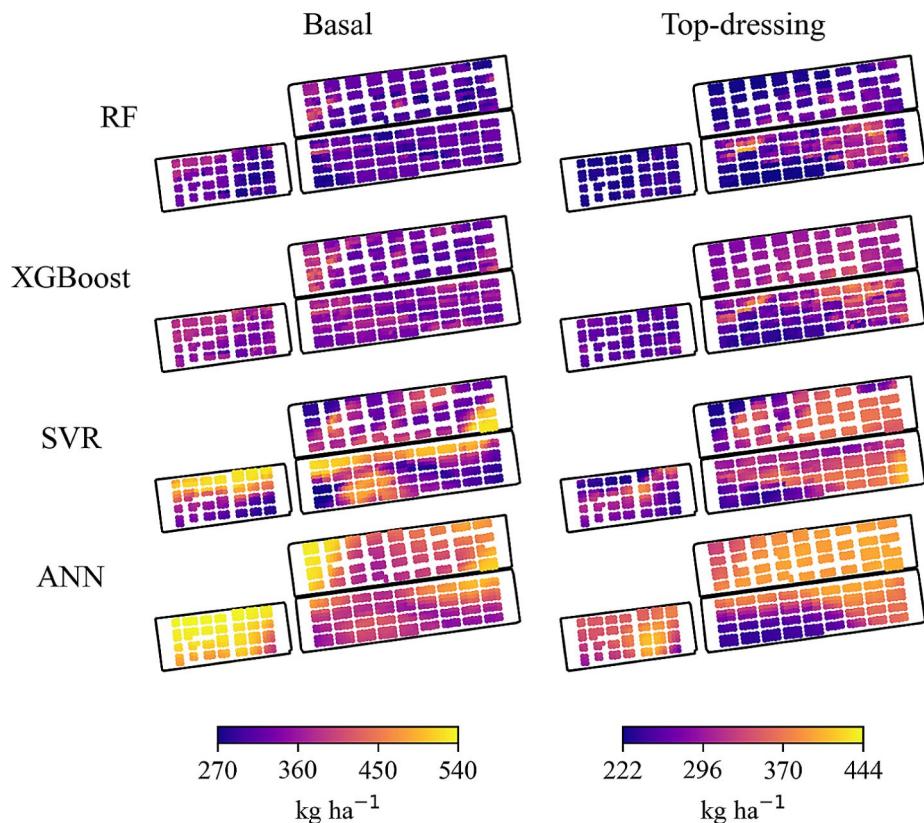


Fig. 7 Spatial distributions EOIRs mean values, for each machine learning algorithm

Discussion

Precise yield prediction and estimation of the causal effects of inputs are essential for the successful implementation of site-specific crop management supported by OFPE. Generally, researchers prefer to select the ‘best’ model based on the metrics of yield prediction accuracies, such as RMSE, mean absolute error, and R^2 values (Barbosa et al., 2020; Wen et al., 2021). But this study has provided valuable information about the impacts of algorithm and covariate selection on fertilization rate management. Overall, machine learning models can predict crop yield well based on RMSE and R^2 values (Figs. 3 and 4). But each machine learning model showed very different EOIR predictions even within a small area (Fig. 7). Predicted site-specific EOIRs were very sensitive to the choice of algorithm and selection of covariates (Figs. 5, 6 and 8). These results highlight that practitioners need a careful consideration of model uncertainty before providing decision makers with fertilizer management recommendations.

The reason that the CV values of EOIRs are not constant in space (Figs. 5 and 6) must be because soil properties are not constant in space (S1), as that is the only input that spatially varies. There are also other factors that influence spatial variability of EOIRs. For instance, spatial variability in established seedlings significantly affect yield (Tanaka et al., 2019),

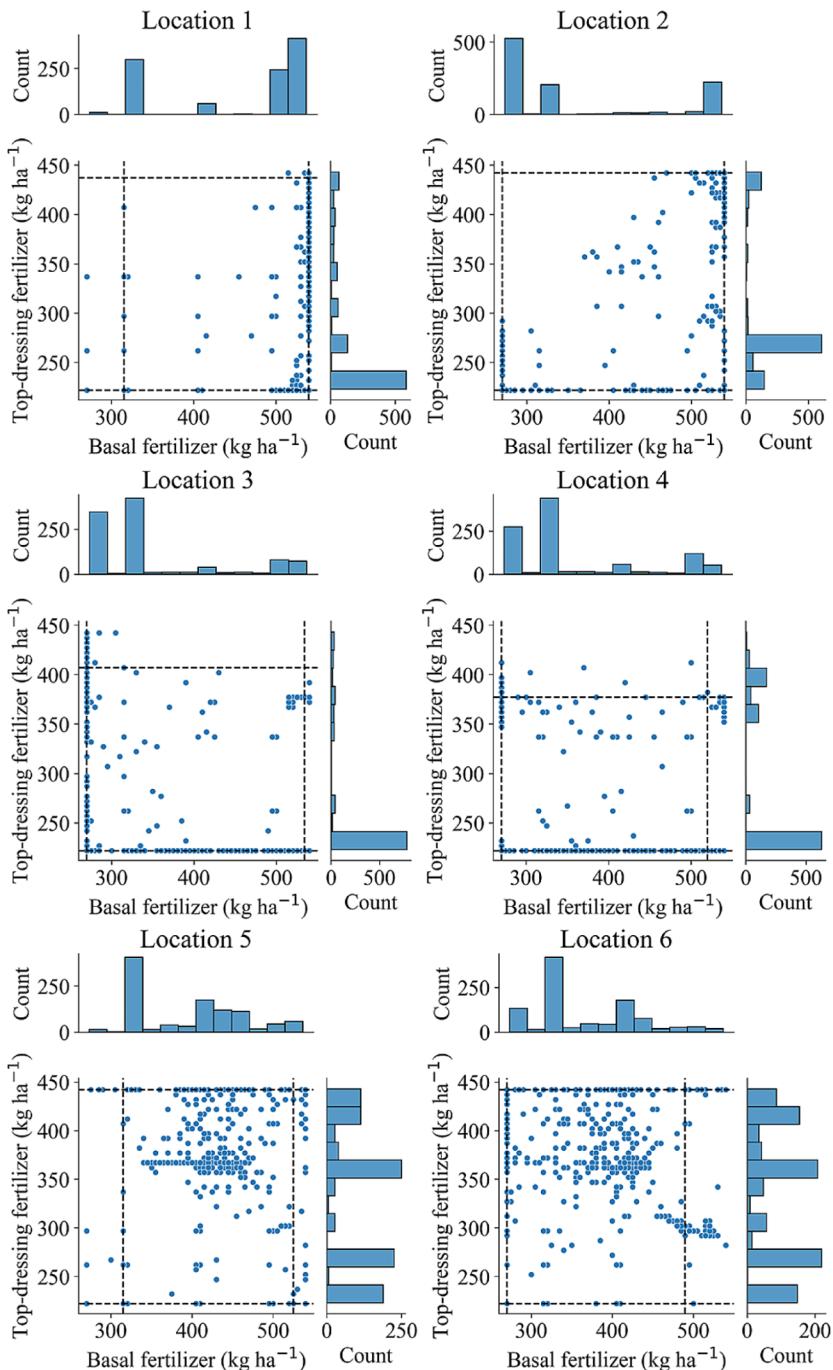


Fig. 8 Scatterplots and histograms of EOIRs of the basal and top-dressing fertilizers in six randomly selected locations (Fig. 1). Data derived from all combinations of algorithms and covariate selection ($n=1,020$). Dashed lines in scatterplots represent borders of the 5th and 95th quantiles

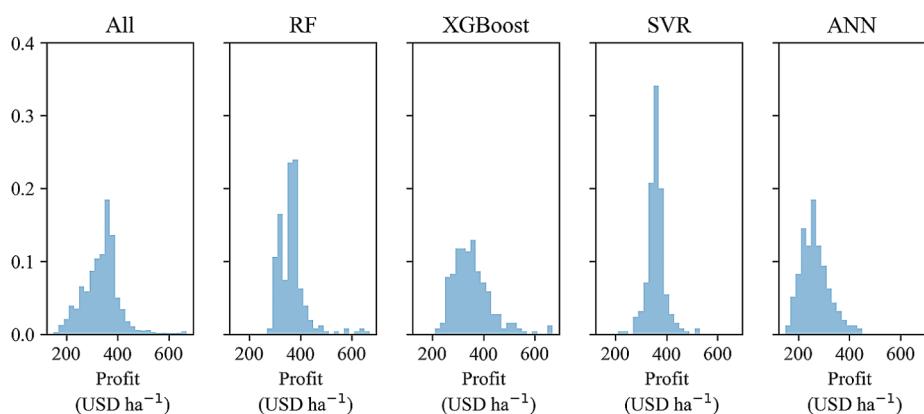


Fig. 9 Histograms of gained profits of entire simulations ($n=1,020$) and each algorithm ($n=255$)

while locations that were trenches before land consolidation had approx. 1.0 t ha^{-1} lower yield than the other parts of the study area, probably due to differences in temporal change in soil moisture conditions (Zhou et al., 2022). However, it is not practical to conduct manual counting of seedlings or to install soil moisture sensors across fields in order to include these factors in the crop yield response modeling. Given the expense of data collection for OFPE, remotely or proximally sensed data might be useful to develop better machine learning models. Although causal factors affecting crop yield could not be identified, either of in-season crop sensing or historical yield map might also enable explaining the pattern of yield response. Given a better management strategy might be established by combining multiple information, it might be necessary to include not only soil data but also in-season and previous proximal/remote sensing data as covariates in crop yield response modeling.

Insufficient consideration on model uncertainty may lead to making highly undesirable input use decisions. The binomial distribution having two peaks at the lowest and highest application rates were found at several locations (Fig. 8), indicating large uncertainty about the effect of fertilizer application on yield. Given little crop yield response to top-dressing fertilizer at the high rate of basal fertilizer (Fig. 2), the highest top-dressing application rates could lead to considerable revenue loss. Although basal fertilizer is a main determinant of yield variation, topdressing fertilizer might be a nuisance covariate for the yield prediction model. Thus, not only outcomes from machine learning but also supplement insights from agronomic knowledge might be important for a sensible EOIR recommendations.

High uncertainty in gained profits was evident depending on model and covariate selection (Fig. 9). For instance, the estimated gained profits ranged from 150 to $660 \text{ \$ ha}^{-1}$. This has major implications on the analysis of cost-benefit performance of PA technology, which in turn affects adoption decision of new technology. Previous studies evaluated either of farming scales or seasons that is necessitated for recovering the purchase cost of variable-rate application equipment (Maine et al. 2010; Tanaka et al., 2023a). In such cases, model uncertainty will affect decision making not only on fertilization but also on the adoption of variable-rate application equipment. ANN had higher uncertainty and lower gained profits. This indicates that ANN provided a more pessimistic and conservative scenario than other models. Therefore, it should be noted that each model tends to provide different uncertainty and recommendation in assessing gained profits. In practice, researchers tend to use only

one model. But this is not without risk, since different models produce different results and lead to different recommendations. Thus, it seems to be reasonable to use ensemble learning approach. However, as discussed in the case of site-specific EOIR predictions, ensemble learning approach might not be able to enhance prediction accuracy for the causal impact derived from the yield prediction models because the predicted EOIR showed a completely different recommended application rate (i.e., either of lowest or highest input rates) (Fig. 8). Practitioners should keep it mind that ensemble learning might be helpful to assess the model uncertainty, but resultant recommendations should not simply be derived from the average.

This study focused only on the effect of model and covariate selection on spatial uncertainty in the EOIRs for fertilizer recommendation. But there are also other sources of uncertainty that deserve attention in future research. For instance, the CV value of exchangeable K was not small (approx. 27%) (S2), indicating substantial uncertainty in model input. While spatial interpolation of soil properties based on geostatistics has been common in producing digital soil maps (Heuvelink & Webster, 2022), it may lead to interpolation errors. If crop yield response models were linear in the soil input, propagation of interpolation errors in digital soil maps could be obtained by simple error propagation rules (Taylor, 1982), but that simple approach is not available for non-linear machine learning models. Due to the high computational cost of Monte Carlo uncertainty propagation methods, this study did not pursue this topic but instead focused on the effect of algorithm and covariate selection on prediction uncertainty. Further study might be needed to explore the effect of machine learning covariate uncertainty on EOIRs and profit prediction.

To train machine learning models, 970 observations with up to eight covariates were used in this study. The data size seemed to be sufficient to achieve the accurate prediction accuracy of site-specific yield levels (Figs. 3 and 4). However, this study used only small-scale on-farm experimental data. The sensitivity of the model uncertainty to the size of the training dataset should be explored in future studies. The differences in predicted EOIRs and profits between model algorithms will be smaller if the training dataset is large. Furthermore, not only spatial but also temporal uncertainty would be essential to consider for better fertilizer recommendation. This study only used annual real on-farm data, while it is difficult to repeat the experiments at the same site for multiple years due to the constraint in the real farm situation. Synthetic data generated by mechanistic models (e.g., the Mitscherlich-Baule function) are not capable of simulating the impact of weather on crop yield. Therefore, a possible solution for the data scarcity problem is to integrate geostatistical simulation and crop simulation model to simulate space-time variability in crop yield (Tanaka et al., 2023b). A surrogate model consisting of a machine learning model trained with synthetic data from a crop simulation model has been proposed to combine biophysical domain knowledge of crop simulation models with data-driven machine learning approaches (Pylianidis et al., 2022). Therefore, integration of Gaussian simulation, crop simulation model, and machine learning approach would provide a chance to assess the space-time model uncertainty.

Conclusions

OFPE are frequently conducted to generate data for the estimation of site-specific crop yield response models. Interest is growing in employing machine learning algorithms to identify spatially heterogeneous and non-linear relationships between agronomic inputs and crop yield. Research has shown that EOIR and gained profit prediction were very sensitive to the selection of machine learning algorithm and covariates. Furthermore, yield response to fertilization could vary substantially from site to site, even in a small area. This might be due to the model uncertainty derived from algorithm and covariate selection. These results highlight the difficulty of providing reliable site-specific input application rate recommendations based on one specific machine learning algorithm and one specific set of covariates. Note that the outcomes of this study were based on small-scale on-farm data conducted in a single season. Further research with data analysis from large-scale OFPEs or synthetic data generated by process-based models should be oriented towards exploring causal inference, thus supporting deriving accurate and robust EOIR predictions.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11119-024-10136-x>.

Acknowledgements The authors wish to thank the farming company ‘Fukue-eino’ for allowing the survey of their fields. This study was supported by a JST ACT-X (JPMJAX20AF), Japan. The study was conducted using the ICP-AES of the Division of Instrument Analysis, Gifu University. This study was funded in part by USDA-NRCS On-Farm Trials Conservation Innovation Grant, “Improving the Economic and Ecological Sustainability of US Crop Production through On-Farm Precision Experimentation,” award number NR213A7500013G021, and by USDA-NIFA Hatch Project 470–362.

Authors contribution Conceptualization, T.S.T.T. and G.B.M.H.; Methodology, T.S.T.T., G.B.M.H., and T.M.; Investigation, T.S.T.T.; Writing – Original Draft, T.S.T.T.; Writing –Review & Editing, T.S.T.T., G.B.M.H., T.M., and D.S.B.; Visualization, T.S.T.T.; Funding Acquisition, T.S.T.T. & D.S.B.; Resources, T.S.T.T.; Supervision, T.S.T.T., G.B.M.H., and D.S.B.

Funding This study was supported by a JST ACT-X (JPMJAX20AF), Japan. This study was funded in part by USDA-NRCS On-Farm Trials Conservation Innovation Grant, “Improving the Economic and Ecological Sustainability of US Crop Production through On-Farm Precision Experimentation,” award number NR213A7500013G021, and by USDA-NIFA Hatch Project 470–362.

Open access funding provided by Aarhus Universitet

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability The codes generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical approval Not available.

Competing interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons

licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. <https://www.tensorflow.org/>. Accessed 7 August 2022.
- Adams, M. L., Cook, S., & Corner, R. (2000). Managing uncertainty in site-specific management: What is the best model? *Precision Agriculture*, 2, 39–54.
- Alesso, C. A., Cipriotti, P. A., Bollero, G. A., & Martin, N. F. (2020). Design of on-farm precision experiments to estimate site-specific crop responses. *Agronomy Journal*, (December 2020), 1–15. <https://doi.org/10.1002/agj2.20572>.
- 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(May 2019), 105197. <https://doi.org/10.1016/j.compag.2019.105197>.
- Boogaard, H., & de Wit, A. (2020). WOFOST: simulation model for quantitative analysis of growth/production of annual crops, (April).
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211–243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>.
- Bullock, D. S., Boerngen, M., Tao, H., Maxwell, B., Luck, J. D., Shiratsuchi, L., et al. (2019). The Data-Intensive Farm Management Project: Changing Agronomic Research through On-Farm Precision Experimentation. *Agronomy Journal*, 111(6), 2736–2746. <https://doi.org/10.2134/agronj2019.03.0165>.
- Cataldo, D. A., Maroon, M., Schrader, L. E., & Youngs, V. L. (1975). Rapid colorimetric determination of nitrate in plant tissue by nitration of salicylic acid. *Communications in Soil Science and Plant Analysis*, 6(1), 71–80. <https://doi.org/10.1080/00103627509366547>.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). New York, NY, USA: ACM. <https://doi.org/10.1145/2939672.2939785>.
- Chollet, F. (2015). and others. Keras. <https://github.com/fchollet/keras>. Accessed 7 August 2022.
- Corner, R., Marinelli, M., & Wright, G. (2008). Error propagation analysis techniques Applied to Precision Agriculture and Environmental models. *Quality aspects in spatial data mining* (pp. 131–145). CRC. <https://doi.org/10.1201/9781420069273.ch11>.
- de Wit, A., Boogaard, H., Fumagalli, D., Janssen, S., Knapen, R., van Kraalingen, D., et al. (2019). 25 years of the WOFOST cropping systems model. *Agricultural Systems*, 168(July 2018), 154–167. <https://doi.org/10.1016/j.agrsy.2018.06.018>.
- R Development Core Team (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.r-project.org>. Accessed 30 March 2022.
- Evans, F. H., Salas, A. R., Rakshit, S., Scanlan, C. A., & Cook, S. E. (2020). Assessment of the use of geographically weighted regression for analysis of large on-farm experiments and implications for practical application. *Agronomy*, 10(11), 1720. <https://doi.org/10.3390/agronomy10111720>.
- Heuvelink, G. B. M. (1998). *Error propagation in Environmental Modelling with GIS*. CRC. <https://doi.org/10.4324/9780203016114>.
- Heuvelink, G. B. M., & Webster, R. (2022). Spatial statistics and soil mapping: A blossoming partnership under pressure. *Spatial Statistics*, 100639. <https://doi.org/10.1016/j.spasta.2022.100639>.
- Heuvelink, G. B. M., Brus, D. J., & Reinds, G. (2010). Accounting for spatial sampling effects in regional uncertainty propagation analysis. *The 9th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Leicester. <https://edepot.wur.nl/160785>. Accessed 19 October 2022.
- Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., et al. (2014). APSIM – Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, 62, 327–350. <https://doi.org/10.1016/j.envsoft.2014.07.009>.
- Hoogenboom, G., Porter, C. H., Boote, K. J., Shelia, V., Wilkens, P. W., Singh, U. (2019). The DSSAT crop modeling ecosystem (pp. 173–216). <https://doi.org/10.19103/AS.2019.0061.10>.
- Inoko, A. (1986). Available nitrogen. In Y. Onikura, et al. (Eds.), *Standard methods of soil analysis and measurement* (pp. 118–121). Hakuyuusha.

- Kakimoto, S., Mieno, T., Tanaka, T. S. T., & Bullock, D. S. (2022). Causal forest approach for site-specific input management via on-farm precision experimentation. *Computers and Electronics in Agriculture*, 199, 107164. <https://doi.org/10.1016/j.compag.2022.107164>.
- Keeney, D. R., & Nelson, D. W. (2015). Nitrogen-Inorganic Forms (pp. 643–698). <https://doi.org/10.2134/agronmonogr9.2.ed.c33>.
- Krause, M. R., Crossman, S., DuMond, T., Lott, R., Swede, J., Arliss, S., et al. (2020). Random forest regression for optimizing variable planting rates for corn and soybean using topographical and soil data. *Agronomy Journal*, 112(6), 5045–5066. <https://doi.org/10.1002/agj2.20442>.
- Lacoste, M., Cook, S., McNee, M., Gale, D., Ingram, J., Bellon-Maurel, V., et al. (2022). On-Farm Experimentation to transform global agriculture. *Nature Food*, 3(1), 11–18. <https://doi.org/10.1038/s43016-021-00424-4>.
- Lark, R. M., Cullis, B. R., & Welham, S. J. (2006). On spatial prediction of soil properties in the presence of a spatial trend: The empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science*, 57, 787–799. <https://doi.org/10.1111/j.1365-2389.2005.00768.x>.
- Maine, N., Lowenberg-DeBoer, J., Nell, W. T., & Alemu, Z. G. (2010). Impact of variable-rate application of nitrogen on yield and profit: A case study from South Africa. *Precision Agriculture*, 11, 448–463. <https://doi.org/10.1007/s11119-009-9139-8>.
- Paccioretti, P., Bruno, C., Gianinni Kurina, F., Córdoba, M., Bullock, D. S., & Balzarini, M. (2021). Statistical models of yield in on-farm precision experimentation. *Agronomy Journal*, 113(6), 4916–4929. <https://doi.org/10.1002/agj2.20833>.
- Pasquel, D., Roux, S., Richetti, J., Cammarano, D., Tisseyre, B., & Taylor, J. A. (2022). A review of methods to evaluate crop model performance at multiple and changing spatial scales. *Precision Agriculture*. <https://doi.org/10.1007/s11119-022-09885-4>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, V., Grisel, O. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>. Accessed 30 March 2022.
- Pylianidis, C., Snow, V., Overweg, H., Osinga, S., Kean, J., & Athanasiadis, I. N. (2022). Simulation-assisted machine learning for operational digital twins. *Environmental Modelling and Software*, 148, <https://doi.org/10.1016/j.envsoft.2021.105274>.
- Ribeiro, P. J., & Diggle, P. J. (2001). The geoR package. *R-NEWS*, 1, 15–18.
- Roques, S. E., Kindred, D. R., Berry, P., & Hellwell, J. (2022). Successful approaches for on-farm experimentation. *Field Crops Research*, 287, 108651. <https://doi.org/10.1016/j.fcr.2022.108651>.
- Saikai, Y., Patel, V., & Mitchell, P. D. (2020). Machine learning for optimizing complex site-specific management. *Computers and Electronics in Agriculture*, 174, <https://doi.org/10.1016/j.compag.2020.105381>.
- Tanaka, T. S. T., Kono, Y., & Matsui, T. (2019). Assessing the spatial variability of winter wheat yield in large-scale paddy fields of Japan using structural equation modelling. *Precision Agriculture*, 19, 751–757. https://doi.org/10.3920/978-90-8686-888-9_93.
- Tanaka, T. S. T., Mieno, T., Tanabe, R., Matsui, T., & Bullock, D. S. (2023a). Toward an effective approach for on-farm experimentation: Lessons learned from a case study of fertilizer application optimization in Japan. *Precision Agriculture*. <https://doi.org/10.1007/s11119-023-10029-5>.
- Tanaka, T. S. T., Yokoyama, Y., Mieno, T., & de Wit, A. (2023b). 27. Synthetic data generation for validating site-specific crop yield response modelling using WOFOST and gaussian geostatistical simulations. *Precision agriculture '23* (pp. 229–235). Wageningen Academic. https://doi.org/10.3920/978-90-8686-947-3_27.
- Taylor, J. R. (1982). *An introduction to Error Analysis: The study of uncertainties in physical measurements*. University Science Books.
- Trevisan, R. G., Bullock, D. S., & Martin, N. F. (2021). Spatial variability of crop responses to agronomic inputs in on-farm precision experimentation. *Precision Agriculture*, 22, 342–363. <https://doi.org/10.1007/s11119-020-09720-8>.
- Truog, E. (1930). The determination of the readily available phosphorus of soils¹. *Agronomy Journal*, 22(10), 874–882. <https://doi.org/10.2134/agronj1930.00021962002200100008x>.
- Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists Second Edition Geostatistics for Environmental Scientists, 2nd Edition*.
- Wen, G., Ma, B. L., Vanasse, A., Caldwell, C. D., Earl, H. J., & Smith, D. L. (2021). Machine learning-based canola yield prediction for site-specific nitrogen recommendations. *Nutrient Cycling in Agroecosystems*, 121(2–3), 241–256. <https://doi.org/10.1007/s10705-021-10170-5>.
- Zhou, X., Heuvelink, G. B. M., Kono, Y., Matsui, T., & Tanaka, T. S. T. (2022). Using linear mixed-effects modeling to evaluate the impact of edaphic factors on spatial variation in winter wheat grain yield in Japanese consolidated paddy fields. *European Journal of Agronomy*, 133, 126447. <https://doi.org/10.1016/j.eja.2021.126447>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Takashi S. T. Tanaka^{1,2,7}  · Gerard B. M. Heuvelink^{3,4} · Taro Mieno⁵ · David S. Bullock⁶

 Takashi S. T. Tanaka
takashi@agro.au.dk

¹ Faculty of Applied Biological Sciences, Gifu University, Yanagido, Gifu 5011193, Japan

² Artificial Intelligence Advanced Research Center, Gifu University, Yanagido, Gifu 5011193, Japan

³ ISRIC-World Soil Information, Wageningen, The Netherlands

⁴ Soil Geography and Landscape Group, Wageningen University, Wageningen, The Netherlands

⁵ Department of Agricultural Economics, University of Nebraska-Lincoln, Lincoln, NE 68583 0922, USA

⁶ Agricultural and Consumer Economics, University of Illinois, 326 Mumford Hall, 1301 W. Gregory, Urbana, IL 61801, USA

⁷ Present address: Department of Agroecology, Faculty of Technical Sciences, Aarhus University, Forsøgsvej 1, Slagelse 4200, Denmark