



Assessing performance of empirical models for forecasting crop responses to variable fertilizer rates using on-farm precision experimentation

Paul B. Hegedus^{1,3} · Bruce D. Maxwell^{1,4} · Taro Mieno^{2,5}

Accepted: 5 October 2022
© The Author(s) 2022

Abstract

Data-driven decision making in agriculture can be augmented by utilizing the data gathered from precision agriculture technologies to make the most informed decisions that consider spatiotemporal specificity. Decision support systems utilize underlying models of crop responses to generate management recommendations, yet there is uncertainty in the literature on the best model forms to characterize crop responses to agricultural inputs likely due for the most part to the variability in crop responses to input rates between fields and across years. Seven fields with at least three years of on-farm experimentation, in which nitrogen fertilizer rates were varied across the fields, were used to compare the ability of five different model types to forecast crop responses and net-returns in a year unseen by the model. All five model types were fit for each field using all permutations of the three years of data where two years were used for training and a third was held out to represent a “future” year. The five models tested were a frequentist based non-linear sigmoid function, a generalized additive model, a non-linear Bayesian regression model, a Bayesian multiple linear regression model and a random forest regression model. The random forest regression typically resulted in the most accurate forecasts of crop responses and net-returns across most fields. However, in some cases the model type that produced the most accurate forecast of grain yield was not the same as the model producing the most accurate forecast of grain protein concentration. Models performed best when the data used for training models was collected from years with similar weather conditions to the forecasted year. The results are important to developers of decision support tools because the underlying models used to simulate management outcomes and calculate net-returns need to be selected with consideration for the spatiotemporal specificity of the data available.

Keywords Decision support · On-farm experimentation · Model assessment · Crop modeling · Predictive modeling

✉ Paul B. Hegedus
paulhegedus@montana.edu

Extended author information available on the last page of the article

Introduction

Farmers and scientists alike have recognized that crop responses to inputs vary between fields and across time, and even within fields (Hegedus & Maxwell, 2022). The digital data revolution in agriculture opens the door for providing insights and understanding of the ecological complexity of agroecosystems (Basso & Antle, 2020). Advances in remote sensing and the advent of big data gathered by precision agriculture technology allow farmers, managing ever greater land areas (Fig. 1), the ability to account for in-field variability and provide information that is otherwise infeasible for farmers to gather from day-to-day observations (Hatfield et al., 2020). Precision agriculture data makes information previously inaccessible to farmers accessible, enabling its use in decision making (Carolan, 2017). Although most farmers use precision agriculture technology, such as global navigation satellite systems (GNSS) and autosteer, more advanced technology and data utilization such as site-specific application of agricultural inputs have received far less adoption, stagnating at around 10% in wheat systems (Schimmelpfennig & Lowenberg-DeBoer, 2020). Large farms are adopting precision agriculture technology at a faster rate than small farms (Schimmelpfennig & Lowenberg-DeBoer, 2020). This introduces the opportunity for policy incentives to encourage small farms to adopt data driven agriculture to address sustainability and efficiency issues related to the application of agricultural inputs. Even using precision agriculture data such as yield and soil maps increases the efficiency of agricultural inputs by 7.2–8.5% (McFadden et al., 2021). A first step towards increasing sustainability on farms that have the capacity for it is precision agriculture technology. Precision agriculture is a data-driven site-specific application of agricultural inputs that increase the efficiency of chemical usage and minimize potential degradation of ecological resources (Gebbers & Adamchuck, 2010).

Although the farm implement industry has wholly embraced big data, there still remains a disconnect between the collection and aggregation of data and utilization into decision making (Carolan, 2017). Agricultural businesses, academia, NGO's and governments all recognize the issues faced in agriculture and are working toward providing solutions to food security in a sustainable manner, yet the approaches and methods for achieving these goals vary (West et al., 2014). The need for decision support systems that integrate data from farms and provide information for farmers to make decisions

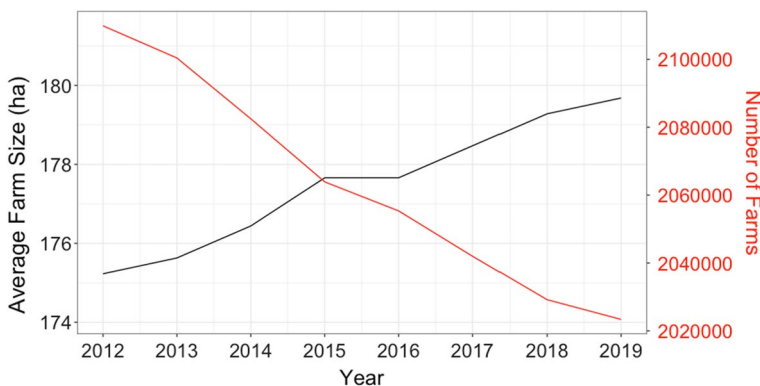


Fig. 1 Average farm sizes in the United States (left y-axis; black line) and number of farms (right y-axis; red line) from 2012 to 2019. Data from USDA NASS, 2019

has long been recognized (McBratney et al., 2005b; Wajid et al., 2021). However, even with industry led efforts into decision support systems and software, adoption of precision agriculture for efficient application of agricultural inputs remains limited (Schimelpennig & Lowenberg-DeBoer, 2020). Sole reliance on industry to provide the technological advances for generating prescription maps that increase the efficiency of agricultural inputs leads to yet another cost associated with farming, wherein producers are required to purchase software from companies that have obfuscated the process in which they generate maps for farmers (Gardner et al., 2021). Thus, farmers are removed from the decision-making matrix and the cost of accessing data intensive prescriptions reduce their net-returns. Even academic techniques for characterizing crop responses remain shrouded behind “black box” methods that obscure the way in which decisions are made, most commonly by machine learning and artificial intelligence approaches.

To make informed management decisions at the time when input decisions need to be made requires forecasting the response of future crops to inputs by simulating across unknown weather and economic conditions. The simulations not only enable farm managers to improve efficiencies of agricultural inputs, but enables understanding of the complexity and nonintuitive outcomes in their agroecological systems (Houlahan et al., 2017). Depending on the forecast model used, the difference in model selection at the field scale has important ramifications. Variation in model selection for enhancing agricultural input efficiency can increase yields by 5.6 to 11.9%, (McFadden et al., 2021). Additionally, de Lara et al. (personal correspondence) state that management recommendations such as for nitrogen (N) fertilizer, vary between model types, highlighting the importance of understanding the implications and differences between models for decision making.

Techniques for characterizing crop responses vary widely, and include detailed process-based crop models (Basso et al., 2011; Sela et al., 2016), parametric and non-parametric frequentist models (Bolton & Friedl, 2013; Johnson, 2014), traditional agronomic based models (Mueller et al., 2014), Bayesian approaches (Lawrence et al., 2015; McFadden et al., 2017), and machine learning methods (Peerlinck et al., 2018, 2019) all utilized in the literature and across industry. This paper addresses the knowledge gap around crop forecast model performance (Thöle et al., 2013) and, as a proof of concept, characterizes the response of rain-fed winter wheat to variable rates of N fertilizer for forecasting crop yield and protein content. A range of approaches to characterize responses (yield and grain protein concentration) to variable N fertilizer management for forecasting, and how this varies across space and time, were assessed. Additionally, how the different models influenced forecasts of net-returns (return on N fertilizer investment) was explored. The objectives were to compare parametric, semi-parametric, machine learning and Bayesian models to (1) assess the accuracy of winter wheat crop response (grain yield and grain protein concentration) forecasts in a hold-out year (representing a future year) from five different types of field specific models, (2) to investigate the influence that the sequence of years used for training field specific models has on forecast accuracy, and (3) to evaluate the efficacy of the five types of models to predict observed net-returns in a hold-out year.

Variable rate application technology along with crop yield and quality monitors has allowed on-field experimentation to produce data that can refine input recommendations specific to each field (Hegedus & Maxwell, 2022). Consistently accurate models to forecast crop responses to agricultural inputs and the return on investment of the inputs are essential for building trusted input recommendations.

Methods

Decision support systems require models for forecasting the response of crops to variable agricultural inputs to assess management decisions when the conditions in the future year are unknown at the decision point in time. Thus, the ability of models to predict crop responses in datasets that have been held out from the model training process are vital for forecasting crop responses and expected net-return in future years.

The two objectives of first identifying the model that best forecasts crop responses and then identifying the influence of training data (years) on making accurate forecasts of crop responses are linked, so they cannot be tested independently or in a logical order. This is because it is quite possible that different forecast models and different data configurations perform differently on different fields. It was assumed a priori that the field-specific model type that performed best for a given field would be identified by averaging forecast performance across data configurations. The influence of the data configuration of training data on prediction performance for each model type was assessed by comparing the mean root mean square error (RMSE) across fields from each model type and data configuration.

Study location

The fields used in this study were from three farms geographically distributed across Montana USA that have been participating in the On-Farm Precision Experiments (OFPE) project (<https://sites.google.com/site/ofpeframework/home>) at Montana State University for the past five years (Fig. 2). All fields selected were rain-fed conventional crop-fallow winter wheat systems that have available data from three years of variable rate N fertilizer rate experiments (Table 1). Variable rates were randomly applied, stratified by yield and grain protein concentration. Experiments in the first two years contained experimental plots where no N was applied. However, on farm B in 2021, an agreement with the farmer meant that the lowest experimental N fertilizer rate (N-rate) would be 56 kg ha⁻¹ instead of the statistically preferred control of zero.

Data collection

Grain yield (kg ha⁻¹) was collected every three seconds directly from farmer's on-combine yield monitors that were calibrated prior to the harvest of each experimental field. Grain protein concentration (%) was collected every ten seconds with a combine mounted Crop-Scan 3000H (Next Instruments, Condell Park, Australia) near infrared monitor (Clancy, 2019). Thus, the response metrics were gathered on different scales, and initially collected as two datasets. As-applied N fertilizer data were collected directly from each farmer's applicator equipment and all response and experimental data were cleaned using standard agronomic practices (Blackmore & Moore, 1999; Sudduth & Drummond, 2007; Sudduth et al., 2012). As-applied N fertilizer rates were georeferenced to the locations of each yield and protein observation in their respective datasets. Additional remotely sensed satellite covariate data were collected from the Google Earth Engine data repository (Gorelick et al., 2017) and georeferenced to the locations of each grain yield and protein concentration observation (Table 2).

Topographic data that do not change over time, such as, slope, elevation, topographic position index (TPI) and soil characteristics were derived from datasets collected in 2015, an arbitrary year across the years collected (Table 2). Temporally varying covariate data

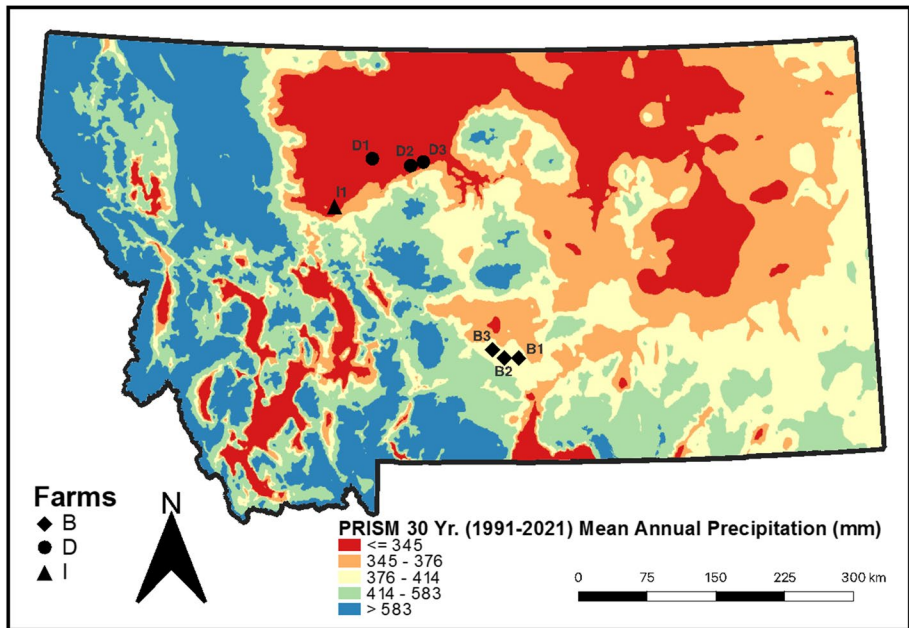


Fig. 2 Map of the seven OFPE field locations for the three selected farmer collaborators with Montana State University. Mean annual precipitation gathered from the University of Oregon's PRISM program across 30 years (1991–2021).

Table 1 Crop histories, field sizes and years in VRA treatment for each field

Farm	Field	Field size (ha)	Crop history: 2014/2015/2016/2017/2018/2019/2020/2021	Years N rate treatment
B	B1	79	SF/WW/CF/WW/CF/WW/CF/WW	2017, 2019, 2021
	B2	94	WW/CF/WW/CF/SF/WW/CF/WW	2016, 2019, 2021
	B3	64	SW/CF/WW/CF/WW/CF/WW/CF	2016, 2018, 2020
D	D1	46	CF/WW/CF/WW/CF/WW/CF/WW	2017, 2019, 2021
	D2	48	WW/SW/WW/CF/WW/CF/WW/CF	2016, 2018, 2020
	D3	20	SW/SW/WW/CF/WW/CF/WW/CF	2016, 2018, 2020
I	I1	94	SW/CF/WW/CF/WW/CF/WW/CF	2016, 2018, 2020

WW winter wheat, CF chemical fallow, SW spring wheat, SF safflower

were collected only through the year prior to the observed harvest to simulate realistic data constraints farmers would have when forecasting future crop responses. Normalized difference vegetation index (NDVI) and normalized difference water index (NDWI) were collected up to December 31st of the prior year and two years prior harvest. For example, observed harvest data from 2021 would be modeled with the maximum vegetation index values (within field pixels) between 01/01/2020 and 12/31/2021 (py=previous year) and between 01/01/2019 and 12/31/2020 (2py=two years prior). Growing degree day (GDD) data were collected across the same timeframe for only the prior year (py). Following the

Table 2 Table of covariate data types gathered from Google Earth Engine to enrich the crop yield and protein datasets gathered from on farms

Data type	Data sources	Resolution	Years collected	Description
NDVI	Landsat 5/7/8	30 m	L5: 1999–2011 L7: 2012–2013 L8: 2014—present	Landsat is an ongoing USGS and NASA collaboration Bands (NIR, red) L5/L7: B4 and B3 L8: B5 and B4
NDWI	Landsat 5/7/8	30 m	L5: 1999–2011 L7: 2012–2013 L8: 2014—present	Landsat is an ongoing USGS and NASA collaboration Bands (NIR, red) L5/L7: B2 and B4 L8: B2 and B5
Elevation	USGS NED	~ 10 m (1/3 arc second), ~ 23 m (3/4 arc second)	1999–present	USGS National Elevation Dataset. Measured in meters
Aspect	USGS NED	~ 10 m (1/3 arc second), 30 m	1999–present	Direction the surface faces, function of neighboring elevations, in radians. Also calculated for each E/W and N/S direction as cosine and sine
Slope	USGS NED	~ 10 m (1/3 arc second), 30 m	1999–present	Rate of change of height from neighboring cells, in degrees. Measured in degrees
TPI	USGS NED	~ 10 m (1/3 arc second), 30 m	1999–present	Measure of divots and low spots as a function of neighboring elevation
Precipitation	DaymetV3	1 km	1999–present	Estimates from the NASA Oak Ridge National Laboratory (ORNL). Measured in mm
GDD	DaymetV3	1 km	1999–present	Estimates from the NASA Oak Ridge National Laboratory (ORNL)
Bulk density	OpenLandMap	250 m	1999–present	Soil bulk density (fine earth) 10×kg/m ³ averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m)
Clay content	OpenLandMap	250 m	1999–present	Clay content in % (kg/kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m)
Sand content	OpenLandMap	250 m	1999–present	Sand content in % (kg/kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m)
pH (phw)	OpenLandMap	250 m	1999–present	Soil pH in H ₂ O averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m)
Water content	OpenLandMap	250 m	1999–present	Soil water content (volumetric %) for 33 kPa and 1500 kPa suction predicted and averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m)
Carbon content	OpenLandMap	250 m	1999–present	Soil organic carbon content in ×5 g/kg averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m)

convention of water years, precipitation data was collected up to October 31st of the prior year. For example, observed harvest data from 2021 would be modeled with the sum of precipitation between 11/01/2019 and 10/31/2020 (py = previous year).

The influence of the magnitude of covariate measurements on model fitting was reduced by centering covariate data, where centering is the subtraction of each covariate observation from the mean of the covariate observations. Soil characteristic and topographic data were centered because there was no expectation that the distribution of the predictor would vary across years. Temporally varying data such as precipitation, GDD and vegetation index (NDVI, NDWI) data were not centered because the distribution of these predictors could vary for a field in different years, meaning that the mean used for centering these data would be different than the mean required for centering in a new dataset from a year not used or not like a year used to train the model.

Experiment design & model assessment

To make decision support systems useful for farmers, the crop response models that they use must be accurate for forecasting future crop responses and allow for quantification of the uncertainty in the forecast. To mimic reality, in which models need to forecast crop responses in a future year not represented in training data, a k-fold cross validation type design was used. Harvest, as applied fertilizer and covariate data for each of the three years that every field underwent N fertilizer experimentation were collected as described above. The crop-fallow system practiced in all the fields means harvest data from experimental years are available every other year on a 2016, 2018, 2020 or 2017, 2019, 2021 schedule. For each field, three split cases with different data configurations were created where two years of data were used for training while data from one year were held out as a test set to evaluate the ability of the model to forecast crop responses in a year with observations unseen by the model (Fig. 3).

Split configuration A represents a scenario where the two oldest years of experimental data (2016 or 2017 and 2018 or 2019, respectively) are used to train models that predicted crop responses in the most recent year (2020 or 2021). Split configuration B uses the first (2016 or 2017) and most recent years (2020 or 2021) of experimental data for training models that predicted the crop responses in the intermediary year (2018 or 2019). Split configuration C uses the most recent year (2020 or 2021) and the intermediary year (2018 or 2019) for training models that predicted crop responses in the first year of experimentation (2016 or 2017).

For each field and crop response (grain yield and grain protein concentration), the design matrix with the three data splits in Fig. 3 was applied to each of the five model

Split	Year(s)		
A	2016/2017	2018/2019	2020/2021
B	2016/2017	2018/2019	2020/2021
C	2016/2017	2018/2019	2020/2021

Train
 Test

Fig. 3 Design matrix for data configurations applied to each field for all 5 model types. Orange cells represent the years used for training in each case and blue cells represent the year used for testing in each case

types. For each of the 210 models (three-split case design across seven fields, with five model types, for two crop responses), the root-mean square error (RMSE) was used as the metric of performance. For each of the models tested, RMSE was calculated from the observed crop responses in the test year (yield or protein) and forecasts (predictions) of crop responses in the test year made by the model fit on the training years.

Models

Five parametric to non-parametric models that range from using frequentist, Bayesian and machine learning approaches were fit to crop responses on each field. The models used were a modified version of a non linear sigmoid model, generalized additive model (GAM), linear and non-linear Bayesian models, and a random forest regression model. For each field, each model type was fit individually using grain yield (kg ha^{-1}) or grain protein concentration (%) as response variables with variable as-applied N fertilizer rates (kg ha^{-1}) and covariates as predictor variables (Table 3). All analysis was performed in R (R Core Team, 2021), where the *OFPE* package was used for all data management and analysis (Hegedus, 2020). All covariate data in training datasets were tested to avoid aliasing and covariance between predictors. No interactions were included in the models, except for in the Bayesian non-linear regression model, because analysis outline in this paper serves as the starting point for identifying the most appropriate model form for crop responses in each field. After identification of the most appropriate model form, further testing and evaluation should be performed where model variations, refinements and further feature selection are tested to generate the most accurate field specific forecasting model.

Modified non-linear sigmoid model (beta function)

The first model type was parametric and modified from a non-linear sigmoidal model called the beta function (Yin et al., 2003). Following the literature, a model that assumes an asymptotic nature of crop response curves to additional N fertilizer was tested here (Anselin et al., 2004; Reynolds et al., 2021; Yin et al., 2003). Initial tests compared the beta function to multiple forms of non-linear models including hyperbolic, Richards, Gompertz, Weibull and logistic forms. Selection of the beta function (Yin et al., 2003) over these other non-linear models was based on three factors; (1) the beta function yielded the lowest mean RMSE across fields for both yield and protein (Table 4), (2) the beta function required the least locking of parameters to facilitate convergence, and (3) the beta function captured the asymptotic response and a downturn of responses at high experimental N fertilizer rates.

The beta model was modified by incorporating an intercept term to represent the response of the crop at zero N-rate. Additionally, a Gaussian spatial correlation structure on the Universal Transverse Mercator (UTM) coordinates of observations to account for the spatial dependence of responses was included (De Bastiani et al., 2015; Xia et al., 2008). The beta function was fit using generalized non-linear least squares regression through the *nlme* package in R (Pinheiro et al., 2021). Bottom-up feature selection was performed where predictors that did not contribute to a decrease in Akaike information criterion (AIC) by two units were omitted from the model. As crop responses differed across fields, not all beta functions fit for yield or protein for a given field and set of training years have the same predictors. All final models for each field were selected from the following beta function used for yield (kg ha^{-1}) or grain protein concentration (%) forecasting, with

Table 3 Identifiers for covariates in the models below, the covariate's data type and a description of the covariate

Covariate ID	Data type	Description
aspect _{cos}	Aspect	Aspect in the E/W direction
aspect _{sin}	Aspect	Aspect in the N/S direction
slope	Slope	Slope in degrees
elev	Elevation	Elevation in meters
tpi	TPI	Topographic Position Index
prec _{py}	Precipitation	Precipitation summed from the water year two years prior to observed data (mm)
gdd _{py}	GDD	Growing degree days summed from the year prior
ndvi _{1py}	NDVI	Maximum Normalized Difference Vegetation Index from across the prior year
ndvi _{2py}	NDVI	Maximum Normalized Difference Vegetation Index from across two years prior
ndwi _{1py}	NDWI	Maximum Normalized Difference Water Index from across the prior year
ndwi _{2py}	NDWI	Maximum Normalized Difference Water Index from across two years prior
bulkdensity	Bulk Density	Soil bulk density (fine earth) 10×kg/m3 averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m)
claycontent	Clay Content	Clay content in % (kg/kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m)
phw	pH of Water	Soil pH in H ₂ O averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m)
watercontent	Water Content	Soil water content (volumetric %) for 33 kPa and 1500 kPa suctions predicted and averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m)
carboncontent	Carbon Content	Soil organic carbon content in ×5 g/kg averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m)

Table 4 Results of the mean RMSE across the seven fields of yield and protein from each of the non-linear models tested with a randomly split training dataset with 70% of observations across three years for model fitting and a hold-out dataset with the other 30% of observations for predicting responses with new data and calculating RMSE

Model	Mean RMSE for Yield (kg ha ⁻¹)	Mean RMSE for Protein (%)
Beta function	1193.13	1.62
Gompertz	1219.14	1.67
Hyperbolic	1216.70	1.68
Logistic 1	1205.73	1.63
Logistic 2	1194.28	1.62
Richards	1205.27	1.63
Weibull 1	1199.54	1.64
Weibull 2	1252.27	1.76

a Gaussian process spatial correlation structure, and the predictors that influence the intercept term (α);

$$R \sim \alpha + \left[(\beta - \alpha) * \left(1 + \frac{\delta_2 - N}{\delta_2 - \delta_1} \right) * \left(\frac{N}{\delta_2} \right)^{\frac{\delta_2}{\delta_2 - \delta_1}} \right] + \text{Gaussian}(x, y) + \epsilon \quad (1)$$

where R is yield (kg ha⁻¹) or grain protein concentration (%), α is the fit minimum yield at a zero N-rate, β is the fit maximum yield (asymptote) at high N-rates, N is the as-applied N-rate in kg ha⁻¹, δ_2 is the fit N-rate where β is approached, δ_1 is the fit N-rate at the inflection point (center of the upturn of the curve), $\text{Gaussian}(x, y)$ is the Gaussian process spatial correlation structure, and ϵ is random error assumed to be normal and independent $N(0, \sigma^2)$. The intercept term, α , was assumed to be influenced by all topographic variables and information from prior years;

$$\begin{aligned} \alpha \sim & \alpha_0 + \alpha_1 \text{aspect}_{\cos} X \text{aspect}_{\sin} + \alpha_2 \text{slope} + \alpha_3 \text{elev} + \alpha_4 \text{tpi} + \alpha_5 \text{prec}_{py} \\ & + \alpha_6 \text{gdd}_{py} + \alpha_7 \text{ndvi}_{py} + \alpha_8 \text{ndvi}_{2py} + \alpha_9 \text{ndwi}_{py} + \alpha_{10} \text{ndwi}_{2py} \\ & + \alpha_{11} \text{bulkdensity} + \alpha_{12} \text{claycontent} + \alpha_{13} \text{phw} + \alpha_{14} \text{watercontent} \\ & + \alpha_{15} \text{carboncontent} \end{aligned} \quad (2)$$

where $\alpha_0 - \alpha_{15}$ were coefficients for the intercept and covariate data listed in Table 3.

Generalized additive model (GAM)

Staying in the non-linear realm but relaxing assumptions on the shape of the model for crop responses, a generalized additive model (GAM) was tested. Generalized additive models are a blend of semi-parametric generalized linear models with additive smoothing terms for modeling non-linear relationships between covariates and the response (Hastie & Tibshirani, 1986; Pinilla & Negrín, 2021; Wood, 2017). Despite their flexibility, the use of GAMs for modeling crop responses has been limited (Chen et al., 2019; Joshi et al., 2021). The GAMs for grain yield and grain protein concentration used gamma family distributions and thin plate splines with shrinkage for the basis functions on most of the covariates in Table 3. A gamma family distribution was selected to ensure that estimates and predictions were above zero in accordance with the reality of grain yield and grain protein concentration measurements. Shrinkage allowed estimated

degrees of freedom to fall to zero, combining the model fitting and feature selection process. The only covariates that did not use a thin plate basis function were the spatial UTM coordinates, which were included as smoothing terms with a Gaussian process basis function to account for spatial autocorrelation (Guisan et al., 2002; Zuur et al., 2012). To account for non-constant variance found in initial model fits, a log link function was used. Models were fit using the *mgcv* package in R (Wood, 2003; Wood et al., 2016). As crop responses differed across fields, not all GAMs for yield or protein for a given field and set of training years have the same predictors but selected from the following initial GAM with a gamma family distribution and log link function;

$$\begin{aligned} \text{Gamma}(R, \text{link} = \log) \sim & I + f_1(x, y) + f_2(\text{aspect}_{\cos}, \text{aspect}_{\sin}) + f_3(N) + f_4(\text{slope}) + f_5(\text{elev}) \\ & + f_6(\text{tpi}) + f_7(\text{prec}_{py}) + f_8(\text{gdd}_{py}) + f_9(\text{ndvi}_{py}) + f_{10}(\text{ndvi}_{2py}) \\ & + f_{11}(\text{ndwi}_{py}) + f_{12}(\text{ndwi}_{2py}) + f_{13}(\text{bulkdensity}) + f_{14}(\text{claycontent}) \\ & + f_{15}(\text{phw}) + f_{16}(\text{watercontent}) + f_{17}(\text{carboncontent}) + \epsilon \end{aligned} \quad (3)$$

where R is the expected yield (kg ha^{-1}) or grain protein concentration (%) with a gamma log link function, I is the intercept, f_1 is the tensor product of the x and y coordinate (x, y) with a Gaussian process basis function, f_2 is the tensor product of centered aspect in the E/W direction ($\text{aspect}_{\cos}$) and aspect in the N/S direction ($\text{aspect}_{\sin}$) with a thin plate shrinkage spline basis function, f_3 – f_{17} are smoothing functions fit with maximum likelihood and thin plate shrinkage splines as the basis function for covariates listed in Table 3, and ϵ is the random error, assumed to be normal and independent $N(0, \sigma^2)$.

Non-linear Bayesian regression model (Bayes NLR)

Two Bayesian models were also tested. Bayesian approaches are commonly applied for modeling crop responses (Huang et al., 2017; Nandram et al., 2014; Ramsey, 2020) but are also used for models that influence policy such as yield gap and potential (Prost et al., 2008) and commodity prices (Drachal, 2019). The first Bayesian model used was adapted from Lawrence et al., (2015). Two modifications were made to fit this dataset and scenario, including replacing electrical conductivity data with clay content, based on the correlation between the two properties (McBratney et al., 2005a). Second, the precipitation term used as a predictor for crop responses was constrained in this model to precipitation summed across the previous year, as only this data would be available when forecasting is required to make management decisions in an upcoming year. Finally, because this data represented scattered points across a field rather than a lattice grid structure as in Lawrence et al., (2015), simultaneous spatial auto regression (SAR) was used rather than conditional auto regression (Ver Hoef et al., 2017; Wall, 2004). The base of this model was identical to Lawrence et al., (2015) where grain yield and protein were assumed to follow a normal distribution;

$$R \sim N(\mu, \sigma_e^2) \quad (4)$$

where R was the grain yield (kg ha^{-1}) or grain protein concentration (%) that followed a normal distribution with a variance, σ_e^2 , that followed an inverse gamma distribution;

$$\sigma_e^2 \sim IG(\alpha, \beta) \quad (5)$$

and the mean crop response, μ , was defined by a non-linear crop model;

$$\mu \sim \frac{\beta_{max} prec_{py}}{1 + \exp(\beta_{shp} - \beta_1 N - \beta_2 claycontent - \beta_3 claycontent XN)} + z \quad (6)$$

where N was the as-applied N-rate (kg ha^{-1}), $prec_{py}$ and $claycontent$ are defined in Table 3, and z was the SAR spatial random effect;

$$z = (I - B)^{-1} \eta \quad (7)$$

where η had a mean $E(\eta) \equiv 0$ and $Cov(\eta) \equiv \sigma_z^2 I$. I was the identity matrix, and B was the $n \times n$ spatial dependence matrix between n_i and n_j where $i \neq j$ and $i, j = 1, \dots, n$ (Hooten et al., 2019).

All priors used here were the same as in Lawrence et al., (2015), where the prior for σ_e^2 followed a gamma distribution with 0.01 used for the shape and scale parameters in Eq. (5). For β terms in Eq. (6), truncated normal distributions, $TN(0, 1000)$, were used for all terms except where $\beta_{max} = N(0, 1000)$. All Bayesian methods were executed in R using the Stan based *brms* package (Bürkner, 2018, 2021, 2017).

Bayesian multiple linear regression model (Bayes MLR)

The second type of Bayesian approach tested was a multiple linear regression (MLR) model. Based on initial exploration of the data (Hegedus & Maxwell, 2022), a normal distribution as the base of a SAR MLR model was used (Anselin et al., 2004; Hooten et al., 2019; Ver Hoef et al., 2018);

$$R \sim N(\mu, \sigma_e^2) \quad (8)$$

where R was the grain yield (kg ha^{-1}) or grain protein concentration (%), σ_e^2 was identical in formation and priors as Eq. (5), and;

$$\mu = X\beta + z \quad (9)$$

where β was the coefficient vector ($\beta_0, \dots, \beta_{p-1}$) with p the number of predictors from Table 3, X is the $n \times p$ covariate matrix of covariate vectors, and z was the SAR random effect. To avoid overfitting, top-down feature selection was applied and omitted covariates where dropping them from the model reduced the AIC by more than two units. Thus, Bayesian MLR models vary in their covariates between fields and splits but are all fit from the same initial model;

$$\begin{aligned} \mu \sim & \beta_0 + \beta_1 N + \beta_2 aspect_{cos} X aspect_{sin} + \beta_3 slope + \beta_4 elev + \beta_5 tpi + \beta_6 prec_{py} \\ & + \beta_7 gdd_{py} + \beta_8 ndvi_{py} + \beta_9 ndvi_{2py} + \beta_{10} ndwi_{py} + \beta_{11} ndwi_{2py} + \beta_{12} bulkdensity \\ & + \beta_{13} claycontent + \beta_{14} pw + \beta_{15} watercontent + \beta_{16} carboncontent + z \end{aligned} \quad (10)$$

where predictor variables are covariates from Table 3 with uniform priors on coefficients $\beta_0 - \beta_{16}$ (Gelman, 2006), and the SAR spatial random error, z , defined in Eq. (7). As with the non-linear modeling, the Stan based *brms* package in R was used for Bayesian analyses (Bürkner, 2018, 2021, 2017).

Random forest regression (RF)

Random forest (RF) regression is an increasingly popular non-parametric machine learning method that has been used to model spatial data (Georganos et al., 2021; Jing et al., 2016), specifically crop response data (Mariano & Mónica, 2021; Marques Ramos et al., 2020; Paccioretti et al., 2021), where crop responses are estimated by the ensemble of regression trees using binary splits of observations based on covariate data. The *ranger* package in R was used for fitting and generating predictions (Wright & Ziegler, 2017). The predictors included in the RF models for yield and protein are in Table 3. To avoid over-fitting, top-down feature selection was performed where covariates were omitted if dropping them from the model resulted in an AIC decrease of over two units. The number of trees (*ntrees*) and the number of covariates sampled at each node (*mtry*) were optimized during the fitting process. To account for spatial effects, UTM coordinates were included as a variable definition following spatial random forest regression practices in the literature (Janatian et al., 2017; Walsh et al., 2017; Wang et al., 2017).

Net-return calculation and prediction

While predicting yield and grain protein concentration are important, the metric of interest for most farmers is net-return, as management decisions are often ultimately based on the expected profit of the given management action. However, a net-return dataset had to be created for each field and year as this data was unavailable. In Montana's wheat systems, grain yield and protein concentration influence net-returns. A nearest neighbor analysis was performed for each split case, field and year where yield and protein were co-located into a test dataset from which the ability of models to predict net-return could be evaluated. To reduce the distance between neighbors of the differing datasets, the nearest yield measurement was georeferenced to each protein measurement because yield observations were collected at a higher density than grain protein concentration observations. To assess the ability of each model type to forecast net-returns, yield and protein were predicted from their respective models to the same locations as in the "observed" net-return dataset, predicted net-return was calculated using the modeled values, and performance was measured by the RMSE between the "observed" and predicted net-returns.

As mentioned above, protein concentration is used to calculate the wheat net-return for farmers in Montana. Farmers receive premiums or dockages to their base price based on the protein concentration of their wheat. Due to data constraints, the protein premium/dockage schedule from 2021 from one elevator in Billings, Montana was assumed for all fields and years. This schedule dictated that farmers received an additional two cents per half percent protein above 11.5% to 14% and lost eight cents per half percent protein below 11.5% to 9.5% off their received base price. Other economic data used to generate values for the economic parameters that influence net-return were gathered for each year from a combination of surveys from the United States Department of Agriculture (USDA) Economic Research Service (ERS) and the farmers in this study;

$$P = Bp + (B0pd + B1pd * protein + B2pd * protein^2) \quad (11)$$

Table 5 Economic parameters used for the calculation of net-return from Eq. 6 derived from USDA ERS and farmer survey information

Year	Base price (<i>Bp</i>)	Cost of N (<i>CA</i>)	Fixed costs (<i>FC</i>)
2016	\$0.15 kg ⁻¹	\$0.75 kg ⁻¹	\$403.13 ha ⁻¹
2017	\$0.21 kg ⁻¹	\$1.10 kg ⁻¹	\$424.23 ha ⁻¹
2018	\$0.19 kg ⁻¹	\$0.71 kg ⁻¹	\$436.51 ha ⁻¹
2019	\$0.16 kg ⁻¹	\$0.75 kg ⁻¹	\$449.18 ha ⁻¹
2020	\$0.16 kg ⁻¹	\$0.62 kg ⁻¹	\$439.67 ha ⁻¹
2021	\$0.26 kg ⁻¹	\$0.88 kg ⁻¹	\$450.74 ha ⁻¹

Parameters vary across years but did not differ across fields for a given year

Table 6 The average RMSE of grain yield (kg ha⁻¹) across data splits from Fig. 3

Field	Beta function	GAM	Bayes NLR	Bayes MLR	RF
B1	2440.91	5015.10	2122.44	2138.59	2412.05
B2	2645.29	2149.64	<i>1883.15</i>	2068.48	1797.65
B3	3382.39	3700.79	2660.58	<i>2606.11</i>	1351.07
D1	1300.23	1139.24	1406.48	<i>1085.34</i>	1023.45
D2	<i>1464.80</i>	3726.03	1856.87	2004.08	1233.95
D3	<i>1333.05</i>	2123.18	1521.88	1537.62	1187.48
I1	3146.71	6568.38	<i>1837.13</i>	6727.25	1002.24

Bolded values indicate the model type with the lowest RMSE for each field, and italic represents the second most accurate model type for forecasting crop responses in each field

$$NR = yield * P - CA * AA - FC - ssAC \quad (12)$$

where *P* was the final price received (\$ kg⁻¹) after imposing the protein premium/dockage on *Bp*, the base price received (\$ kg⁻¹), *B0pd* was the intercept of the protein premium/dockage function set by the grain elevator, *B1pd* was the coefficient on the grain protein concentration (%), and *B2pd* was the coefficient on the squared protein term. The net-return, *NR* (\$ ha⁻¹), received was a function of the product of the *yield* (kg ha⁻¹) and *P*, minus the cost of the applied input (*CA*) multiplied by the as-applied N-rate of the input (*AA*), the fixed costs (*FC*) associated with production (\$ ha⁻¹) that do not include the input, and the cost per hectare of the site-specific application (*ssAC*).

No cost of site-specific application (*ssAC*) was included since each farmer in this study used their own already available equipment to enact the varied rates. The parameters for the protein premium and dockage, *B0pd*, *B1pd*, and *B2pd*, were derived from fitting a smoothed step function to the protein premium schedule. The *Bp*, *CA*, and *FC* values used for each year are in Table 5.

Results

The RMSE values for grain yield, grain protein concentration and net-return were reported for each of the five model types, for each field and for each data split case (Fig. 3; Table 13). The first objective was to assess the accuracy of five different types of winter

Table 7 The average RMSE of grain protein concentration (%) across data splits from Fig. 3

Field	Beta function	GAM	Bayes NLR	Bayes MLR	RF
B1	2.50	3.54	6.49	3.79	3.21
B2	3.06	3.78	7.80	3.13	2.72
B3	3.87	1.72	4.44	4.89	1.74
D1	1.57	2.01	1.74	2.15	1.48
D2	3.53	4.22	4.60	7.30	1.35
D3	1.90	1.34	3.75	2.08	1.71
I1	3.74	11.29	7.40	9.57	3.11

Bolded values indicate the model type with the lowest RMSE for each field, and italic represents the second most accurate model type for forecasting crop responses in each field

wheat crop response (grain yield and grain protein concentration) forecast models using all combinations of training and holdout years (Fig. 3). The mean of the RMSE of observed and predicted grain yields were taken across the three data configurations for an average RMSE per model for each field (Table 6). These results indicate that in six out of seven fields, across any year forecasted, the RF regression resulted in the most accurate forecast of grain yield in a given test dataset with one year of data held out. Bayesian approaches were the next most accurate where, in one out of seven fields, the Bayesian NLR model had the lowest RMSE out of all the model types. In two out of seven fields the Bayesian NLR had the second lowest RMSE out of all model types. Additionally, in three out of seven fields the Bayesian MLR model had the second best RMSE. The GAM was the poorest predictive model type, with the highest RMSE out of all models in five out of seven fields.

The mean RMSE of observed and predicted grain protein concentration were calculated across the three data cases in Fig. 3 for each field (Table 7). In four out of seven fields, predictions of grain protein concentration from the RF regression had the lowest mean RMSE across data configurations and in three out of seven fields had the second lowest RMSE. The non-linear beta function had the lowest RMSE for one out of seven fields and second lowest RMSE in four out of seven fields. These results indicate that, like the RF regression models for grain yield, the RF models generated the most accurate forecasts of grain protein concentration, averaged across data configurations. On average across fields, the second-best model for forecasting grain protein concentration was the beta function. In contrast to results of grain yield models, the GAM was the best predictive model type for grain protein concentration in two out of seven fields, while the Bayesian approaches were the least accurate at forecasting protein.

The second objective was to investigate the influence of the sequence of years used for training field specific models on forecast accuracy. The mean RMSE across fields from each crop response model for each split data case in Fig. 3 was calculated where: A=first two years of experimental data were used for training and the most recent year of experimental data was the test set, B=the first and last years of experimental data were used for training and the intermediary year was the test set, C=the most recent years of experimental data were used for training and the first year of experimentation was used for testing.

The mean RMSE of grain yield for each model using each data split case were taken across fields to assess which set of training years (split cases) resulted in the most accurate forecasts (Table 8). For the 2016, 2018, 2020 data collection schedules, split case C had the lowest mean RMSE of grain yield across fields for four out of the five models, with the exception

Table 8 Mean RMSE of grain yield (kg ha^{-1}) across fields for each model type, where the given model type was fit for each field using the split case scenario in Fig. 3

Model	Data schedule 2016–2020			Data schedule 2017–2021		
	A	B	C	A	B	C
Beta function	1524.73	1324.24	4146.24	2462.23	2605.47	1318.73
GAM	7987.63	2533.45	1567.68	1386.21	3788.27	3129.49
Bayes NLR	2073.44	2182.08	1651.83	1789.51	1876.77	1745.79
Bayes MLR	2869.12	5036.14	1751.03	1285.15	2220.84	1786.43
RF	1205.85	1418.36	956.84	1635.32	2379.04	1218.79

The mean RMSE across fields for each model and split case were divided by the crop-fallow data schedule to assess differences between a 2016, 2018 and 2020 schedule and a 2017, 2019 and 2021 schedule. Bolded cells indicate the split data case (A, B or C) that had the lowest mean RMSE across fields for each model type

Table 9 Mean RMSE of grain protein concentration (%) across fields for each model type, where the given model type was fit for each field using the split case scenario in Fig. 3

Model	Data schedule 2016–2020			Data schedule 2017–2021		
	A	B	C	A	B	C
Beta function	5.94	1.43	2.42	2.67	2.26	2.20
GAM	6.85	3.48	3.58	3.39	3.75	2.20
Bayes NLR	6.12	4.30	4.73	5.94	6.17	3.92
Bayes MLR	6.81	6.467	4.61	3.09	3.79	2.19
RF	1.98	1.77	2.19	2.50	2.19	2.72

being that split case B resulted in the lowest RMSE across fields using a beta function model. In the 2017, 2019, 2021 data schedule, split case C yielded the lowest mean RMSE of grain yield across fields for three out of the five models, while split case A yielded the lowest mean RMSE across fields when the GAM or Bayesian MLR model were used (Table 8).

The same assessment across data splits (Fig. 3) was performed where the models were used to forecast grain protein concentration and the mean RMSE was taken across fields (Table 9). For the 2016, 2018, 2020 data collection schedules, split case B had the lowest mean RMSE of protein across fields for four out of the five models with the exception being that split case C resulted in the lowest RMSE across fields when the Bayesian MLR was used. In the 2017, 2019, 2021 data schedule, split case C yielded the lowest mean RMSE for protein across fields for four out of the five models, while split case B yielded the lowest mean RMSE across fields when the RF regression model was used.

Accurate predictions of grain yield and protein concentration are necessary for calculating net-return, which is the metric of interest for most farmers when making decisions. Both the grain yield and protein models fit for a given field with a given model type were used to predict yield and protein in a dataset with calculated net-returns from the test year. The same tables as above were generated for summarizing the RMSE of net-returns.

The mean RMSE values for net-return were calculated across the three data split cases for each model type in each field (Table 10). The RF regression model produced the lowest mean RMSE of net-return across data split cases in four out of seven fields, and the second lowest mean RMSE in two out of seven fields. The second-best model for forecasting net-returns was the beta function, which produced the lowest RMSE of net-return across data

Table 10 The average RMSE of net-return (\$ ha⁻¹) across data splits from Fig. 3

Field	Beta function	GAM	Bayes NLR	Bayes MLR	RF
B1	1551.89	1787.89	3509.56	1785.57	1585.83
B2	<i>1102.01</i>	1354.92	3672.17	1399.27	816.45
B3	<i>1391.43</i>	1450.85	2032.71	3283.46	545.12
D1	439.47	443.25	456.42	<i>430.38</i>	340.17
D2	3058.18	2814.71	<i>1884.18</i>	3501.24	561.16
D3	<i>770.06</i>	644.03	1604.63	950.49	731.64
I1	1193.06	5667.15	3414.23	2478.43	<i>1307.34</i>

Bolded values indicate the model type with the lowest RMSE for each field, and Italic represents the second most accurate model type for forecasting crop responses in each field

Table 11 Mean RMSE of net-return (\$ ha⁻¹) across fields for each model type, where the given model type was fit for each field using the split case scenario in Fig. 3

Model	Data schedule 2016–2020			Data schedule 2017–2021		
	A	B	C	A	B	C
Beta Function	3438.38	632.68	738.49	943.45	913.63	1236.30
GAM	5095.38	1192.54	1644.63	1180.54	1053.49	1352.03
Bayes NLR	1716.67	2425.12	2560.02	2340.62	3249.38	2048.14
Bayes MLR	4117.98	2125.07	1417.16	1065.65	1440.25	1109.31
RF	935.39	728.91	694.65	676.48	915.61	1150.36

The mean RMSE across fields for each model and split case were divided by the crop-fallow data schedule to assess differences between a 2016, 2018 and 2020 schedule and a 2017, 2019 and 2021 schedule. Bolded cells indicate the split data case (A, B or C) that had the lowest mean RMSE across fields for each model type

split cases in two out of seven fields and produced the second lowest mean RMSE across data types in three out of seven fields.

As done for grain yield and protein above, the mean RMSE across fields was calculated for each data split case from Fig. 3 and model type (Table 11). For both data schedules there was no consistent split case for training and testing data across models based on the mean of net-return RMSEs across fields. In both data schedules the split test case where the intermediary year was used as the test scenario resulted in the lowest RMSE for two out of five models. Split case A and C generated the most accurate forecasts from Bayesian NLR models in the 2016–2020 and 2017–2021 data schedules, respectively. Split case C and A generated the most accurate forecasts from Bayesian MLR and RF regression models in the 2016–2020 and 2017–2021 data schedules, respectively, in juxtaposition to that of the Bayesian MLR and RF regression models.

Discussion

Due to the variation in crop responses across fields, and the drivers of responses and net-returns, between fields and across years (Hegedus & Maxwell, 2022), it was expected that the forecast model type most appropriate for characterizing the crop response to variable

N fertilizer rates would vary across fields and depend on the years used for training the model. For both grain yield and grain protein concentration, the results indicate that the RF regression most consistently reduced the uncertainty of forecasts, based on the mean RMSE across the three test years. This contrasts with the assumption that the most appropriate model for predicting crop responses to variable N rates would vary across fields and crop responses. Yet the RF regression model should not be considered a silver bullet, despite performing best on average across fields. Despite performing best on average across fields, this simplifies field-specific variation in the most predictive model type and variation in the most predictive model for different crop responses in a given field. A RF model seems a reasonable assumption for a default model that a selection of other model types are compared against, yet the importance of testing multiple model types is demonstrated in field B1, where the Bayesian NLR regression model performed better than the RF at forecasting grain yield (Table 6). Additionally, the Beta model performed better than the RF model at predicting grain protein concentration in field B1 (Table 7). This is important for two reasons; the first being that the RF was not the most predictive model type for this field, despite performing best on average across fields, and the second that the most predictive model type for field B1 varied between the crop responses. The GAM outperformed the RF at predicting grain protein concentration in fields B3 and D3 (Table 7), indicating two other cases where, without conducting model selection on a field-specific basis, assuming a default (e.g., RF) model type would not result in the most accurate forecasts of grain protein concentration and subsequent net-return.

Variation in the most predictive model type for different crop responses was further illustrated when assessing the second-best model types for each field. The second-best model at predicting grain yield was typically one of the Bayesian methods (Table 6) while the beta function was typically the second-best model type at predicting grain protein concentration (Table 7).

Based on the influence of grain yield and grain protein on net-return and the results from objective 1, it was unsurprising that using RF regression for modeling both crop responses to calculate net-return tended to result in the most accurate net-return forecasts across field (Table 10). However, like the results for grain protein concentration, the second-best model across fields for generating an accurate forecast of net-returns was the beta function. In fields B1 and D3, the model type that produced the best forecasts of grain protein also produced the best forecasts of net-return. This indicates that when using models to forecast net-return there is a higher importance of generating accurate predictions of grain protein concentration compared to grain yield. If accurate grain yield predictions had a higher influence on forecasting net-return, it would be expected that the second-best model type for net-return would be the Bayesian methods. Field I1 represents an outlier to this assessment however, as the RF regression produced the most accurate forecasts of grain yield and grain protein independently (Tables 6, 7), yet when the beta function model was used for predicting both yield and protein to calculate net-return resulted in a more accurate forecast compared to when the RF regression was used for predicting yield and protein to calculate net-return (Table 10).

Net-return forecasts are important for simulating management recommendations and in forming the basis of farmer decision making. Despite the RF regression model's dominance in forecasting, these results should not detract from the importance of the field-specificity in model selection and the difference in the most appropriate model for the different crop responses, which require confronting data from on-farm experimentation with multiple models. For example, in a hypothetical situation for a field where a given model type produces the best forecasts for yield and an alternate model type produces the best

forecasts for protein, forecasting net-returns and simulating management recommendations in a decision support tool should be based on separate models for grain yield and grain protein concentration. In the data shown in this paper, an example of this concept arises in fields B1, B3 and D3. Highlighting field B1, the Bayesian NLR model produced the best forecast for grain yield while the beta function model produced the best forecast for grain protein concentration, indicating not only that the RF regression should not be used at all in this situation, but that two separate model forms should be used for yield and protein. Thus, mixing, and matching models for grain yield and grain protein might produce more accurate forecasts of net-return.

When selecting models, consideration must be given to the weather conditions in the years that data is available for training, as some models produce better predictions in certain weather conditions than others. It was predicted that if weather conditions of the years used for training differed from the conditions of the year in the test dataset forecasts of crop responses would be less accurate compared to when the conditions in the years used for training resembled the conditions in the year of the test dataset. Since the data used in this study came from rain-fed winter wheat production fields, amount of precipitation received was judged to be the best weather condition to assess year suitability for predictions. The wettest years in the data set included 2018 and 2019, while 2020 and 2021 were, on average, the driest (Table 12).

Averaged across fields, most of the five model types generated the most accurate forecasts of grain yield in split configuration C, where the most recent two years of experimental data were used to train models that made predictions on yields in the first year of experimentation (Table 8). The out-of-sequence result is interpreted as that the most accurate forecasts of grain yield were made when training data contained a wet and dry year, which meant that the relatively dry year in the test set was reasonably represented in the training data. The least accurate forecasts of yield were made in split configuration B, where the two drier years were used for training and forecasts were made for the wet year. Having a range of weather conditions (wet and dry years) in the training dataset meant that forecasts in an unknown year were stronger compared to when data from similar weather conditions were used to train a model.

The fields analyzed were collected on two crop-fallow schedules, with the 2016 schedule containing data collected in 2016, 2018 and 2020 and the 2017 schedule containing data collected in 2017, 2019 and 2021. The influence of the configuration of training and testing data on the predictive performance of grain protein models depended on whether training and testing data came from the 2016 or 2017 data schedule (Table 9). When data came from the 2017 schedule, split configuration C resulted in the most accurate forecasts across most models, like for grain yield. However, when the data came from the 2016

Table 12 The total precipitation from the AgriMet station setup by the Bureau of Land Reclamation at the Central Ag Research Station in Conrad, MT

Year	Total precipitation (mm)	Mean precipitation (mm)
2016	340	287
2017	234	
2018	474	497
2019	520	
2020	294	274
2021	256	

schedule the most accurate forecasts for most models came from split configuration B, where 2016 and 2020 were used for training to predict protein concentration in 2018. The 2017 schedule was the most extreme, with both the wettest and driest precipitation years (Table 12). Additionally, the difference in precipitation between the two driest years in the 2017 schedule, 2021 and 2017, was 21.6 mm which indicates that conditions in 2021 and 2017 were very similar. Thus, if the training dataset contained either 2021 or 2017 and the test set was either 2017 or 2021, the conditions of the test set would be represented in the training data (configuration A or C). On the other hand, precipitation in the 2016 schedule had a smaller range of precipitation, but precipitation varied between years more so than in the 2017 schedule where the difference in precipitation between two driest years of the 2016 schedule, 2020 and 2016, was 47.1 mm (Table 12). This indicates that in the 2016 schedule, there is not a pair of years that are as similar as 2021 and 2017, so no data configuration existed where conditions in the test year were almost identical to conditions in a year from the training dataset.

Analysis on Montana dryland conventionally managed agroecosystems has shown that variable rate experimentation of N fertilizer inputs requires 6–8 crop years of data for model coefficients to converge and enable accurate predictions of outcomes (Lawrence et al., 2015). This further highlights the need for public research funding of long-term agricultural studies on how precision agriculture technologies can improve management recommendations that promote sustainability (DeLonge et al., 2020). The results in this paper were based off three years of crop data, below the recommended threshold, and further illustrate the importance of a larger set of data from which to train models. More accurate forecasts in unknown years are expected with more than three years of data, because more weather conditions in the training datasets will increase the likelihood that conditions in the next year will be represented in the model fitting process. The importance of representation of conditions in an upcoming year in the training data is demonstrated using the situation where the best forecast for grain yield was observed: field D1 where the RF model was used with data configuration C (Table 13). Assuming perfect representation of conditions in the next year, achieved by training the model with a subset of data containing all three years and testing in the holdout set containing all three years, the forecast accuracy was almost halved compared to when 2019 and 2021 data were used to train the model and

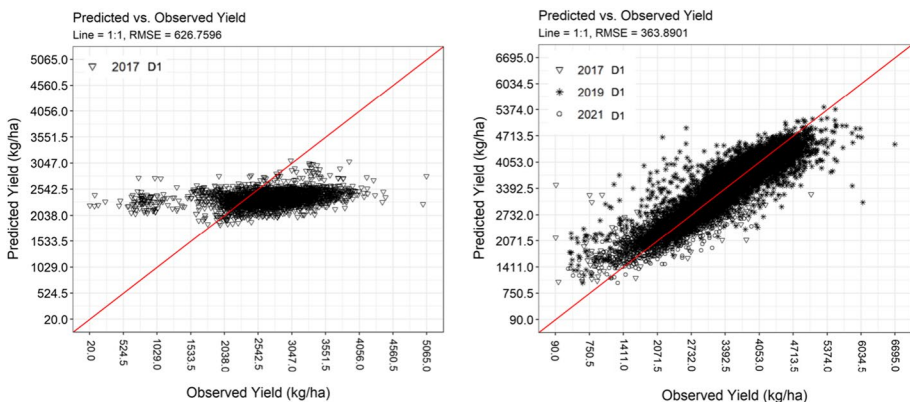


Fig. 4 Predicted vs. observed grain yields (kg ha^{-1}) from field D1 using the RF regression model where 2019 and 2021 data were used for training and tested in 2017 (left) and training and test datasets were derived from all three years (right). RMSE reported in kg ha^{-1}

tested on 2017 data (Fig. 4). As a backlog of data from varying weather years is collected, it is expected that forecasts will iteratively improve year after year until uncertainty in crop response forecasts is constrained to model structure and selection rather than limitations in the conditions represented in the training data.

A further concern over interpretation of these results, especially in selection of model training data, is the influence of the fallow year or rotation crop inclusion. The interaction of these factors with weather will take more years to refine the models and characterize the certainty in recommendations. As more data is gathered and synthesized into decision support systems via modeling efforts, researchers and scientists need to recognize potential pitfalls of their modeling approaches beyond just model selection, such as issues of scaling model results and the data used in models (Fritsch et al., 2020).

Crop response models used in decision support systems directly influence management recommendations. The certainty of input management recommendations has been sharply improved with on-farm experimentation facilitated by precision technologies, site-specific data availability and modern analytics. These results suggest continued use of on-farm experimentation to produce a backlog of data for training models in various weather conditions, crop rotations and practices so the crop response signal can emerge from the noise caused by within field crop response and weather variation. It must also be recognized that field-specific data might require confrontation with a suite of models to make the best assessment of which model to base input recommendations. These results demonstrate that modern farms have the equipment and technology to inform adaptive management through on-farm experimentation, and that adoption will rely on development of decision support systems, demonstration of the profitability of variable rate application and possible policy incentives.

Conclusion

Data generated from on-farm experimentation on seven fields were used to evaluate five different types of crop response models. The RF regression typically resulted in the most accurate forecasts of crop responses and net-returns across most fields, but not all. In some cases, the model type that produced the most accurate forecast of grain yield was not the same as the model producing the most accurate forecast of grain protein concentration. When the data used for training models was collected in years with varying weather conditions, forecasts of crop responses in a new year tended to be more accurate than when models were trained on data from similar weather conditions and used to forecast crop responses in a year with dissimilar weather conditions. When selecting crop response models for generating forecasts of net-returns and simulating management outcomes in a decision support system, field specificity in the most appropriate model type for each individual crop response should be considered. Additionally, the most appropriate model for a given crop response in each field may depend on the range of weather conditions in the years that data is available to train that model. These results are important to developers of decision support tools because selection of the underlying models used to simulate management outcomes and calculate net-returns need to consider the spatiotemporal specificity of the data available.

Appendix

See Table 13.

Table 13 Table with the RMSE for each crop response and net-return, type of model, and permutation (A, B, C) of data used for training or testing as shown in Fig. 3

RMSE for grain yield (kg ha ⁻¹)				RMSE for protein concentration %				RMSE for net-return (\$ ha ⁻¹)			
Beta function				Beta function				Beta function			
Field	A	B	C	Field	A	B	C	Field	A	B	C
B1	3148.90	2935.09	1238.74	B1	3.19	2.36	1.95	B1	977.28	1345.10	2333.30
B2	2940.79	3334.18	1660.90	B2	3.37	3.14	2.67	B2	1250.88	1040.66	1014.49
B3	1584.93	1690.53	6871.70	B3	8.06	1.72	1.82	B3	2587.74	541.11	1045.44
D1	1297.01	1547.13	1056.54	D1	1.45	1.27	1.98	D1	602.19	355.12	361.10
D2	1622.90	1567.17	1204.33	D2	8.10	1.02	1.47	D2	8157.14	553.68	463.72
D3	1744.54	926.49	1328.11	D3	2.04	1.45	2.22	D3	581.17	587.50	1141.51
I1	1146.55	1112.78	7180.80	I1	5.54	1.52	4.16	I1	2427.47	848.42	303.29
GAM				GAM				GAM			
B1	2418.55	5080.02	7546.72	B1	2.84	6.10	1.67	B1	862.55	1511.10	2990.01
B2	981.91	4583.06	883.94	B2	5.79	3.41	2.15	B2	2315.97	1160.35	588.45
B3	7771.55	2362.69	968.14	B3	2.02	1.60	1.53	B3	2760.57	1150.34	441.62
D1	758.17	1701.73	957.82	D1	1.53	1.75	2.76	D1	363.11	489.02	477.62
D2	7681.69	2653.18	843.23	D2	7.28	3.83	1.53	D2	6523.76	1461.54	458.82
D3	1390.75	2172.43	2806.35	D3	1.52	0.96	1.53	D3	533.45	564.66	833.98
I1	15,106.63	2945.50	1653.00	I1	16.60	7.53	9.73	I1	10,563.73	1593.62	4844.12
Bayes NLR				Bayes NLR				Bayes NLR			
B1	2147.47	1976.54	2243.31	B1	7.71	7.44	4.31	B1	3699.99	3725.88	3102.81
B2	1848.52	2153.62	1647.32	B2	8.40	9.34	5.67	B2	2810.97	5602.23	2603.31
B3	2974.07	2823.95	2183.72	B3	5.39	3.93	3.99	B3	1718.17	2244.68	2135.27
D1	1372.56	1500.15	1346.72	D1	1.70	1.74	1.78	D1	510.91	420.02	438.31
D2	1816.12	2219.94	1534.55	D2	5.58	4.27	3.95	D2	1024.08	2314.66	2313.80
D3	1374.78	1919.44	1271.42	D3	4.07	3.85	3.33	D3	802.84	2040.95	1970.09

Table 13 (continued)

Bayes NLR			Bayes NLR			Bayes NLR					
II	2128.79	1764.97	1617.64	II	9.42	5.15	7.64	II	3321.58	3100.18	3820.93
Bayes MLR			Bayes MLR			Bayes MLR					
B1	2010.19	2578.19	1827.39	B1	3.40	5.68	2.29	B1	1034.45	2355.51	1966.75
B2	1019.07	2506.86	2679.51	B2	3.79	3.24	2.36	B2	1789.97	1379.80	1028.03
B3	4280.07	2239.78	1298.46	B3	10.06	2.32	2.29	B3	8177.02	946.31	727.07
D1	826.17	1577.48	852.38	D1	2.07	2.45	1.93	D1	372.53	585.44	333.15
D2	2591.85	2246.79	1173.60	D2	7.59	4.17	10.15	D2	5912.85	1657.33	2933.55
D3	1518.81	1575.44	1518.61	D3	2.00	1.86	2.38	D3	721.41	905.55	1224.52
II	3085.74	14,082.57	3013.43	II	7.58	17.52	3.61	II	1660.66	4991.10	783.52
RF			RF			RF					
B1	2693.10	2313.15	2229.90	B1	3.42	2.39	3.82	B1	809.90	1417.41	2530.17
B2	1425.97	3167.28	799.69	B2	2.64	2.86	2.65	B2	812.64	939.62	697.09
B3	1350.20	1676.90	1026.12	B3	1.69	1.74	1.78	B3	639.70	540.88	454.77
D1	786.89	1656.70	626.76	D1	1.43	1.31	1.70	D1	406.90	389.80	223.81
D2	822.63	1851.63	1027.60	D2	1.53	1.20	1.32	D2	639.13	637.55	406.80
D3	1450.04	989.93	1122.47	D3	1.57	1.35	2.21	D3	584.27	510.87	1099.79
II	1200.55	1154.99	651.18	II	3.12	2.77	3.44	II	1878.46	1226.32	817.23

Acknowledgements The authors wish to thank the team members of the On-Field Precision Experiment (OFPE) project and the Agroecology Lab at Montana State University. 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), the 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), and the Montana Fertilizer Advisory Council from 2016 to 2021. Additionally, the authors would like to thank the Montana INBRE funded Statistical Consulting & Research Services at Montana State University, specifically Dr. Mark Greenwood and Caroline Hardy.

Author contributions PBH and BDM authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by PBH. The first draft of the manuscript was written by PBH, and all authors commented on versions of the manuscript. All authors read and approved the final manuscript.

Funding 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), the 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), and the Montana Fertilizer Advisory Council from 2016 to 2021. Paul B. Hegedus reports financial support was provided by Montana Fertilizer Advisory Committee. Paul B. Hegedus reports a relationship with University of Illinois at Urbana-Champaign that includes employment and travel reimbursement. Bruce D. Maxwell reports a relationship with University of Illinois at Urbana-Champaign that includes funding grants and travel reimbursement.

Data availability The datasets generated during and/or analyzed during the current study are not publicly available without each farm’s approval but may be obtained through reasonable request from the corresponding author.

Declarations

Conflict of interest The authors have no competing interests 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

- Anselin, L., Bongiovanni, R., & Lowenberg-DeBoer, J. (2004). A spatial econometric approach to the economics of site-specific nitrogen management. *American Journal of Agricultural Economics*, 86(8), 675–687.
- Basso, B., & Antle, J. (2020). Digital agriculture to design sustainable agricultural systems. *Nature Sustainability*, 3(4), 254–256. <https://doi.org/10.1038/s41893-020-0510-0>
- Basso, B., Ritchie, J. T., Cammarano, D., & Sartori, L. (2011). A strategic and tactical management approach to select optimal N fertilizer rates for wheat in a spatially variable field. *European Journal of Agronomy*, 35(4), 215–222. <https://doi.org/10.1016/j.eja.2011.06.004>
- Blackmore, S., & Moore, M. (1999). Remedial correction of yield map data. *Precision Agriculture*, 1(1), 53–66. <https://doi.org/10.1023/A:1009969601387>

- Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, 173, 74–84. <https://doi.org/10.1016/j.agrformet.2013.01.007>
- Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P. C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- Bürkner, P. C. (2021). Bayesian item response modeling in R with brms and stan. *Journal of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- Carolan, M. (2017). Publicising food: Big data, precision agriculture, and co-experimental techniques of addition. *Sociologia Ruralis*, 57(2), 135–154. <https://doi.org/10.1111/soru.12120>
- Chen, K., O'leary, R. A., & Evans, F. H. (2019). A simple and parsimonious generalised additive model for predicting wheat yield in a decision support tool. *Agricultural Systems*, 173, 140–150. <https://doi.org/10.1016/j.agsy.2019.02.009>
- Clancy, M. (2019). *Real-time protein monitoring*. Next Instruments.
- De Bastiani, F., de Aquino, M., Cysneiros, A. H., Uribe-Opazo, M. A., & Galea, M. (2015). Influence diagnostics in elliptical spatial linear models. *TEST*, 24(2), 322–340. <https://doi.org/10.1007/s11749-014-0409-z>
- DeLonge, M., Robbins, T., Basche, A., & Haynes-Mawlow, L. (2020). The state of sustainable agriculture and agroecology research and impacts: a survey of US scientists. *Journal of Agriculture, Food Systems, and Community Development*, 9(2), 1–26. <https://doi.org/10.5304/jafscd.2020.092.009>
- Drachal, K. (2019). Analysis of agricultural commodities prices with new Bayesian model combination schemes. *Sustainability (switzerland)*. <https://doi.org/10.3390/su11195305>
- Fritsch, M., Lischke, H., & Meyer, K. M. (2020). Scaling methods in ecological modelling. *Methods in Ecology and Evolution*, 11(11), 1368–1378. <https://doi.org/10.1111/2041-210X.13466>
- Gardner, G., Mieno, T., & Bullock, D. S. (2021). An economic evaluation of site-specific input application Rx maps: Evaluation framework and case study. *Precision Agriculture*, 22(4), 1304–1316. <https://doi.org/10.1007/s11119-021-09785-z>
- Gebbers, R., & Adamchuck, V. I. (2010). Precision agriculture and food security. *Science*, 327(5967), 828–831.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis*, 1(3), 515–534. <https://doi.org/10.1214/06-BA117A>
- Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuyse, S., et al. (2021). Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, 36(2), 121–136. <https://doi.org/10.1080/10106049.2019.1595177>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*, 157(2–3), 89–100. [https://doi.org/10.1016/S0304-3800\(02\)00204-1](https://doi.org/10.1016/S0304-3800(02)00204-1)
- Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3), 297–310. <https://doi.org/10.1214/ss/1177013604>
- Hatfield, J. L., Cryder, M., & Basso, B. (2020). Remote sensing: advancing the science and the applications to transform agriculture. *IT Professional*, 22(3), 42–45. <https://doi.org/10.1109/MITP.2020.2986102>
- Hegedus, P. B., & Maxwell, B. D. (2022). Rationale for field-specific on-farm precision experimentation. *Agriculture, Ecosystems and Environment*, 338, 1–14. <https://doi.org/10.1016/j.agee.2022.108088>
- Hegedus, P. B. (2020). OFPE: On-Farm Precision Experiments (OFPE) Data Management and Analysis Tools. R package version 1.8.11. <https://github.com/paulhegedus/OFPE.git>
- Hooten, M. B., Ver Hoef, J. M., & Hanks, E. M. (2019). Simultaneous autoregressive model (SAR). *Wiley StatsRef*. https://doi.org/10.1007/978-0-387-35973-1_1217
- Houlahan, J. E., McKinney, S. T., Anderson, T. M., & McGill, B. J. (2017). The priority of prediction in ecological understanding. *Oikos*, 126(1), 1–7. <https://doi.org/10.1111/oik.03726>
- Huang, X., Huang, G., Yu, C., Ni, S., & Yu, L. (2017). A multiple crop model ensemble for improving broad-scale yield prediction using Bayesian model averaging. *Field Crops Research*, 211(6), 114–124. <https://doi.org/10.1016/j.fcr.2017.06.011>

- Janatian, N., Sadeghi, M., Sanaeinejad, S. H., Bakhshian, E., Farid, A., Hasheminia, S. M., et al. (2017). A statistical framework for estimating air temperature using MODIS land surface temperature data. *International Journal of Climatology*, 37(3), 1181–1194. <https://doi.org/10.1002/joc.4766>
- Jing, W., Yang, Y., Yue, X., & Zhao, X. (2016). A comparison of different regression algorithms for downscaling monthly satellite-based precipitation over North China. *Remote Sensing*, 8(10), 1–17. <https://doi.org/10.3390/rs8100835>
- Johnson, D. M. (2014). An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, 141, 116–128. <https://doi.org/10.1016/j.rse.2013.10.027>
- Joshi, V. R., Kazula, M. J., Coulter, J. A., Naeve, S. L., & Garcia y Garcia, A. (2021). In-season weather data provide reliable yield estimates of maize and soybean in the US central Corn Belt. *International Journal of Biometeorology*, 65(4), 489–502. <https://doi.org/10.1007/s00484-020-02039-z>
- Lawrence, P. G., Rew, L. J., & Maxwell, B. D. (2015). A probabilistic Bayesian framework for progressively updating site-specific recommendations. *Precision Agriculture*, 16(3), 275–296. <https://doi.org/10.1007/s11119-014-9375-4>
- Mariano, C., & Mónica, B. (2021). A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. *Computers and Electronics in Agriculture*, 184, 106094. <https://doi.org/10.1016/j.compag.2021.106094>
- Marques Ramos, A. P., Prado Osco, L., Furuya, E. G., D., Nunes Gonçalves, W., Cordeiro Santana, D., Pereira Ribeiro Teodoro, L., et al. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, 178, 105791. <https://doi.org/10.1016/j.compag.2020.105791>
- McBratney, A. B., Minasny, B., & Whelan, B. M. (2005a). Obtaining Useful High-resolution Soil Data from Proximally-Sensed Electrical Conductivity/Resistivity (PSEC / R) Surveys. In Stafford, J. V. (Ed.) Proceedings of the 4th European Conference on Precision Agriculture *Precision Agriculture 5* pp 503–510: Wageningen Academic Publishers.
- McBratney, A., Whelan, B., Ancew, T., & Bouma, J. (2005b). Future directions of precision agriculture. *Precision Agriculture*, 6(1), 7–23. <https://doi.org/10.1007/s11119-005-0681-8>
- McFadden, B. R., Brorsen, B. W., & Raun, W. R. (2017). Nitrogen fertilizer recommendations based on plant sensing and Bayesian updating. *Precision Agriculture*, 19(1), 1–14. <https://doi.org/10.1007/s11119-017-9499-4>
- McFadden, J. R., Rosburg, A., & Njuki, E. (2021). Information inputs and technical efficiency in mid-west corn production: Evidence from farmers' use of yield and soil maps. *American Journal of Agricultural Economics*, 104(2), 589–612. <https://doi.org/10.1111/ajae.12251>
- Mueller, N. D., West, P. C., Gerber, J. S., Macdonald, G. K., Polasky, S., & Foley, J. A. (2014). A tradeoff frontier for global nitrogen use and cereal production. *Environmental Research Letters*, 9(5), 054002. <https://doi.org/10.1088/1748-9326/9/5/054002>
- Nandram, B., Berg, E., & Barboza, W. (2014). A hierarchical Bayesian model for forecasting state-level corn yield. *Environmental and Ecological Statistics*, 21(3), 507–530. <https://doi.org/10.1007/s10651-013-0266-z>
- 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>
- Peerlinck, A., Sheppard, J., & Maxwell, B. (2018). Using deep learning in yield and protein prediction of winter wheat based on fertilization prescriptions in precision agriculture. In *14th International Conference on Precision Agriculture*. <https://www.ispag.org/proceedings/?action=abstract&id=5209&title=Using+Deep+Learning+in+Yield+and+Protein+Prediction+of+Winter+Wheat+Based+on+Fertilization+Prescriptions+in+Precision+Agriculture>
- Peerlinck, A., Sheppard, J., & Senecal, J. (2019). AdaBoost with neural networks for yield and protein prediction in precision agriculture. *Proceedings of the International Joint Conference on Neural Networks*. <https://doi.org/10.1109/IJCNN.2019.8851976>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2021). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–153, <https://www.CRAN.R-project.org/package=nlme>
- Pinilla, J., & Negrín, M. (2021). Non-parametric generalized additive models as a tool for evaluating policy interventions. *Mathematics*, 9(4), 1–12. <https://doi.org/10.3390/math9040299>
- Prost, L., Makowski, D., & Jeuffroy, M. H. (2008). Comparison of stepwise selection and Bayesian model averaging for yield gap analysis. *Ecological Modelling*, 219(1–2), 66–76. <https://doi.org/10.1016/j.ecolmodel.2008.07.026>
- R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL <https://www.R-project.org/>.

- Ramsey, A. F. (2020). Probability distributions of crop yields: A Bayesian spatial quantile regression approach. *American Journal of Agricultural Economics*, 102(1), 220–239. <https://doi.org/10.1093/ajae/aaz029>
- Reynolds, W. D., Drury, C. F., Phillips, L. A., Yang, X., & Agomoh, I. V. (2021). An adapted weibull function for agricultural applications. *Canadian Journal of Soil Science*, 101(4), 680–702. <https://doi.org/10.1139/cjss-2021-0046>
- Schimmelpfennig, D., & Lowenberg-DeBoer, J. (2020). Farm types and precision agriculture adoption: Crops, regions, soil variability, and farm size. *SSRN Electronic Journal*, 1–38. <https://doi.org/10.2139/ssrn.3689311>
- Sela, S., van Es, H. M., Moebius-Clune, B. N., Marjerison, R., Melkonian, J., Moebius-Clune, D., et al. (2016). Adapt-N outperforms grower-selected nitrogen rates in northeast and midwestern united states strip trials. *Agronomy Journal*, 108(4), 1726–1734. <https://doi.org/10.2134/agronj2015.0606>
- Sudduth, K. A., & Drummond, S. T. (2007). Yield editor: Software for removing errors from crop yield maps. *Agronomy Journal*, 99(6), 1471–1482. <https://doi.org/10.2134/agronj2006.0326>
- Sudduth, K. A., Drummond, S. T., & Myers, D. B. (2012). Yield editor 2.0: Software for automated removal of yield map errors. Paper No. 121338243.: ASABE <https://doi.org/10.13031/2013.41893>
- Thöle, H., Richter, C., & Ehlert, D. (2013). Strategy of statistical model selection for precision farming on-farm experiments. *Precision Agriculture*, 14(4), 434–449. <https://doi.org/10.1007/s11119-013-9306-9>
- Ver Hoef, J. M., Hanks, E. M., & Hooten, M. B. (2017). On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. *Spatial Statistics*, 25, 68–85. <https://doi.org/10.1016/j.spasta.2018.04.006>
- Ver Hoef, J. M., Petersen, E. E., Hooten, M. B., Hanks, E. M., & Fortin, M.-J. (2018). Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs*, 88(1), 36–59.
- Wajid, A., Hussain, K., Ilyas, A., Habib-Ur-rahman, M., Shakil, Q., & Hoogenboom, G. (2021). Crop Models: Important Tools in Decision Support System to Manage Wheat Production under Vulnerable Environments. *Agriculture (switzerland)*. <https://doi.org/10.3390/AGRICULTURE11111166>
- Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121(2), 311–324. [https://doi.org/10.1016/S0378-3758\(03\)00111-3](https://doi.org/10.1016/S0378-3758(03)00111-3)
- Walsh, E. S., Kreakie, B. J., Cantwell, M. G., & Nacci, D. (2017). A Random Forest approach to predict the spatial distribution of sediment pollution in an estuarine system. *PLoS ONE*, 12(7), 1–18. <https://doi.org/10.1371/journal.pone.0179473>
- Wang, Y., Wu, G., Deng, L., Tang, Z., Wang, K., Sun, W., et al. (2017). Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. *Scientific Reports*, 7(1), 1–10. <https://doi.org/10.1038/s41598-017-07197-6>
- West, P. C., Gerber, J. S., Engstrom, P. M., Mueller, N. D., Brauman, K. A., Carlson, K. M., et al. (2014). Leverage points for improving global food security and the environment. *Food Security*, 345(6194), 1–5.
- Wright, M. N., & Ziegler, A. (2017). Ranger: A Fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1), 1–17. <https://doi.org/10.18637/jss.v077.i01>
- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society*, 65(1), 95–114. <https://doi.org/10.1111/1467-9868.00374>
- Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516), 1548–1563. <https://doi.org/10.1080/01621459.2016.1180986>
- Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Hall/CRC.
- Xia, H., Ding, Y., & Wang, J. (2008). Gaussian process method for form error assessment using coordinate measurements. *IIE Transactions (institute of Industrial Engineers)*, 40(10), 931–946. <https://doi.org/10.1080/07408170801971502>
- Yin, X., Goudriaan, J., Lantinga, E. A., Vos, J., & Spiertz, H. J. (2003). A flexible sigmoid function of determinate growth. *Annals of Botany*, 91(3), 361–371. <https://doi.org/10.1093/aob/mcgg029>
- Zuur, A. F., & Camphuysen, Kees, C. (2012). 6 Generalized Additive Models applied on northern gannets. In *A Beginners Guide to Generalized Additive Models with R* (p. 15): Highland Statistics Ltd.

Authors and Affiliations

Paul B. Hegedus^{1,3}  · **Bruce D. Maxwell**^{1,4}  · **Taro Mieno**^{2,5} 

Bruce D. Maxwell
bmax@montana.edu

Taro Mieno
tmieno2@unl.edu

¹ Department of Land Resources and Environmental Sciences, Montana State University, 334 Leon Johnson Hall, Bozeman, MT 59717, USA

² Department of Agricultural Economics, University of Nebraska-Lincoln, 102 Filley Hall, Lincoln, NE 68583, USA

³ Montana State University, 717 Leon Johnson Hall, Bozeman, MT 59717, USA

⁴ Montana State University, 718 Leon Johnson Hall, Bozeman, MT 59717, USA

⁵ University of Nebraska-Lincoln, 209 Filley Hall, Lincoln, NE 68583, USA