Agricultural and Forest Meteorology 311 (2021) 108668

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Unraveling uncertainty drivers of the maize yield response to nitrogen: A Bayesian and machine learning approach

Adrian A. Correndo [a,*], Nicolas Tremblay [b], Jeffrey A. Coulter [c], Dorivar Ruiz-Diaz [a], David Franzen [d], Emerson Nafziger [e], Vara Prasad [f], Luiz H. Moro Rosso [a], Kurt Steinke [g], Juan Du [h], Carlos D. Messina [i], Ignacio A. Ciampitti [a,*]

[a] Department of Agronomy, Kansas State University, 1712 Claflin Rd., Manhattan, KS 66506, USA.
[b] Agriculture and Agri-Food Canada, Horticultural R&D Center, St-Jean-sur-Richelieu J3B 3E6, Canada
[c] Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, Minnesota 55108, USA.
[d] Department of Soil Science, School of Natural Resource Sciences, North Dakota State University, Dept. 7180, PO Box 6050, Fargo, ND 58108, USA.
[e] Department of Crop Sciences, University of Illinois, 1102 S. Goodwin Ave., Urbana, IL 61821, USA.
[f] Sustainable Intensification Lab, Kansas State University, Manhattan, KS 66506, USA.
[g] Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA.
[h] Department of Statistics, Kansas State University, Manhattan, KS 66506, USA.
[i] Corteva Agriscience, 7250 NW 62nd Ave., Johnston IA 50310, USA

ARTICLE INFO

Keywords:
Uncertainty
Fertilizer N response
Maize
Bayesian models
Response curves

ABSTRACT

Development of predictive algorithms accounting for uncertainty in processes underpinning the maize (*Zea Mays* L.) yield response to nitrogen (N) are needed in order to provide new N fertilization guidelines. The aims of this study were to unravel the relative importance of crop management, soil, and weather factors on both the estimate and the size of uncertainty (as a risk magnitude assessment) of the main components of the maize yield response to N: i) yield without N fertilizer (B0); ii) yield at economic optimum N rate (YEONR); iii) EONR; and iv) the N fertilizer efficiency (NFE) at the EONR. Combining Bayesian statistics to fit the N response curves and a machine learning algorithm (extreme gradient boosting) to assess features importance on the predictability of the process, we analyzed data of 730 response curves from 481 site-years (4297 observations) in maize N rate fertilization studies conducted between 1999 and 2020 in the United States and Canada. The EONR was the most difficult attribute to predict, with an average uncertainty of 50 kg N ha^{-1}, increasing towards low (<100 kg N ha^{-1}) and high (>200 kg N ha^{-1}) EONR expected values. Crop management factors such as previous crop and irrigation contributed substantially (~50%) to the estimation of B0, but minorly to other components of the maize yield response to N. Weather contributed about two-thirds of explained variance of the estimated values of YEONR, EONR, and NFE. Additionally, weather factors governed the uncertainty (72% to 81%) of all components of the N response process. Soil factors provided a consistent but limited (10% to 23%) contribution to explain both expected N response as well as its associated uncertainties. Efforts to improve N decision support tools should consider the uncertainty of models as a type of risk, potential in-season weather scenarios, and develop probabilistic frameworks for improving this data-driven decision-making process of N fertilization in maize crop.

1. Introduction

Nitrogen (N) is probably the most complex plant nutrient to study due to an intricate set of spatio-temporal interactions governing plant growth dynamics, soil biogeochemical cycling, and environmental effects on the plant-soil system (Mesbah et al., 2017; Briat et al., 2020). Despite decades of research, addressing the uncertainty on the growth and demand of N in maize (*Zea mays* L.) is still a major concern (Babcock, 1992; Morris et al., 2018, Raun et al., 2019), as indicated by the collateral impacts of misuse of fertilizer and low N use efficiency due to uncertainty of fertilizer rate needed (Sela et al., 2018a,b). Refining the management of a complex system such as the one governing soil-plant N

* Corresponding authors.
E-mail addresses: correndo@ksu.edu (A.A. Correndo), ciampitti@ksu.edu (I.A. Ciampitti).

https://doi.org/10.1016/j.agrformet.2021.108668
Received 14 May 2021; Received in revised form 6 August 2021; Accepted 1 October 2021
Available online 21 October 2021
0168-1923/

dynamics requires understanding the processes generating the yield response to N using multiple perspectives (Lory and Scharf, 2003; Martinez-Feria et al., 2018; Correndo et al., 2021a).

Yield responses to N fertilizer are often modeled using non-linear regression models, which are considered a practical way to provide summaries of the N response. Field trials with various fertilizer N rates are used to estimate optimum rates, mostly under the economic return criteria. Under this scenario, uncertainty is inevitable due to the multiple interactions between the crop with the agronomic management, soil processes, and weather factors (Kyveryga et al., 2007). The degree of uncertainty on the parameters describing the N response functions (*e.g.*, intercept, slope, curvature) and derived quantities (*e.g.*, intercept, optimum N rate, maximum yield, efficiencies) represent a measure of risk (Babcock, 1992), however, they are typically overlooked in the scientific literature (Hernandez and Mulla, 2008). The unpredictable nature that environment has on N dynamics and crop yield dictate the need for models accounting for stochastic components (Tmusiime et al., 2011; Raun et al., 2019).

From the statistical standpoint, the N response curves have been mostly studied using a frequentist approach, for which only the data are considered random, and unknown parameters of interest are treated as fixed variables. In contrast, the Bayesian approach treats the unknown model parameters and derived quantities as random variables. Within the Bayesian framework, we aim to estimate the best model parameters given two main components: i) prior knowledge of the process of interest (the N response curve), and ii) the available observed data (Wakefield, 2013). Literature or expert knowledge is used to define prior distributions of model parameters, and the data are used as new evidence to update our prior beliefs through inferences based on probability distributions (posteriors). Therefore, not only the estimates of the parameters but also their uncertainties are components of interest in a Bayesian framework. Given the increasing computational power and development of new algorithms, Bayesian methods are becoming more common, and are increasingly being used in agricultural research (Lacasa et al., 2020; Ciampitti et al., 2021; Laurent et al., 2020).

Machine learning algorithms are suitable to identify complex association patterns in large datasets (Jordan and Mitchel, 2015). Belonging to the family of supervised learning techniques, classification and regression tree-based methods such as decision trees, random forests and boosting are robust and powerful techniques as recently demonstrated in agricultural research (Shahhosseini et al., 2019; Schwalbert et al., 2020). For example, Correndo et al. (2021a) used conditional random forests to compare forecast frameworks for predicting maize yield without N fertilization while evaluating the contribution of management, soil, and weather features on those predictions. Alternatively, boosting methods consist of fitting multiple decision trees to the data, where each tree is *sequentially* grown using the residuals from previous trees (James et al., 2013). Extreme gradient boosting (xgboost; Chen and Guestrin, 2016), one of the implementations of gradient boosting machines (Friedman, 2001), is known as one of the best performing algorithms for both regression and classification problems (Osman et al., 2021; Park and Kim, 2021). Besides its prediction capabilities, xgboost allows the estimation of a permutation-based feature importance, which serves as a useful interpretation tool to examine the decrease in a scoring rule (*e.g.*, mean square error) when features values are randomly shuffled (Breiman, 2001).

The objectives of this research were to study the importance of crop management, soil, and weather factors on both estimate and the magnitude of uncertainty of the main components—a yield without N (B0), economic optimum N rate (EONR), yield at the EONR (YEONR), and N fertilizer efficiency (NFE) at the EONR— describing the maize yield response to N.

2. Materials and methods

2.1. Review

A database was built by including experimental data that met certain requirements as follows:

1) Collected on experiments during the last two decades (1999-2020) in order to reduce the yield variability associated with genetic advancement of yield potential (Woli et al., 2016); 2) Collected from replicated field trials with N treatments in small plots or field strips; 3) A minimum of four N rate treatments, including a control (zero-N) and a maximum rate of at least 168 kg N ha^{-1} in order to limit the chances of N limitation for achieving an environmentally attainable yield maximum; 4) Trials with positive response but without reaching a yield-plateau were removed from the analysis since EONR and YEONR expected values resulted out of the data range and their uncertainties extremely high, and no associations with specific soil or weather conditions were found, indicating a most likely experimental design limitation (data not shown); 5) Absolute yield data; 6) Planting date; 7) Topsoil crop nutrient analysis results and/or soil series; 8) Previous crop and tillage system; 9) Latitude and longitude coordinates, or report the nearest town in order to retrieve archived weather and/or missing soil series data; 10) No manure as treatments or as a past management input. General crop management (hybrid, row spacing, other nutrients, weed and pest management) was assumed to have been chosen to maximize yield under each site-specific condition.

Published manuscripts were the first source of data, accessed using an engine-search in Web of Science® filtering by the following keywords: “corn/maize” and “nitrogen fertilizer” or “nitrogen fertilization” and “United States” and/or “Canada”. In order to reduce publication bias effect (Dickersin and Min, 1993), unpublished data (*e.g.*, dissertations, field reports, unpublished experiments) that met criteria were also included. After filtering and selection processes, 481 site-years distributed across United States and Canada (**Figure 1A**) resulting in 730 N response curves of maize were gathered from 32 different data sources (**Supp. Table 1**), including published and unpublished studies. Grain yield was standardized at a water content of 155 g kg^{-1}, and each yield point at a given N rate was the average value of 3 to 5 replications.

The database used partially coincides with a previous study by Correndo et al., (2021a), who focused solely on developing a predictive algorithm to forecast maize yield without N. However, this current study is dissimilar mainly in the following points: 1) this study pays special attention on assessing the uncertainty of the entire N response process rather than focusing on a single component, such as the maize yield without N (Correndo et al., 2021a); 2) the present work uses function parameters of the N response process (Bayesian regression analysis) and their uncertainties as the object of study, while Correndo et al. (2021a) solely used observed yields without N as the response variable; 3) the current manuscript considers only experiments with a set of fertilizer N rate treatments satisfying certain minimum requirements (481 site-years, 4297 yield observations), while Correndo et al. (2021a) only considered studies presenting observed maize yields without N fertilization (679 site-years, 1031 yield observations); and lastly, 4) this study only uses trials that reported “sowing date” in order to produce more refined weather variables with the planting date as a reference to adjust the weather summaries (**Table 1**), while Correndo et al. (2021a) used only weather variables summarized by calendar months.

2.2. Meta-data

Soil related variables were topsoil (0-15 cm) soil organic matter (SOM, %) and soil texture (clay, silt and sand, %). Soil data were collected from original sources, accessed from authors’ records when not reported in manuscripts, or retrieved from gridded POLARIS soil data engine (Chaney et al., 2016), a raster optimization based on SSURGO data with a spatial resolution of 1 km^2. When SOM data were reported at

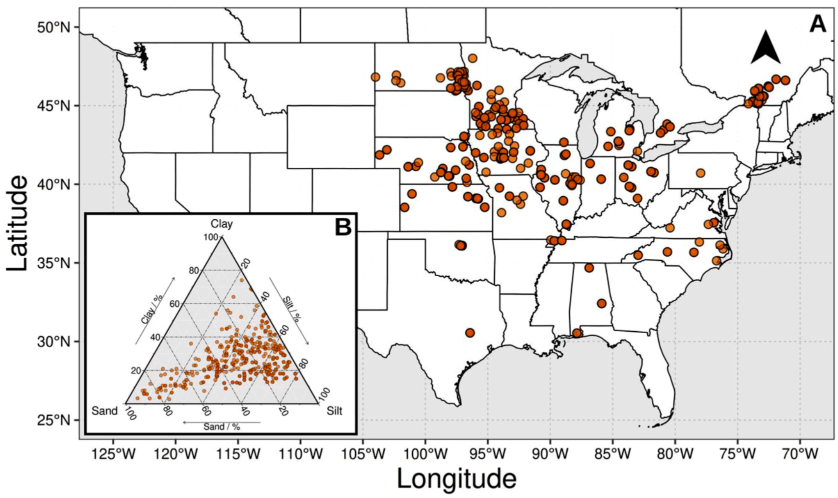

Figure 1. A: Geographical distribution of maize nitrogen fertilization trials under study (730 response curves from 481 site-years) performed in the USA and Canada during the period 1999-2020. B: soil texture distribution (0-15 cm).

Table 1
Meta-data included for studying their influence on selected descriptors of the N response process in maize. *CHU = crop heat units (Tremblay et al., 2012). †das = days after sowing.

MANAGEMENT		
Variable	*Levels*	
Previous Crop	alfalfa; annual legumes (soybean); others (maize, sorghum, wheat, barley, rye, sunflower)	
Tillage system	Tilled; no-till	
Irrigation	Irrigated; Rainfed	
SOIL		
Variable	*Units*	*Depth*
SOM = Soil Organic Matter	%	0-15 cm
Clay		
Silt		
Sand		
WEATHER		
Variable	*Units*	*Periods*
Pp = Precipitations (sum)	mm	pre = -30 – -1 das†, 1 = 0-30 das, 2 = 31-60 das, 3 = 61-90 das, 4 = 91-120 das,
Tm = Mean Temperature (average)	°C	
SDI = Shannon Diversity Index	0-1 (uneven - even)	
EPE = Extreme PP Events (count)	# days PP > 25 mm	
vpd = Vapor Pressure Deficit (average)	KPa	
Rad = Incident radiation (sum)	MJ m^{-2}	
Q = Photothermal quotient	MJ m^{-2} / CHU*	
ETE = Extreme T Events (count)	Days w T_{max} > 30°C	2,3,4

0-20 or 0-30 cm, values were standardized to 0-15 cm using stratification factors based on data from previous research on grain crops trials (Al-Kaisi et al., 2005; Varvel and Wilhelm, 2011; Franzluebbers, 2010; Villamil et al., 2015).

Daily weather data were accessed via the *Daymet* (Thornton et al., 2019) API-client source developed for R-software (package *daymetr*) using reported latitude-longitude coordinates of the trials or nearest town. With a spatial resolution of 1 km^2, precipitation (PP), maximum and minimum temperature (T, °C,), vapor pressure deficit (vpd, kPa), incident shortwave solar radiation during daylight period (Rad, MJ m^{-2}) and day-length. Weather data were transformed into monthly basis (as sum or average) using reported sowing dates as the reference (das: days after sowing). We divided the weather data into five main periods: i) pre-sowing, as the weather of 30 days before sowing; ii) 1st month after sowing (0-30 das), as proxy of the establishment period; ii) 2nd month after sowing (31 to 60 das), as a proxy of the most active growth vegetative period; iii) 3rd month after sowing (61 to 90 das), as proxy of the period around flowering; and iv) 4th month after planting (91-120 dfs), as a proxy of the grain filling period.

A series of additional weather variables were calculated to capture environmental differences that might not have been captured by analyzing standard weather information. For example, the Shannon Diversity Index (SDI) as described by Tremblay et al. (2012) was included to describe the distribution of PP during each period. Extreme PP events were included as the number of days with precipitations greater than 25 mm as a proxy of excessive rainfall events (Puntel et al., 2019; Correndo et al., 2021a). Crop development was described by crop heat units (CHU; Tremblay et al., 2012). Extreme temperature events (ETE, defined as the number of days with maximum temperature greater than 30°C) were also included as a proxy of heat stress risk (Butler and Huybers, 2013; Ye et al., 2017). The photo-thermal quotient (Q) was calculated as the ratio between cumulative Rad and CHU, as an indicator of the solar radiation available to the crop per unit of thermal time during each period, related to yield potential (Bannayan et al., 2018).

2.3. *Data analysis*

2.3.1. *Nitrogen response process*

We fit quadratic and quadratic-plateau regression models using grain yield as the response variable and N rate as the explanatory variable. The quadratic and the quadratic-plateau models are the most extensively used in the literature as they have parameters with a clear interpretation for developing N recommendations (Cerrato and Blackmer, 1990; Wortmann et al., 2011; Kyveryga et al., 2007). Besides its simplicity, the quadratic model presents a great flexibility in terms of possible shapes of the response including flat-, negative or positive linear-, and bell-shaped curves. For each particular case, we selected the model that resulted in the best performance (>R^2 median from Bayesian posteriors). Overall, we have observed a better performance of the quadratic model in the majority of cases (n=513) (**Supp. Figure 1**), with less uncertain estimates, especially in terms of key descriptors of interest such as EONR and NFE (**Supp. Figure 2**). The quadratic-plateau model resulted the best option in 217 cases, particularly when the response curve was very well defined -high R^2 for both models- (**Supp. Figure 1**). In contrast, with less defined N responsiveness patterns, the quadratic-plateau model may result in a more erratic convergence. Since many of the sites used in this study are in areas of the USA where high wind is possible, the quadratic response for maize may also be most appropriate due to yield decrease due to high-wind-induced 'green snap' under high N conditions, while the quadratic-plateau model assumes there is no penalty for greater N rates.

The N response measured through a quadratic model is normally estimated as follows:

$$y_i = B_0 + B_1 x_i - B_2 {x_i}^2 \tag{1}$$

where, for the ith observation, y represents the maize yield (Mg ha^{-1}), x represents the N rate (kg N ha^{-1}), B_0 is the intercept (yield without N fertilizer), B_1 is the linear slope, as the response in yield per unit of change in initial N availability, and B_2 is the quadratic coefficient.

Sharing the same parameters than (1), the N response measured through a quadratic-plateau model is normally estimated as follows:

$$y_i = B_0 + B_1 x_i - B_2 {x_i}^2, if x_i < AONR, \\ B_0 + B_1 AONR - B_2 AONR^2, if x_i \geq AONR \tag{2}$$

where, $AONR$ stands for the agronomic optimum N rate, which corresponds to the level of x_i when the first derivative of the function is equal to zero (AONR = B_1 / $2B_2$)).

Once the model was selected, we considered four main descriptors of the N response process (**Figure 2:** i) B0 (yield without N fertilizer); ii) the grain yield at the economic optimum N rate (YEONR, Mg ha^{-1}); iii) the EONR (kg N ha^{-1}); and iv) N fertilizer efficiency at the EONR (NFE, kg yield (kg applied N)$^{-1}$). The B0 was estimated as the intercept of the response curves. The YEONR and the EONR were estimated as the level of y, and x, respectively, when the first derivative of Eq. (1) is equal to the nitrogen:maize prices ratio (Mg grain kg N^{-1}). Lastly, the NFE was estimated as the quotient between the N responsiveness (YEONR - B0) and the EONR.

A novelty of this analysis was to include an uncertainty component associated with maize grain and fertilizer N prices. For this purpose, we considered the prices ratio as a random variable. Thereby, each time the EONR was estimated, instead of considering a fixed ratio, the value was sampled from a probability distribution. Including a gamma prior distribution into the Bayesian framework, we simulated the historical prices ratio variability observed during the period 1998-2018 (USDA-ERS, 2021a,(USDA-ERS 2021b)). For maize grain we considered the future price at each April, whereas for fertilizer N we considered the average price of anhydrous ammonia (82-0-0) and urea (46-0-0) at each April or March. The historical average prices of maize were 152 $ Mg grain^{-1} and 0.763 $ kg N^{-1}, with an average prices ratio of 0.0053 Mg grain kg N^{-1} (standard deviation of 0.0014 Mg grain kg N^{-1}). With a mean of 0.0055 kg grain kg N^{-1} and a standard deviation of 0.0016 Mg grain kg N^{-1}, the simulated prior for the prices ratio (~ *gamma* (shape = 11, rate = 2)) showed a distribution equivalent to the actual historical PR variability (**Supp. Figure 3**).

2.3.2. Bayesian N response models

The four descriptors of interest were obtained by fitting the simple quadratic or the quadratic-plateau regression model under a hierarchical Bayesian framework using the following priors:

$$y_i \sim Gaussian\left(\mu_i, \sigma_i^2\right) \tag{3}$$

$$u_{i-SQ} = B_0 + B_1 x - B_2 x^2 \tag{4a}$$

$$u_{i-QP} = B_0 + B_1 x_i - B_2 {x_i}^2, if\ x_i < AONR, \\ B_0 + B_1 AONR - B_2 AONR^2, if\ x_i \geq AONR \tag{4b}$$

$$B_0 \sim U(0, 18); \tag{5}$$

$$B_1 \sim U(0, 0.2); \tag{6}$$

$$B_2 \sim gamma(1, 10); \tag{7}$$

$$\sigma_i^2 \sim gamma(2, 2); \tag{8}$$

where for each trial, y_i represents the yield at the ith N rate, u_i represents the underlying process (4a if simple quadratic, 4b if quadratic-plateau), σ_i^2 is the variance of the process, and *Gaussian, U*, and *gamma* stands for normal, uniform, and gamma distributions for priors. Weakly informative priors were defined following previous experience on maize observed yield without N fertilization (B_0) (Correndo et al., 2021a), linear response to N of quadratic models (B_1), and curvature (B_2) (Correndo et al., 2021b; Lacasa et al., 2020). Uniformly distributed priors for B_0 and B_1 were used to ensure adaptability of the priors to each case. In the case of B_2 and σ_i^2, gamma priors were used to support positive values of the parameters, similar to Lacasa et al. (2020) in a study maize yield response to plant density. Particularly for σ_i^2, a gamma prior provides a more suitable alternative than uniform priors, which are proven to lead to a positive miscalibration (overestimation) of the variance (Gelman,

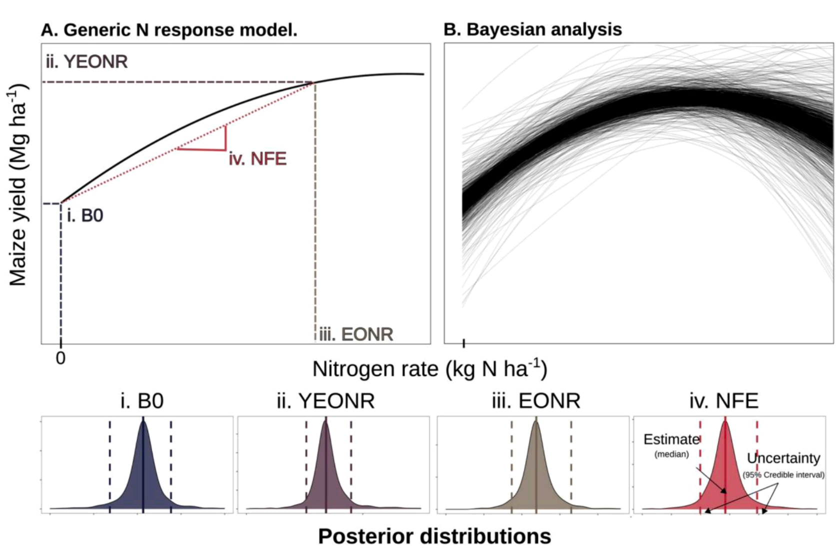

Figure 2. Conceptual representation of the N response process descriptors of interest (A), and the applied Bayesian analysis framework (B) to obtain both estimates and uncertainties from posterior distributions of: i. B0 (yield without N, Mg ha^{-1}), ii) YEONR (yield at EONR, Mg ha^{-1}), iii) EONR (kg N ha^{-1}), and iv) NFE (kg yield response kg N^{-1} at EONR).

2006).

From each model, the expected estimates of the descriptors were retrieved as the median (50th percentile) of the posterior distributions. Similarly, the magnitude of uncertainty for each descriptor was obtained as the length of the 95%-credible intervals (2.5th to 97.5th percentile) from the posterior distributions.

Bayesian models were fit in R-software (R Core Team, 2020), using the *rjags* package v4-10 (Plummer et al., 2019), which applies Gibbs sampling (Geman and Geman, 1984), a Markov Chain Monte Carlo (MCMC) algorithm to generate a sequence of samples approximated to a posterior probability distribution function of parameters. We used 4 parallel chains with 20,000 iterations, including 5,000 as burn-in, and a thinning interval of 10.

2.3.3. Feature importance assessment

In order to reproduce complex association patterns between the descriptors of the N response process and crop management, soil, and weather variables, we applied the xgboost algorithm (Chen and Guestrin, 2016). The target variables were eight, as both the estimate (median) and the uncertainty (95%-credible interval length) of the four N response descriptors: B0, YEONR, EONR, and NFE (**Figure 2**). The model inputs were the crop management, soil, and weather variables described in **Table 1**. Since xgboost only handles numerical matrices, categorical variables such as previous crop, tillage and irrigation were transformed using one-hot-encoding. As a result, previous crop -containing three levels (**Table 1**)-was split into two dummy variables: i) ALF, equal to 1 if previous crop was “alfalfa” (*Medicago sativa* L.), or equal to 0 if not, and ii) LEG, equal to 1 if previous crop was an “annual legume”, equal to zero if previous crop was “other”, otherwise (ALF = 1) always equal to 0.

Since the main purpose of using xgboost here was to assess features importance rather than developing a forecasting model, we considered the entire seasonal weather as if these data were known or perfectly predictable.

For each model, a nested cross-validation (CV) that encompassed the use of an inner-loop for optimization and an outer-loop to assess the generalization performance (Krstajic et al., 2014). We used an outer 20-fold scheme, setting aside a different 5% of observations at a time to be used later as the *testing* data. At the inner loop, a 10-fold-CV was applied over each outer-training set, dividing 90% for training and 10% for validation. For each model, we performed a grid-search to optimize the hyper-parameters of interest: i) *nrounds,* as the number of trees in the forest, ii) *eta*, as the gradient or learning rate, iii) *maxdepth*, as the maximum depth of trees in the forest, iv) *alpha,* as the L1 (LASSO) regularization coefficient, and v) *lambda*, as the L2 (Ridge) regularization coefficient. Regularization through *alpha* and *lambda* was used to reduce the influence of collinearity due to the presence of correlated covariates (**Supp. Figure 4**). We fixed *ncolsamples* at 0.7 (70% of features randomly selected) and *early_stopping_rounds* at 3. The rest of parameters were set to default options. Best combinations were selected based on average performance on the inner-validation set. With the optimized hyper-parameters, performance metrics and features importance were assessed using the outer-training sets (20) to predict the observations on the outer-testing sets. The importance of data input features was quantified using permutation tests (Breiman, 2001).

To evaluate models performance, we used: i) the root mean square error (RMSE, Mg ha^{-1}) as an average squared errors-based statistic that penalizes large residuals; ii) the normalized or relative RMSE (RRMSE, %) as a metric of percentage deviation from the average yield (Yang et al., 2014); iii) the mean bias error (MBE) as the average difference of predicted values with respect to observed, for which positive values mean a systematic over-prediction while negative mean under-prediction; iv) the Nash–Sutcliffe (ME), and v) the Kling-Gupta model efficiencies, as a normalized analog to the coefficient of determination (Nash and Sutcliffe, 1970; Kling et al., 2012); vi) the concordance correlation coefficient (CCC) as a normalized metric that weighs the correlation coefficient (precision) by an index of accuracy (Lin, 1989); and vii) the classical coefficient of determination (R^2) that represents a measure of precision (not accuracy). Formulae of metrics can be found at **Supp. Table 2**. The medians (50th percentile) of each metric based on the 20-folds-CV were selected as their unbiased central-tendency statistic.

3. Results

3.1. Database description

Maize N rate trials under study were distributed across 19 US states (AL, IA, IL, IN, KS, MI, MN, MO, NC, ND, NE, OH, OK, PA, SD, TN, TX, VA, and WI) and two Canada provinces (ON and QC) (**Figure 1A**). The majority of experiments were concentrated during the period 2004-2014 (n=499, 68%), 103 trials were conducted between 1999-2003 (14%), and 128 between 2015-2020 (18%). A total of 601 (82%), and 129 (18%) trials were under rainfed and irrigated conditions, respectively. In terms of tillage management, 466 trials were reported under conventional tillage (64%), and 264 under no-tillage (36%). Previous crops were alfalfa (n=82), annual legumes (n=368), and cereals or others (n=280). Sowing dates (day of the year) varied from days 61 (March 1st) to 155 (June 4th), with a median at day 120, and inter-quartile range (IQR, percentiles 25th to 75th) between days 103 to 136. From sowing to 120 das, mean temperature ranged from 15.7 °C to 25.9 °C, with a median of 20.6 °C (IQR = 19.4 °C – 21.9 °C), and accumulated precipitations ranged from 117 mm to 727 mm, with a median of 420 mm (IQR = 350 mm – 480 mm). Soils represented 11 soil textural classes (Soil Survey Staff, 2018) (**Figure 1B**), and SOM values (%) ranged from 0.5% to 7.9% with a median of 3.6% (IQR = 2.4%-5.2%).

Maximum N rates varied from 168 to 366 kg N ha^{-1}, with a median of 248 kg N ha^{-1}. The reported metadata concerning N fertilization strategy resulted incomplete in the majority of cases. Thus, only 284, 490, and 642, and 284 reported details on fertilizer application form, source, and timing, respectively. Reported forms of N application were broadcasted (n=94), injected (n=176), banded (n=7), and incorporated (n=7). Reported N sources were urea-ammonium nitrate (32-0-0, n=238), ammonium nitrate (34-0-0; n= 169), urea (46-0-0, n=35), calcareous-ammonium nitrate (15-0-0, n=24), and anhydrous-ammonia (82-0-0; n=23). Lastly, reported N application timings were between V2-V6 (n=280), at sowing (n=198), pre-sowing (n=81), split applications between planting and V4-V6 (n=58), and between V7-V9 (25).

Observed yields varied from a minimum of 0.35 Mg ha^{-1} to a maximum of 19.0 Mg ha^{-1}. Yield without N averaged 7.93 (IQR = 5.31-10.24 Mg ha^{-1}), maximum yield averaged 12.23 (IQR = 10.61-14.16 Mg ha^{-1}), and the apparent N responsiveness (maximum yield minus yield without N) averaged 4.30 Mg ha^{-1} (IQR = 2.28-6.20 Mg ha^{-1}).

3.2. Bayesian analysis of N response descriptors

The analysis of the 730 regression curves using the Bayesian approach produced most probable values of the four maize N response descriptors as well as their corresponding uncertainties (**Figure 3A-D**). For B0, the estimates ranged from 0.5 to 17.5 Mg ha^{-1} with a median of 7.9 Mg ha^{-1}, and uncertainty ranged from 0.3 to 4.0 Mg ha^{-1} with a median of 1.4 Mg ha^{-1}. For YEONR, estimates ranged from 1.5 to 19.0 Mg ha^{-1} with a median of 12.4 Mg ha^{-1}, and uncertainty ranged from 0.2 to 8.2 Mg ha^{-1} with a median of 1.6 Mg ha^{-1}. For EONR, estimates ranged from 0 to 368 kg N ha^{-1} with a median of 158 kg N ha^{-1}, and uncertainty ranged from 8 to 261 kg N ha^{-1} with a median uncertainty magnitude of 49 kg N ha^{-1}. Lastly, NFE estimates ranged from 4.2 to 68.8 kg yield kg N^{-1} with a median of 22.1 kg yield kg N^{-1}, and uncertainty ranged from 2.1 to 39.5 kg yield kg N^{-1} with a median uncertainty of 7.1 kg yield kg N^{-1}. The uncertainty in B0, YEONR and NFE showed a poor association with the estimated values (**Figure 3E, F, H**),

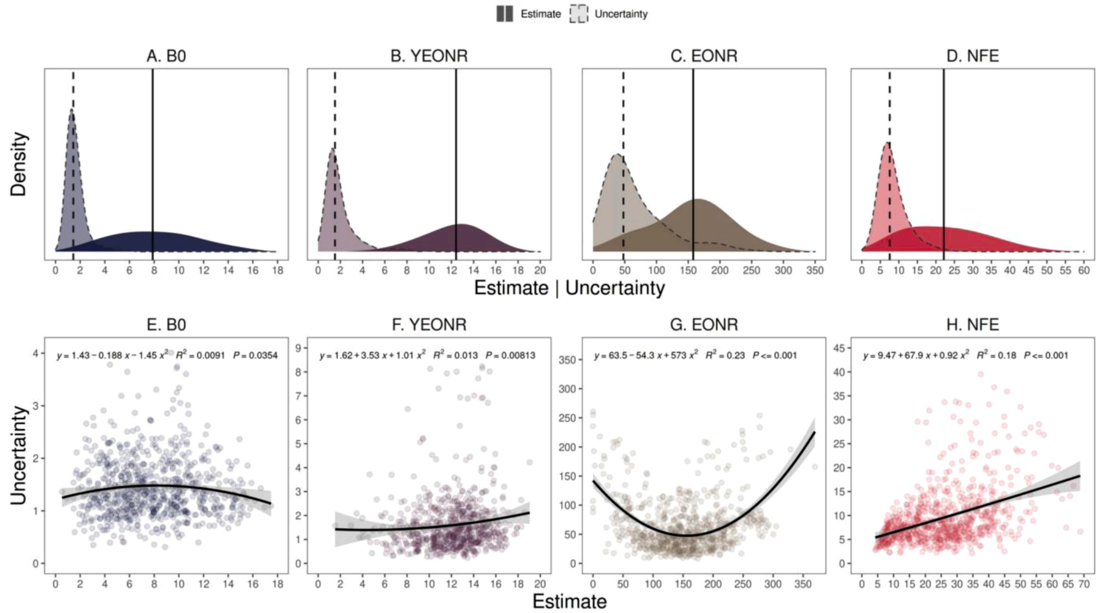

Figure 3. Distribution of estimates and their corresponding uncertainties (A-D) of the four selected maize nitrogen response descriptors (A: intercept (B0, Mg ha^{-1}); B: yield at the economic optimum N rate (EONR)(YEONR, Mg ha^{-1}); C: EONR (kg N ha^{-1}); and D: N fertilizer efficiency to the EONR (NFE, kg yield kg fertilizer N^{-1}); and the relationship between the uncertainty of each descriptor and its estimate (E-H). In A to D, vertical lines indicate the medians of the distributions (solid: estimate, dashed: uncertainty).

while EONR uncertainty was more closely related to estimates, with a trend of higher uncertainties with at both low as well as at high EONR estimated values (Figure 3**G**).

3.3. Features importance

3.3.1. Prediction performance

As expected, the xgboost algorithm showed better performance in predicting estimated values than in predicting uncertainties (Figure 4). The prediction of estimates showed RMSE medians of 1.90 Mg ha^{-1} for B0 (RRMSE = 24%), 1.68 Mg ha^{-1} for YEONR (RRMSE = 14%), 52 kg N ha^{-1} for EONR (RRMSE = 34%), and 10.2 kg yield kg fertilizer N $^{-1}$ for NFE (RRMSE = 40%). The prediction of uncertainties resulted in RMSE of 0.52 Mg ha^{-1} for B0 (RRMSE = 36%), 1.17 Mg ha^{-1} for YEONR (RRMSE = 72%), 44 kg N ha^{-1} for EONR (RRMSE = 69%), and 6.2 kg yield kg fertilizer N $^{-1}$ for NFE (RRMSE = 66%). The rest of the dimensionless metrics, although with different error penalization rules, indicated that the uncertainty magnitudes were much more complex to predict than the estimates (Figure 4). For example, R^2 ranged from 0.36 to 0.71 when predicting estimates, while varied from 0.08 to 0.22 in the

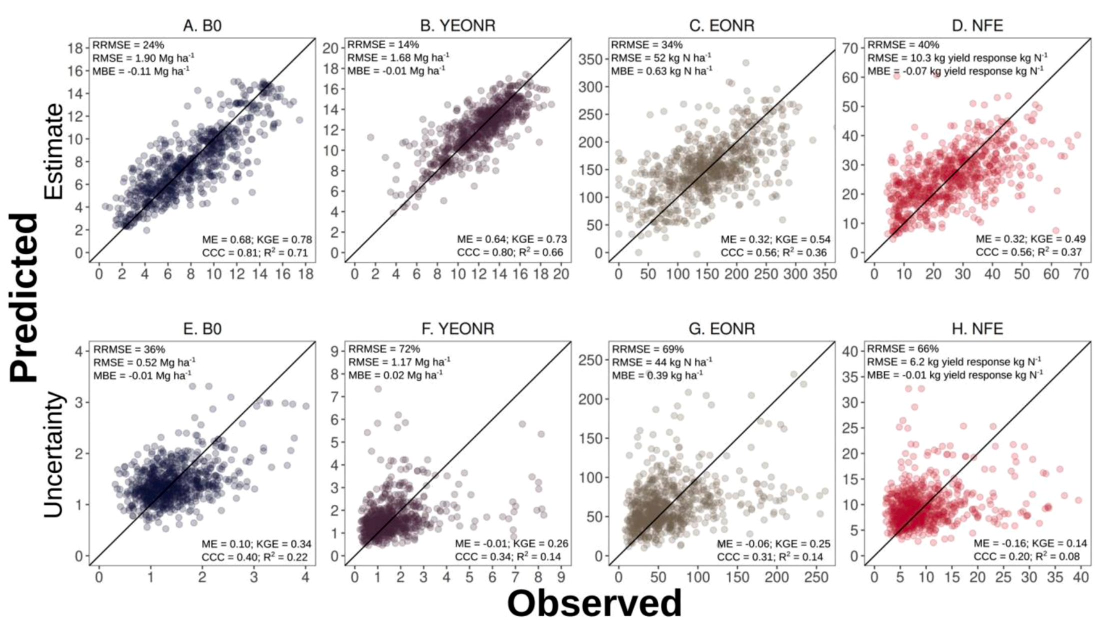

Figure 4. Extreme gradient boosting performance for the prediction of estimates (A-D) and uncertainties (E-H) of four descriptors of the maize nitrogen response process: i) intercept (B0, Mg ha^{-1}), ii) yield at economic optimum nitrogen rate (EONR) (YEONR, Mg ha^{-1}), iii) EONR (kg N ha^{-1}), and iv) nitrogen fertilizer efficiency (NFE = kg yield response kg N^{-1}). Data points are pooled from 20 out-of-bag (OOB) testing samples from cross-validation procedure. Metric values represent the medians of the OOB samples. *RMSE:* root mean square error*; RRMSE:* relative *RMSE; MBE*: mean bias error*; ME:* Nash-Sutcliffe model efficiency*; KGE: Kling-Gupta model efficiency; CCC:* concordance correlation coefficient*; R^2:* coefficient of determination.

case of predicting uncertainties.

3.3.2. Features contribution

The permutation importance test served as an indicator of the relative contribution of features (**Figure 5**) to explained variability by the xgboost algorithm (**Figure 4**). Results indicate that the crop management factors under study were more relevant to predict the estimates rather than the uncertainties of the N response process. Particularly for the estimate of B0, crop management contributed 50% of explained variability (**Figure 4A**), while it only contributed about 1% of explained uncertainty (**Figure 4B**). For YEONR, EONR and NFE estimates, crop management contributed 16%, 19%, and 12%, respectively, of explained variance. In contrast, crop management contributed only 4%, 4%, and 3% of explained variance of YEONR, EONR, and NFE uncertainties, respectively. Regardless of the descriptor estimate or uncertainty, the contribution of soil variables to explained variances was more consistent, ranging from 10% to 23% of explained variance of the N response. Lastly, and as expected, the contribution of weather variables to explained variance was more relevant for prediction of uncertainties than for prediction of estimates. Regardless of the N response descriptor, weather contributed from 72% to 81% of explained variance of uncertainties. In the case of estimates, weather was particularly useful for the prediction of YEONR (64%), EONR (67%), and NFE (78%) components.

Among the crop management components affecting B0 value prediction, previous crop contributed about 37% of explained variance, while irrigation contributed about 13%. In terms of B0 uncertainty, however, previous crop and irrigation showed a negligible influence (**Figure 5**). The most influential weather variable for the B0 expected value was vpd during the pre-sowing period (6%), while precipitations during the late vegetative period (Pp_2) and vpd during grain filling (vpd_4), and clay content, respectively, were the most relevant variables, each contributing about 6% of explained variance of B0 uncertainty. In the case of YEONR estimates, the most relevant feature resulted SOM with ca. 11% of explained variance, while irrigation and previous crop contributed with ca. 14%. Although precipitations and radiation around the flowering period (Pp_3, 6%; Rad_3, 5%) and distribution of precipitations during the late-vegetative period (SDI_2, 3%) resulted among the most important weather features, importance patterns highlight an evenly distributed contribution of evaluated weather variables. Similarly, although precipitations during late vegetative period (PP_2) resulted the most important weather variable (ca. 6%), most of features evenly contributed to explain YEONR uncertainty. Lastly, soil variables contributed with ca. 23% of explained variance, with silt (9%) and SOM (6%) as the most important features, while crop management variables showed an insignificant contribution to YEONR uncertainty (5%).

Previous crop was the most important variable to predict expected EONR values, explaining about 17% of variance. Nonetheless, as stated above, crop management showed a trivial contribution to explain EONR uncertainty. The silt fraction contributed with ca. 5% of explained variance of EONR estimates and ca. 10% of EONR uncertainty. However, we were not able to observe a clear set of most important weather variables defining either estimates or uncertainties, denoting the complex association patterns involving EONR. Similarly, previous crop exerted the most important influence on NFE estimates (ca. 11%), however, the rest of evaluated variables evenly contributed to predict NFE. In terms of the NFE uncertainty, as the most important variables, precipitations around grain filling period (Pp_4) contributed with ca. 5% of explained variance, while clay, SOM, and clay contributed with ca. 5%, 4%, and 4% of explained NFE-uncertainty variance, respectively.

4. Discussion

This study provides an unprecedented assessment of the N response in maize, combining Bayesian statistics with machine learning to unraveling the contribution of crop management, soil, and weather factors to the prediction of both the expected response and its related uncertainties. Highlighting the stochastic nature of the process, this work offers a decomposition of the N response into simple and interpretable components (**Figure 2A**). In the scientific literature, scarce attempts can be highlighted related to considering the parameters of the yield response to N supply as random variables (Hernandez and Mulla, 2008; Tembo et al., 2008; Boyer et al., 2013). However, none of the existing literature on this topic has addressed the investigation of the drivers behind the uncertainty magnitude in the estimated parameters of

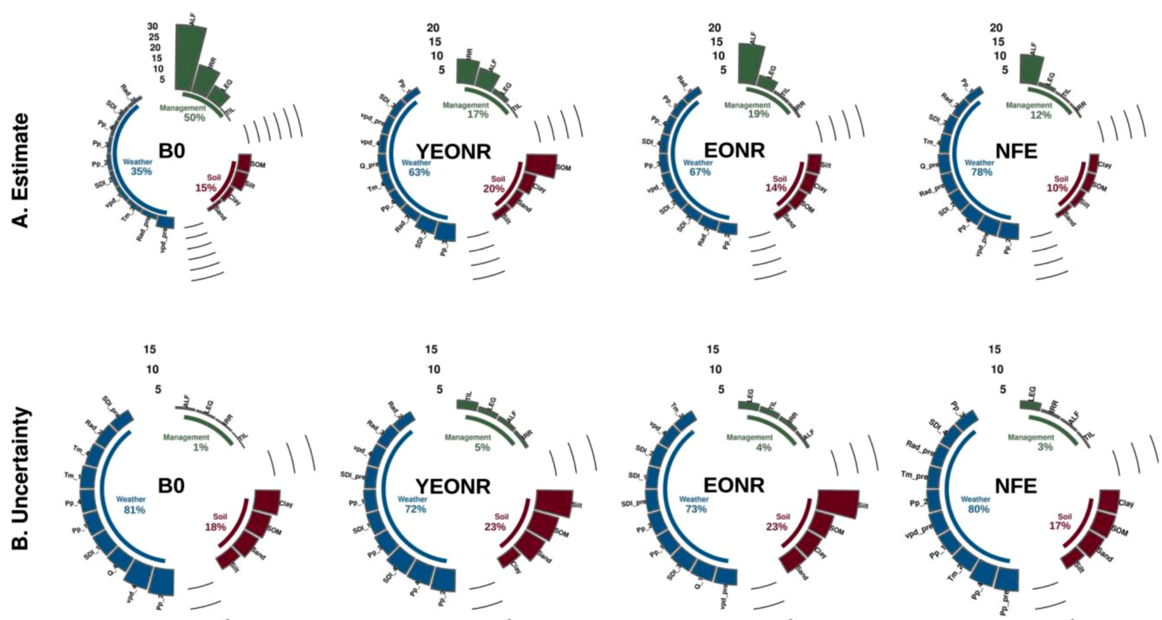

Figure 5. Relative contribution (%) of crop management, soil, and weather variables (10 most important) to expected estimates (A) and uncertainties (B) of main descriptors of the maize nitrogen responsiveness process: i) intercept (B0), ii) yield at the economic optimum N rate -EONR- (YEONR), iii) EONR, and iv) nitrogen fertilizer efficiency (NFE = (YEONR – B0) / EONR).

the maize yield to N supply responses for a given environment.

Improving the awareness of the uncertainties is critical to convey wise information to stakeholders, moving from a static/experience based to a more dynamic/data-driven decision-making process. Similar insights have been recently offered by Laurent et al. (2020) when discussing the benefits of reporting credibility intervals and probability of mean effect size for on-farm network trials. Enhancing the capability of current models to capture uncertainty and to provide sensitivity analysis is the foundation for deploying Bayesian frameworks (Makowski et al., 2004; van de Schoot et al., 2014) to become one of the new pillars for the improved crop N recommendation systems around the globe.

A valuable novelty in our approach is considering the stochastic nature of prices when estimating the EONR. In this regards, although the uncertainties in maize and fertilizer N prices are both major factors deciding fertilization strategy (Blackmer and Kyveryga, 2012), the clear majority of the literature studying the EONR only considered fixed prices for simplicity (*e.g.*, Scharf et al., 2005; Kyveryga et al., 2007; Alotaibi et al., 2018), missing a relevant random component when developing N guidelines. Hence information on historical series of prices at local and/or regional levels should be considered when estimating the EONR (Yost et al., 2014; Nigon et al., 2019). From an economic standpoint, an *ex-ante* approach is the most adequate when estimating EONR (Bullock and Bullock, 2000; Hernandez and Mulla, 2008), for which the evaluation of uncertainties is crucial. In addition to model the uncertainty on the yield response components, we have demonstrated that employing Bayesian statistics also allows to model the variability on the prices ratio by using available historical prices data as a prior.

From the main factors linked to the estimates of the maize N response curves, previous crop (alfalfa) and irrigation have been already identified as critical for B0 (Correndo et al., 2021a) mainly due to the effect of soil N supply and soil physical conditions (Yost et al., 2014; Riedell, 2014) and water as critical factor limiting productivity for field crops and maize (Mueller et al., 2012; Meng et al., 2016). Likewise, for YEONR, water and previous crop were quite relevant from a management standpoint, but with a larger role of SOM with a minor influence of clay as key soil factors for attaining high yields (Lal, 2004; Tremblay et al., 2012). For both EONR and NFE, the influence of weather on the estimation of these factors is more relevant in agreement with previous reports highlighting the impact of this factor on N supply and demand (Soufizadeh et al., 2018).

This study provides relevant insights on the importance of weather (72 to 81%) for improving forecast models that enhance our ability to predict the uncertainties in N response. From a weather standpoint, the most relevant features were evident 60 days after planting time. On the one hand, this emphasizes the importance of further improving the in-season diagnosis tool for crop N status (Scharf et al., 2011). In that regard, combining sensor data with machine learning techniques appears as a promising approach (Wang et al., 2021a). On the other hand, this denotes the need for improving our understanding of future weather conditions and our ability of developing probability scenarios (using historical weather as a proxy)when deciding optimal N rates to be applied in our diverse (*e.g.*, different crop rotation, tillage, management) farming systems. Lack of adequate spatial resolution weather data is a large constraint not only for developing more precise forecasts but also for improving relevant decision support tools (Van Wart et al., 2015). The main challenges for estimating the economic production potential for large field regions is generally linked to the uncertainty of weather forecasts and changes of agricultural landscapes (Jones et al., 2000). In addition, the uncertainty of weather data and its interpolation greatly depend on the density and distribution of weather stations within a region (Mourtzinis et al., 2017). However, future challenges for weather data are mainly connected to the ease of access, data quality, and comprehensiveness/evaluation for this information for relevant use on decision tools and research in agriculture (Overpeck et al., 2011).

The minor role of the evaluated crop management factors (previous crop, irrigation and tillage) on predicting the uncertainties of the yield response to N implies the need of testing the relevancy of other practices (*e.g.*, fertilization strategy, hybrid, plant density) for explaining residual variance. Our results also remark a consistent but limited contribution of the evaluated soil features to understand the uncertainty of N response process. Nonetheless, the inference related to SOM and soil texture fractions results limited since they are not direct and perhaps inconsistent estimators of soil N mineralization (Schomberg et al., 2009; Ros et al., 2011; Cai et al., 2016) and they may also carry confounding effects regarding geographical differences. Thus, further soil indicators regarding soil N supply may be valuable inputs for N response prediction models (Franzluebbers, 2018; McDaniel et al., 2020). Both results may reflect limitations in our approach. As described in Correndo et al. (2021a), our database presents unevenly reported metadata and a lack of relevant features such as pre-plant and/or in-season soil N supply, plant growth determinations, N concentration in tissues or overall plant N uptake, maturity dates, genetic material, among other data descriptors that could ultimately result on improved prediction performance.

Historically, fertilizer N recommendations have been based on estimated production functions treated as the “true” underlying model, largely ignoring the inherent uncertainty existent in any relationship and its associated errors. The estimated uncertainties in our research serve as a measure of risk magnitude when modeling the yield response functions to generate N recommendations. Our main finding is that the EONR presents an significant inherent uncertainty, typically about 50 kg N ha^{-1}, with an increasing risk of erratic estimates at both low ($<$ 100 kg N ha^{-1}) and high ($>$ 200 kg N ha^{-1}) expected EONR values. Empirical evidence indicate that the majority of US maize farmers prefer in-field management strategies as a method of adapting to climate-based risk (Mase et al., 2017). Thus, producers commonly consider N fertilizer as a risk reducing factor (Babcock, 1992; Scharf et al., 2005). In other words, the uncertainty derived from weather or soil N supply most likely leads to increased N rates as risk-neutral farmers perceive profitable to reduce the probability of being caught short of N. Notwithstanding, our evidence indicates that increasing the N fertilizer rate “just in case” (without a clear rationale) under high EONR uncertainty would result both environmentally and economically riskier.

Lastly, this study also provides insights on the opportunity to model the uncertainties of the yield response functions, with the challenge of enhancing the quality of in-season weather forecasts and generating robust prediction frameworks. Crop simulation tools such as DSSAT (Jones et al., 2003), APSIM (Holzworth et al., 2018), or more specific models such as Adapt-N (Melkonian et al., 2008; (Sela et al., 2016)) combining crop management, soil, and in-season weather data to optimize split applications of N fertilizer are a robust foundation. For example, Adapt-N offers estimates of the uncertainty around the recommended rate and allows to use a set of risk considerations related to market prices and N dynamics. Still, the challenge is to transform the simulation frameworks from deterministic to more probabilistic. Decision support tools focusing on N recommendations should ideally provide potential seasonal weather scenarios and their probabilities to understand the level of risk taking by agronomists, farmers, and stakeholders. The data-fusion approach of integrating observed weather data during the early vegetative period and historical weather to create potential scenarios during the late vegetative and reproductive periods represent a unique opportunity to evaluate risks when deciding the N rate (Wang et al., 2021b). The inclusion of stochastic dominance analysis (studying conditional distributions instead of just means) may also provide valuable insights about key factors to manage risks on N decisions, as it has been made for other production factors such as genetics (Nolan and Santos, 2019).Therefore, still major efforts on risk research should be the main focus when fine-tuning decision tools for input utilization in farmer fields, conducting scenarios for combinations of types of risk based on probabilities of historical data (Pannel, 1997) and/or based on better seasonal weather forecasts.

5. Conclusions

This study provides relevant insights on understanding the estimation of the N response for maize, with the additional component of assessing the level of uncertainty for those parameters of the response models. One of the main conclusions of this work is that the expected values of N response components and, although more challenging, their related uncertainties are both susceptible to be modeled. More precisely, yield without N (B0), YEONR and NFE are the most predictable components of N response, while the biggest difficulties were found for predicting the EONR component. Although challenging, broadly variable and susceptible to change by weather, we foresee that uncertainties can be modeled, especially for the B0 and NFE components.

Weather features contributed with roughly two-thirds of explained variance of YEONR and NFE. In addition, weather variables were, undeniably, the most relevant metadata (72% to 81%) to predict the uncertainty of N response (mainly reflected in the EONR). Crop management factors largely affected the prediction of the expected B0, but slightly influenced all the other parameters of the maize N response model. Likewise, crop management displayed a trivial influence (<5%) on the uncertainty of the N response components. Soil factors exerted a consistent but limited contribution to explain both expected N response as well as their uncertainties.

Overall, this research suggests that improvement on the decision support tools should consider the uncertainty of yield response to N supply models as a type of risk, potential in-season weather scenarios, and develop probabilistic frameworks for improving the data-driven decision-making process for N fertilization in maize. The combination of improved modeling approaches along with artificial intelligence tools and advance statistical frameworks (*e.g.*, Bayesian) can provide more dynamic options for N management in maize and other major field crops.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2021.108668.

References

Al-Kaisi, M.M., Yin, X., Licht, M.A., 2005. Soil carbon and nitrogen changes as influenced by tillage and cropping systems in some Iowa soils. Agric. Ecosyst. Environ. 105, 635–647. https://doi.org/10.1016/j.agee.2004.08.002.

Alotaibi, K.D., Cambouris, A.N., St. Luce, M., Ziadi, N., Tremblay, N., 2018. Economic optimum nitrogen fertilizer rate and residual soil nitrate as influenced by soil texture in corn production. Agron. J. 110, 2233–2242. https://doi.org/10.2134/agronj2017.10.0583.

Babcock, B., 1992. The effects of uncertainty on optimal nitrogen applications. Rev. Agric. Econ. 14 (2), 271–280. https://doi.org/10.2307/1349506.

Bannayan, M., Hoogenboom, G., Crout, N.M.J., 2018. Photothermal impact on maize performance: a simulation approach. Ecological Modelling 180 (2-3), 277–290. https://doi.org/10.1016/j.ecolmodel.2004.04.022.

Blackmer, T.M., Kyveryga, P., 2012. Digital imagery guides statewide nutrient management survey. International Society of Precision Agriculture. In: Khosla, R. (Ed.), Proc. 11th Int. Conf. on Precision Agriculture. Indianapolis, IN, pp. 15–18. July.

Boyer, C.N., Larson, J.A., Roberts, R.K., McClure, A., Tyler, D.D., Zhou, V., 2013. Stochastic corn yield response functions to nitrogen for corn after corn, corn after cotton, and corn after soybeans. J. Agric. Applied Econ. 45 (4), 1–12.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Briat, J.F., Gojon, A., Plassard, C., Rouached, H., Lemaire, G., 2020. Reappraisal of the central role of soil nutrient availability in nutrient management in light of recent advances in plant nutrition at crop and molecular levels. Eur. J. Agron. 116, 126069 https://doi.org/10.1016/j.eja.2020.126069.

Bullock, D.S., Bullock, D.G., 2000. From Agronomic Research to Farm Management Guidelines: A Primer on the Economics of Information and Precision Technology, 2. Precision Agriculture, pp. 71–101.

Butler, EE, Huybers, P., 2013. Adaptation of US maize to temperature variations. Nat. Clim. Chang. 3N, 68–72. https://doi.org/10.1038/nclimate1585.

Cai, A., Xu, H., Shao, X., Zhu, P., Zhang, W., Xu, M., Murphy, D.V., 2016. Carbon and Nitrogen Mineralization in Relation to Soil Particle-Size Fractions after 32 Years of Chemical and Manure Application in a Continuous Maize Cropping System. PloS one 11 (3), e0152521. https://doi.org/10.1371/journal.pone.0152521.

Cerrato, M.E., Blackmer, A.M., 1990. Comparison of models for describing corn yield response to fertilizer nitrogen. Agron. J. 82, 138–143. https://doi.org/10.2134/agronj1990.00021962008200010030x.

Chaney, N., Minasny, B., Herman, J., Nauman, T., Brungard, C., Morgan, C., Mcbratney, A., Wood, E., Yimam, Y., 2016. POLARIS soil properties: 30-meter probabilistic maps of soil properties over the contiguous United States. Water Resources Research. https://doi.org/10.1029/2018WR022797.

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, San Francisco, CA, USA, 13–17 August 2016. New York, NY, USA. ACM, pp. 785–794.

Ciampitti, I.A., Fernandez, J., Tamagno, S., Zhao, B., Lemaire, G., Makowski, D., 2021. Does critical N dilution curves for maize crop vary across genotype x environment x management scenarios? - A Bayesian analysis. Eur. J. Agron. 123, 126202 https://doi.org/10.1016/j.eja.2020.126202.

Correndo, A.A., Gutiérrez-Boem, F.H., García, F.O., Alvarez, C., Álvarez, C., Angeli, A., Barbieri, P., Barraco, M., Berardo, A., Boxler, M., Calviño, P., Capurro, J.E., Carta, H., Caviglia, O., Ciampitti, I.A., Díaz-Zorita, M., Díaz-Valdéz, S., Echeverría, H.E., Espósito, G., Ferrari, M., Ferraris, G.N., Gambaudo, S., Gudelj, V., Ioele, J.P., Melchiori, R.J.M., Molino, J., Orcellet, J.M., Pagani, A., Pautasso, J.M., Reussi Calvo, N.I., Redel, M., Rillo, S., Rimski-Korsakov, H., Sainz-Rozas, H.R., Saks, M., Tellería, M.G., Ventimiglia, L., Zorzín, J.L., Zubillaga, M.M., Salvagiotti, F., 2021b. Attainable yield and soil texture as drivers of maize response to nitrogen: A synthesis analysis for Argentina. Field Crops Res 273, 108299. https://doi.org/10.1016/J.FCR.2021.108299.

Correndo, A.A., Rotundo, J.L., Tremblay, N., Archontoulis, S., Coulter, J.A., Ruiz-Diaz, D., Franzen, D., Franzluebbers, A.J., Nafziger, E., Schwalbert, R., Steinke, K., Williams, J., Messina, C.D., Ciampitti, I.A., 2021a. Assessing the uncertainty of maize yield without nitrogen fertilization. Field Crops Res 260. https://doi.org/10.1016/j.fcr.2020.107985.

Dickersin, K., Min, Y., 1993. Publication bias: the problem that won't go away. Ann N Y Acad Sci 703, 135–146. https://doi.org/10.1111/j.1749-6632.1993.tb26343.x, 1993 Dec 31, discussion 146-8.

Franzluebbers, A., 2010. Depth distribution of soil organic carbon as a signature of soil quality. In: 19th World Congress of Soil Science, Soil Solutions for a Changing World, 1 –6 August 2010, Brisbane, Australia. https://www.iuss.org/19th%20WCSS/Symposium/pdf/2164.pdf.

Franzluebbers, A.J., 2018. Soil-Test Biological Activity with the Flush of CO2: III. Corn Yield Responses to Applied Nitrogen. Soil Sci. Soc. Am. J. 82, 708–721. https://doi.org/10.2136/sssaj2018.01.0029.

Friedman, J., 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics 29 (5), 1189–1232. https://statweb.stanford.edu/~jhf/ftp/trebst.pdf.

Geman, S., Geman, D., 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (6), 721–741. https://doi.org/10.1109/TPAMI.1984.4767596.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Anal 1 (3), 515–534. https://doi.org/10.1214/06-BA117A.

Hernandez, J.A., Mulla, D.J., 2008. Estimating uncertainty of economically optimum fertilizer rates. Agron. J. 100, 1221–1229. https://doi.org/10.2134/agronj2007.0273.

Holzworth, D., Huth, N.I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Verrall, S., Herrmann, N.I., Zheng, B., Snow, V., 2018. APSIM Next Generation: Overcoming challenges in modernising a farming systems model. Environ. Model. Softw. 103, 43–51. https://doi.org/10.1016/j.envsoft.2018.02.002.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. Chapter In: An Introduction to Statistical Learning with Applications in R, 1st ed., 8. Springer, New York. https://doi.org/10.1007/978-1-4614-7138-7.

Jones, J.W., Hansen, J.W., Royce, F.S., Messina, C.D., 2000. Potential benefits of climate forecasting to agriculture. Agric. Ecosyst. Environ. 82 (1-3), 169–184. https://doi.org/10.1016/S0167-8809(00)00225-5.

Jones, J.W, Hoogenboom, G., Porter, C., Boote, K.J., Batchelor, W.D., Hunt, L.A., 2003. DSSAT cropping system model. Eur. J. Agron. 18, 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. Science 349, 255–260. https://doi.org/10.1126/science.aaa8415.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J. Hydrol. 424-425, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011.

Krstajic, D., Buturovic, L.J., Leahy, D.E., Thomas, S., 2014. Cross-validation pitfalls when selecting and assessing regression and classification models. J. Cheminformatics 6, 10, 2014. http://www.jcheminf.com/content/6/1/10.

Kyveryga, P., Blackmer, A.M., Morris, T.F., 2007. Disaggregating model bias and variability when calculating economic optimum rates of nitrogen fertilization for corn. Agron. J. 99, 1048–1056. https://doi.org/10.2134/agronj2006.0339.

Lacasa, J., Gaspar, A., Hinds, M., Don, S.J., Berning, D., Ciampitti, I.A., 2020. Bayesian approach for maize yield response to plant density from both agronomic and

economic viewpoints in. North America. Sci. Rep. 10, 15948. https://doi.org/10.1038/s41598-020-72693-1.
Lal, R., 2004. Soil carbon sequestration impacts on climate change and food security. Science 304, 1623–1627. https://doi.org/10.1126/science.1097396.
Laurent, A., Miguez, F., Kyveryga, P., Makowski, D., 2020. Going beyond mean effect size: Presenting prediction intervals for on-farm network trial analyses. Eur. J. Agron. 120, 126127 https://doi.org/10.1016/j.eja.2020.126127.
Lin, L.I.-K, 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 45 (1), 255–268. https://doi.org/10.2307/2532051.
Lory, J.A., Scharf, P.C., 2003. Yield goal versus delta yield for predicting fertilizer nitrogen need in corn. Agron. J. 95, 994–999. https://doi.org/10.2134/agronj2003.0994.
Makowski, D., Jeuffroy, M., Guérif, M., 2004. Bayesian methods for updating crop-model predictions, applications for predicting biomass and grain protein content. Frontis 57–68. https://library.wur.nl/ojs/index.php/frontis/article/view/858.
Martinez-Feria, R.A, Castellano, M.J., Dietzel, R.N., Helmers, M.J., Liebman, M., Huber, I., Archontoulis, S.V., 2018. Linking crop- and soil-based approaches to evaluate system nitrogen-use efficiency and tradeoffs. Agric. Ecosyst. Environ. 256, 131–143. https://doi.org/10.1016/j.agee.2018.01.002.
Mase, A.S., Grami, B.M., Prokopy, L.S., 2017. Climate change beliefs, risk perceptions, and adaptation behavior among Midwestern U.S. crop farmers. Climate Risk Manag 15, 8–17. https://doi.org/10.1016/j.crm.2016.11.004.
McDaniel, M., Walters, D, Bundy, L., Li, X., Drijber, R.A., Sawyer, J.E., Castellano, M.J., Laboski, C.A.M., Scharf, P.C., Horwath, W.R, 2020. Combination of biological and chemical soil tests best predict maize nitrogen response. Agron. J. 112, 1263–1278. https://doi.org/10.1002/agj2.20129.
Melkonian, J.J., van Es, H.M., Joseph, L., 2008. Adapt-N: Adaptive nitrogen management for maize using high-resolution climate data and model simulations. In: Koshla, R. (Ed.), Proceedings of the 9th International Conference on Precision Agriculture. Monticello, IL, USA. Denver, CO. International Society f Precision Agriculture.
Meng, Q., Chen, X., Lobell, D., Cui, Z., Zhang, Y., Yang, H., Zhang, F., 2016. Growing sensitivity of maize to water scarcity under climate change. Sci Rep 6, 19605. https://doi.org/10.1038/srep19605.
Mesbah, M., Pattey, E., Jego, G., 2017. A model-based methodology to derive optimum nitrogen rates for rainfed crops a case study for corn using STICS in Canada. Comput. Electron. Agric. 142, 572–584. https://doi.org/10.1016/j.compag.2017.11.011.
Morris, T.F., Murrell, T.S., Beegle, D.B., Camberato, J.J., Ferguson, R.B., Grove, J., Ketterings, Q., Kyveryga, P.M., Laboski, C.A., McGrath, J.M., Meisinger, J.J., Melkonian, J., Moebius-Clune, B.N., Nafziger, E.D., Osmond, D., Sawyer, J.E., Scharf, P.C., Smith, W., Spargo, J.T., van Es, H.M., Yang, H., 2018. Strengths and limitations of nitrogen rate recommendations for corn and opportunities for improvement. Agron. J. 110, 1–37. https://doi.org/10.2134/agronj2017.02.0112.
Mourtzinis, S., Rattalino-Edreira, J.I., Conley, S.P., Grassini, P., 2017. From grid to field: assessing the quality of gridded weather data for agricultural applications. Eur. J. Agron. 82 (A), 163–172. https://doi.org/10.1016/j.eja.2016.10.013.
Mueller, N., Gerber, J.S., Johnston, M., Ray, D.K., Ramankutty, N., Foley, J.A., 2012. Closing yield gaps through nutrient and water management. Nature 490, 254–257. https://doi.org/10.1038/nature11420.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I - A discussion of principles. J. Hydrol. 10 (3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.
Nigon, T.J., Yang, C., Mulla, D.J, Kaiser, D.E., 2019. Computing uncertainty in the optimum nitrogen rate using a generalized cost function. Comp. Electr. Agric. 167, 105030 https://doi.org/10.1016/j.compag.2019.105030.
Nolan, E., Santos, P., 2019. Genetic modification and yield risk: a stochastic dominance analysis of corn in the USA. PLOS ONE 14 (10), e0222156, 10.137/journal.pone.0222156.
Osman, A.I.A., Ahmed, A.N., Chow, M.F., Huang, Y.F., El-Shafie, A., 2021. Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. Ain. Shams. Eng. J. https://doi.org/10.1016/j.asej.2020.11.011.
Overpeck, J.T., Meehl, G.A., Bony, S., Easterling, D.R., 2011. Climate data challenges in the 21st century. Science 331, 700–702. https://doi.org/10.1126/science.1197869. Issue 618.
Pannel, D.J., 1997. Sensitivity analysis of normative economic models: theoretical framework and practical strategies. Agric. Econ. 16 (2), 139–152. https://doi.org/10.1016/S0169-5150(96)01217-0.
Park, S., Kim, J., 2021. The Predictive Capability of a Novel Ensemble Tree-Based Algorithm for Assessing Groundwater Potential. Sustainability 13 (5), 2459. https://doi.org/10.3390/su13052459.
Plummer, M., Stukalov, A., Denwood, M., 2019. Package 'rjags'. CRAN Repository . 19pp. ftp://journal.r-project.org/pub/R/web/packages/rjags/rjags.pdf.
Puntel, L.A., Pagani, A., Archontoulis, S.V., 2019. Development of a nitrogen recommendation tool for corn considering static and dynamic variables. Eur. J. Agron. 105, 189–199. https://doi.org/10.1016/j.eja.2019.01.003.
Core Team, R, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL. http://www.R-project.org/.
Raun, W.R., Dhillon, J., Aula, L., Eickhoff, E., Weymeyer, G., Figueirdeo, B., Lynch, T., Omara, P., Nambi, E., Oyebiyi, F., Fornah, A., 2019. Unpredictable Nature of Environment on Nitrogen Supply and Demand. Agron. J. 111, 2786–2791. https://doi.org/10.2134/agronj2019.04.0291, 2019.
Riedell, W.E., 2014. Nitrogen fertilizer applications to maize after alfalfa: grain yield, kernel composition, and plant mineral nutrients. J. Plant Nutr. 37 (12), 2026–2035. https://doi.org/10.1080/01904167.2014.911892.
Ros, G.H., Hanegraaf, M.C., Hoffland, E., van Riemsdijk, W.H., 2011. Predicting soil N mineralization: Relevance of organic matter fractions and soil properties. Soil Biology and Biochemistry 43 (8), 1714–1722. https://doi.org/10.1016/j.soilbio.2011.04.017.
Scharf, P.C., Kitchen, N.R., Sudduth, K.A., Davis, J.G., Hubbard, V.C., Lory, J.A., 2005. Field-scale variability in optimal nitrogen fertilizer rate for corn. Agron. J. 97, 452–461. https://doi.org/10.2134/agronj2005.0452.
Scharf, P.C., Shannon, D.K., Palm, H.L., Sudduth, K.A., Drummond, S.T., Kitchen, N.R., Mueller, L.J., Hubbard, V.C., Oliveira, L.F., 2011. Sensor-based nitrogen applications out-performed producer-chosen rates for corn in on-farm demonstrations. Agron. J. 103, 1683–1691. https://doi.org/10.2134/agronj2011.0164.
Schomberg, H.H., Wietholter, S., Griffin, T.S., Reeves, D.W., Cabrera, M.L., Fisher, D.S., Endale, D.M., Novak, J.M., Balkcom, K.S., Raper, R.L., Kitchen, N.R., Locke, M.A., Potter, K.N., Schwartz, R.C., Truman, C.C., Tyler, D.D., 2009. Assessing indices for predicting potential nitrogen mineralization in soils under different management systems. Soil Sci. Soc. Am. J. 73, 1575–1586. https://doi.org/10.2136/sssaj2008.0303.
Schwalbert, R., Amado, T., Nieto, L., Corassa, G., Rice, C., Peralta, N., Schauberger, B., Gornott, C., Ciampitti, I.A., 2020. Mid-season county-level corn yield forecast for US Corn Belt integrating satellite imagery and weather variables. Crop Sci 60, 739–750. https://doi.org/10.1002/csc2.20053.
Sela, S., van Es, H.M., Moebius-Clune, B.N., Marjerison, R., Melkonian, J., Moebius-Clune, D., Schindelbeck, R., Gomes, S., 2016. Adapt-N outperforms grower-selected nitrogen rates in Northeast and Midwestern United States strip trials. Agron. J. 108, 1726–1734. https://doi.org/10.2134/agronj2015.0606.
Sela, S., van Es, H.M., Moebius-Clune, B.N., Marjerison, R., Kneubuhler, G., 2018a. Dynamic model-based recommendations increase the precision and sustainability of N fertilization in midwestern US maize production. Comput. Electron. Agric. 153, 256–265. https://doi.org/10.1016/j.compag.2018.08.010.
Sela, S., Woodbury, P.B, van Es, H.M., 2018b. Dynamic model-based N management reduces surplus nitrogen and improves the environmental performance of corn production. Environ. Res. Lett. 13, 054010 https://doi.org/10.1088/1748-9326/aab908.
Shahhosseini, M., Martinez-Feria, R.A., Hu, G., Archontoulis, S.V., 2019. Maize yield and nitrate loss prediction with machine learning algorithms. Environ. Res. Lett/ 14, 124026. https://doi.org/10.1088/1748-9326/ab5268.
Soufizadeh, S., Munaro, E., McLean, G., Massignam, A., van Oosterom, E.J., Chapman, S. C., Messina, C., Cooper, M., Hammer, G., 2018. Modelling the nitrogen dynamics of maize crops – Enhancing the APSIM maize model. Eur. J. Agron. 100, 118–131. https://doi.org/10.1016/j.eja.2017.12.007.
Tembo, G., Brorsen, B., Epplin, F., Tostão, E., 2008. Crop Input Response Functions with Stochastic Plateaus. American Journal of Agricultural Economics 90 (2), 424–434. http://www.jstor.org/stable/30139594.
Thornton, P.E., Thornton, M., Mayer, B., Wei, Y., Devarakonda, R, Vose, R., Cook, R.B., 2019. Daymet: daily surface weather data on a 1-km Grid for North America, Version3. ORNL DAAC, Oak Ridge, Tennessee, USA. https://daymet.ornl.gov/.
Tremblay, N., Bouroubi, Y.M., Bélec, C., Mullen, R.W., Kitchen, N.R., Thomason, W.E., Ebelhar, S., Mengel, D.B., Raun, W.R., Francis, D.D., Vories, E.D., Ortiz-Monasterio, I., 2012. Corn response to nitrogen is influenced by soil texture and weather. Agron. J. 104, 1658–1671. https://doi.org/10.2134/agronj2012.0184.
Tumusiime, E., Brorsen, B.W., Mosali, J., Johnson, J., Locke, J., Biermacher, J.T., 2011. Determining optimal levels of nitrogen fertilizer using random parameter models. J. Agric. Appl. Econ. 43 (4), 541–552. https://doi.org/10.1017/S1074070800000067.
USDA-ERS, 2021a. Season-average price forecasts. USDA-Economic Research Service, Washington, DC. https://www.ers.usda.gov/data-products/season-average-price-forecasts (last accessed on 30 April 2021).
USDA-ERS, 2021b. Fertilizer use and price statistics. USDA-Economic Research Service, Washington, DC. https://www.ers.usda.gov/data-products/fertilizer-use-and-price/documentation-and-data-sources (last accessed on 30 April 2021).
van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J.B., Neyer, F.J., van Aken, M.A., 2014. A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. Child Dev 85, 842–860. https://doi.org/10.1111/cdev.12169.
Van Wart, J., Grassini, P., Yang, H., Claessens, L., Jarvis, A., Cassman, K., 2015. Creating long-term weather data from thin air for crop simulation modeling. Agric. Forest. Meteor 209-210, 49–58. https://doi.org/10.1016/j.agrformet.2015.02.020.
Varvel, G.E., Wilhelm, W.W., 2011. Soil carbon levels in irrigated Western Corn Belt rotations. Agron. J. 100 (4), 1180–1184. https://doi.org/10.2134/agronj2007.0383.
Villamil, M.B., Little, J., Nafziger, E.D., 2015. Corn residue, tillage, and nitrogen rate effects on soil properties. Soil Tillage Res 151, 61–66. https://doi.org/10.1016/j.still.2015.03.005.
Wang, X., Miao, Y., Dong, R., Zha, H., Xia, T., Chen, Z., Kusnierek, K., Mi, G., Sun, H., Li, M. 2021a. Machine learning-based in-season nitrogen status diagnosis and side-dress nitrogen recommendation for corn. 10.1016/j.eja.2020.126193.
Wang, X., Miao, Y., Batchelor, W.D., Dong, R., Kusnierek, K., 2021b. Evaluating model-based strategies for in-season nitrogen management of maize using weather data fusion. 10.1016/j.agrformet.2021.108564.
Wakefield, J., 2013. Bayesian and Frequentist Regression Methods. Springer Series in Statistics. Springer, New York, New York, NY. https://doi.org/10.1007/978-1-4419-0925-1.
Woli, K.P., Boyer, M.J., Elmore, R.W., Sawyer, J.E., Abendroth, L.J., Barker, D.W., 2016. Corn era hybrid response to nitrogen fertilization. Agron. J. 108, 473–486. https://doi.org/10.2134/agronj2015.0314.
Wortmann, C.S., Tarkalson, D., Shapiro, C., Dobermann, A., Ferguson, R., Hergert, G., Walters, D., 2011. Nitrogen use efficiency of irrigated corn for three cropping systems in Nebraska. Agron. J. 103, 76–84. https://doi.org/10.2134/agronj2010.0189.

Yang, J.M., Yang, J.Y., Liu, S., Hoogenboom, G., 2014. An evaluation of the statistical methods for testing the performance of crop models with observed data. Agric. Syst. 127 (May 2014), 81–89. https://doi.org/10.1016/j.agsy.2014.01.008.

Ye, Q., Lin, X., Adee, E., Min, D., Assefa Mulisa, Y., O'Brien, D., Ciampitti, I.A., 2017. Evaluation of climatic variables as yield-limiting factors for maize in Kansas. International Journal of Climatology 37 (S1), 464–475. https://doi.org/10.1002/joc.5015.

Yost, M.A., Morris, T.F., Russelle, M.P., Coulter, J.A., 2014. Second-Year Corn after Alfalfa Often Requires No Fertilizer Nitrogen. Agron. J. 106, 659–669. https://doi.org/10.2134/agronj2013.0362.