Field Crops Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Field Crops Research

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

Assessing variation in maize grain nitrogen concentration and its implications for estimating nitrogen balance in the US North Central region

Fatima A.M. Tenorio[a], Alison J. Eagle[b], Eileen L. McLellan[b], Kenneth G. Cassman[a], Reka Howard[c], Fred E. Below[d], David E. Clay[e], Jeffrey A. Coulter[f], Allen B. Geyer[g], Darin K. Joos[d], Joseph G. Lauer[h], Mark A. Licht[i], Alexander J. Lindsey[g], Bijesh Maharjan[a], Cameron M. Pittelkow[d], Peter R. Thomison[g], Charles S. Wortmann[a], Victor O. Sadras[j], Patricio Grassini[a,*]

[a] Department of Agronomy and Horticulture, University of Nebraska–Lincoln, Lincoln, NE, 68583-0915, USA
[b] Environmental Defense Fund, Washington, DC, 20009, USA
[c] Department of Statistics, University of Nebraska–Lincoln, Lincoln, NE 68583-0963, USA
[d] Department of Crop Sciences, University of Illinois, Urbana, IL, 61801, USA
[e] Agronomy, Horticulture, and Plant Science Department, South Dakota State University, Brookings, SD, 57007, USA
[f] Department of Agronomy and Plant Genetics, University of Minnesota, MN, 55108-6026, USA
[g] Department of Horticulture and Crop Science, The Ohio State University, OH, 43210, USA
[h] Department of Agronomy, University of Wisconsin-Madison, Madison, WI, 53706, USA
[i] Department of Agronomy, Iowa State University, Agronomy Hall, IA, 50011-1010, USA
[j] South Australian Research and Development Institute (SARDI), Waite Campus, Australia

ARTICLE INFO

Keywords:
Grain nitrogen concentration
Grain nitrogen removal
Nitrogen balance
Maize

ABSTRACT

Accurate estimation of nitrogen (N) balance (a measure of potential N losses) in producer fields requires information on grain N concentration (GNC) to estimate grain-N removal, which is rarely measured by producers. The objectives of this study were to (i) examine the degree to which variation in GNC can affect estimation of grain-N removal, (ii) identify major factors influencing GNC, and (iii) develop a predictive model to estimate GNC, analyzing the uncertainty in predicted grain-N removal at field and regional levels. We compiled GNC data from published literature and unpublished databases using explicit criteria to only include experiments that portray the environments and dominant management practices where maize is grown in the US North Central region, which accounts for one-third of global maize production. We assessed GNC variation using regression tree analysis and evaluated the ability of the resulting model to estimate grain-N removal relative to the current approach using a fixed GNC. Across all site-year-treatment cases, GNC averaged 1.15%, ranging from 0.76 to 1.66%. At any given grain yield, GNC varied substantially and resulted in large variation in estimated grain-N removal and N balance. However, compared with GNC, yield differences explained much more variability in grain-N removal. Our regression tree model accounted for 35% of the variation in GNC, and returned physiologically meaningful associations with mean air temperature and water balance in July (*i.e.*, silking) and August (*i.e.*, grain filling), and with N fertilizer rate. The predictive model has a slight advantage over the typical approach based on a fixed GNC for estimating grain-N removal for individual site-years (root mean square error: 17 *versus* 21 kg N ha^{-1}, respectively). Estimates of grain-N removal with both approaches were more reliable when aggregated at climate-soil domain level relative to estimates for individual site-years.

Abbreviations: ANOVA, analysis of variance; ET_O, grass-based reference evapotranspiration (mm); GNC, grain nitrogen concentration (%); ME, absolute mean error; NIR, near infrared; RMSE, root mean square error; SS, sum of squares; TED, technology extrapolation domain; T_{max}, maximum temperature (ºC); T_{mean}, mean temperature (ºC); T_{min}, minimum temperature (ºC); US, United States

* Corresponding author.
E-mail address: pgrassini2@unl.edu (P. Grassini).

https://doi.org/10.1016/j.fcr.2018.10.017
Received 6 September 2018; Received in revised form 24 October 2018; Accepted 28 October 2018
0378-4290/

Please cite this article as: Tenorio, F.A.M., Field Crops Research, https://doi.org/10.1016/j.fcr.2018.10.017

1. Introduction

Nitrogen (N) fertilizer is an essential input to sustain high cereal yields (Cassman et al., 2002). However, mismatches between N inputs and crop N demand could result in N losses to the environment (Erisman et al., 2013). As a result, there is growing interest in developing cost-effective indicators to evaluate the degree to which N fertilizer inputs are congruent with crop N requirements (Zhang et al., 2015). A simplified N balance, calculated as the difference between N inputs (including fertilizer, manure, symbiotic N_2 fixation, deposition) and grain-N removal, can be used to assess potential for N losses in producer fields (McLellan et al., 2018 and references cited therein). However, estimating N balance depends on the calculation of grain-N removal, and while maize producers usually know the grain yield achieved on each of their fields, they rarely measure grain nitrogen concentration (GNC). Lack of GNC measurements reflects that most maize grain produced in the US is used for livestock feed, and its value derives from its energy rather than its protein content. Some maize crop models (*e.g.*, CERES-Maize; Jones and Kiniry, 1986) can simulate grain-N removal, but they require calibration and copious amounts of data inputs (*i.e.*, daily weather, soil properties, cultivar coefficients, and management practices) to be useful for predicting grain-N removal in individual fields. Additionally, previous studies have shown that these models performed relatively poor at reproducing measured GNC in field-grown maize (*e.g.*, Liu et al., 2010; Yakoub et al., 2017). Hence, at issue is how reliable the estimation of grain-N removal can be in the absence of measured GNC.

Average maize GNC has declined over time as an unintended consequence of breeders' selection for higher yields (Duvick and Cassman, 1999; Ciampitti and Vyn, 2012; DeBruin et al., 2017), and a number of published studies have aimed to understand the associated physiological drivers (Chen and Vyn, 2017 and references cited therein). Early in the 1970s, Welch (1971) used an average GNC of 1.61% to estimate grain-N removal. Later, Boone et al. (1984) reported a mean of 1.33% based on measured data across commercial maize hybrids grown in the Midwestern US at different plant densities. A review paper by Ciampitti and Vyn (2012) reported the same mean GNC of 1.33% for maize hybrids released between 1940 to 1990, with mean GNC decreasing to 1.20% for hybrids released between 1991 and 2011. The GNC values reported here are all expressed at a standard 15.5% moisture content. Besides the long-term decline in GNC, prior studies on maize have reported substantial variation in GNC due to climate and management practices (Viets and Domingo, 1948; Zuber et al., 1954; Genter et al., 1956; Lang et al., 1956; Boone et al., 1984; Feil et al., 1990; Liang et al., 1996). In the absence of measured GNC data, the typical approach is to assume a fixed GNC from the literature. For example, the International Plant Nutrition Institute (IPNI) recommended using an average GNC of 1.2% for estimating grain-N removal in absence of measured data (http://www.ipni.net/article/IPNI-3296). However, the degree to which variation in GNC would affect the estimation of grain-N removal and N balance in individual field has not been explicitly evaluated.

There are many studies aiming to model sources of variation in GNC for winter cereals such as wheat and barley (Correll et al., 1994; Smith and Gooding, 1999; Hansen et al., 2002; Zhao et al., 2005). For instance, Correll et al. (1994) developed a predictive model based on seasonal air temperature and precipitation to explain variation in GNC for wheat and barley in South Australia. Later, Smith and Gooding (1999) reported a model showing that cultivar and N fertilizer rate were also important factors influencing GNC in wheat. Although both environmental and management factors have been reported to influence GNC in maize, no attempt has been made to synthesize and analyze existing GNC data to generate a predictive model for maize GNC. Such a model would be useful for estimating grain-N removal and N balance in producer fields in the absence of directly measured GNC data.

In the present study, we collected existing maize GNC data from experiments conducted across the US North Central region (Fig. 1), which is an area that accounts for *ca.* 33% of global maize production. Only data that portray the range of dominant on-farm management practices and hybrids were used for the analysis. The specific objectives were to (i) examine the degree to which variation in GNC can affect estimation of grain-N removal in maize, (ii) identify major factors influencing GNC and model these sources of variation, and (iii) evaluate an approach to estimate GNC as an alternative to a fixed GNC, analyzing the uncertainty in predicted grain-N removal at field as opposed to regional level.

2. Materials and methods

2.1. Database description and criteria

Published articles and online databases were screened to compile experimental data on GNC from field-grown maize across the US North Central region. Major climate, soil, and management features of maize-based agroecosystems in the US North Central region are described elsewhere (Grassini et al., 2014). The search was restricted to experiments conducted during the 1999–2016 period to represent recent hybrids and management practices. Our database included observations from nine states: Illinois (IL), Indiana (IN), Iowa (IA), Kansas (KS), Minnesota (MN), Nebraska (NE), Ohio (OH), South Dakota (SD), and Wisconsin (WI) (Fig. 1; Supplementary Table S1). Only data from replicated experiments that meet two criteria were included: (i) field-grown grain maize crops managed with current crop and soil management practices in the region, and (ii) reported data on grain yield, GNC, N fertilizer rate, and water regime (irrigated or rainfed). We thus excluded experiments sown for silage or hybrid seed production, with experimental hybrids, with outdated practices (*e.g.*, moldboard plow), or with unrealistic treatments (*e.g.*, N omission plots). Likewise, we excluded experiments in which maize was grown after alfalfa because only a very small fraction of US maize follows alfalfa and potential soil N supply following this perennial legume crop can be large and difficult to calculate. Experiments receiving manure were also excluded given the difficulties to quantify N inputs from the manure. A total of 1307 site-year-treatment cases met our criteria, which were used for the subsequent analyses. The database included rainfed and irrigated crops (43 and 57% of total cases, respectively).

Since GNC and grain yield were reported across studies either at oven-dry or standard moisture content, all grain yield and GNC data were standardized to 15.5% moisture content for analysis. Reported oven-dry moisture content was assumed to be zero. GNC was measured using combustion and near infrared (NIR) in 70% and 30% of total observations, respectively. Although we did not have side-by-side data to rigorously compare GNC measured with different methods (NIR *versus* combustion), we did not find strong evidence that this would bias the analysis because average GNC (± standard deviation) differed little among experiments using NIR (1.19 ± 0.16%) *versus* combustion (1.13 ± 0.16%) to determine GNC. Additionally, results from the statistical analyses using the database with NIR- *versus* combustion-measured GNC were almost identical; hence, we showed the results using the pooled database (see Section 2.2). For half of the sites, geographic coordinates were available; county or nearby city were reported for the remaining sites. Other variables were available for a reasonable number of experiments (> 40%), including plant density, previous crop, artificial drainage, tillage method, N fertilizer source, N split application (yes/no), and N application timing (spring only or fall and spring). Analytical methods that can handle missing values, such as the regression tree analysis followed in this study, allowed inclusion of the full suite of data (see Section 2.2).

Daily maximum (T_{max}) and minimum (T_{min}) air temperature and precipitation were retrieved for each field from DAYMET (https://daymet.ornl.gov/) while incident solar radiation was retrieved from the Prediction of Worldwide Energy Resources (NASA POWER, https://power.larc.nasa.gov/) based on the coordinates or approximate site

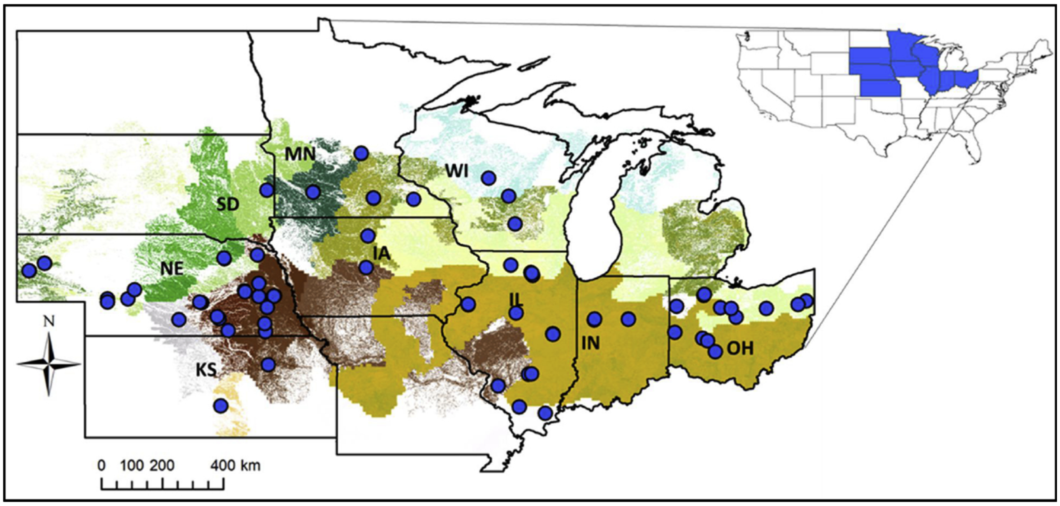

Fig. 1. Map of US North Central region showing the sites of the experiments used in the analysis (circles). Each color represents a climate-soil combination (Technology Extrapolation Domain [TED], Rattalino Edreira et al., 2018). Experiments were located in TEDs that account for 58% of total US maize harvested area. Acronyms are: Illinois (IL), Indiana (IN), Iowa (IA), Kansas (KS), Minnesota (MN), Nebraska (NE), Ohio (OH), South Dakota (SD), and Wisconsin (WI).

reported for each experiment. Both DAYMET and NASA POWER provide gridded weather data (resolution: 1 km^2 and 12,000 km^2, respectively). The DAYMET weather has shown good agreement with measured data for average temperature and total precipitation when summed over several months or an entire growing season (Mourtzinis et al., 2017), while NASA POWER incident solar radiation has shown strong agreement with measured records in agricultural areas with flat terrain, as in the US North Central region (van Wart et al., 2013). Informed by physiological principles (Cantarero et al., 1999; Cicchino et al., 2010; Lobell et al., 2013), key weather variables influencing crop growth and grain yield were investigated in relation to their influence on GNC. For July, which roughly coincides with silking, and for August, corresponding to grain filling in the target region, we calculated mean air temperature (T_{mean}), number of days with $T_{max} \geq 32$ °C, number of days with $T_{min} \geq 22$ °C, mean incident solar radiation, and total water balance, calculated as the difference between total precipitation and reference grass-based evapotranspiration (ETo; Allen et al., 1998). Thresholds of 22 °C (T_{min}) and 32 °C (T_{max}) were chosen for stressful high air temperatures for maize (Herrero and Johnson, 1980; Prasad et al., 2006a; Cicchino et al., 2010; Lobell et al., 2013). Unfortunately, dates of silking and physiological maturity were not recorded in most experiments; hence it was not possible to derive means of weather variables for specific crop phases rather than on a calendar basis. For irrigated crops, water balance was assumed to be zero as irrigation ensures adequate water supply during the entire crop season. Because coordinates were not available for *ca.* half of the experiments, and given the large spatial variability in soil properties, we did not attempt to retrieve site-specific soil parameters.

Experiments were assigned to technology extrapolation domains (TEDs; Rattalino Edreira et al., 2018). Each TED corresponds to a climate-soil domain, within which crop growth and nutrient cycling are expected to be similar. In those cases in which field coordinates were not available, experiments were assigned to the prevalent TED in the area around/within the near town/county where the study was conducted. Experiments used for the analysis were located within TEDs that account for 58% of the total US maize harvested area (Fig. 1). Because of data imbalance among states, with higher number of experiments in NE and MN, the regression tree was repeated 20 times using resampling of 50 observations in these two states, to obtain a balanced experimental design. The test indicated that using either a balanced *versus* unbalanced number of observations or different subsets of randomly selected fields had little impact on the results. Hence, in the present study, we reported only the results derived from the regression trees using the entire database.

2.2. Data analysis

Nitrogen removed with harvested grain was estimated based on reported grain yield and GNC. To evaluate sensitivity of grain-N removal to variation in GNC at a given yield level, we plotted grain-N removal *versus* grain yield and fitted boundary functions using quantile regression for the 5th and 95th percentiles (Koenker and Basset, 1978) *via* the "quantreg" package (Koenker, 2017) in R (Fig. 2). Additionally, analysis of variance (ANOVA) was performed to determine the percentage of total variance in grain-N removal explained by grain yield and GNC.

Whereas GNC varies with hybrid (Genter et al., 1956; Boone et al., 1984; Uribelarrea et al., 2004), the large number of hybrids available in the market and their fast turn over precluded adding hybrid as an explanatory factor for prediction purposes. Here we used ANOVA to

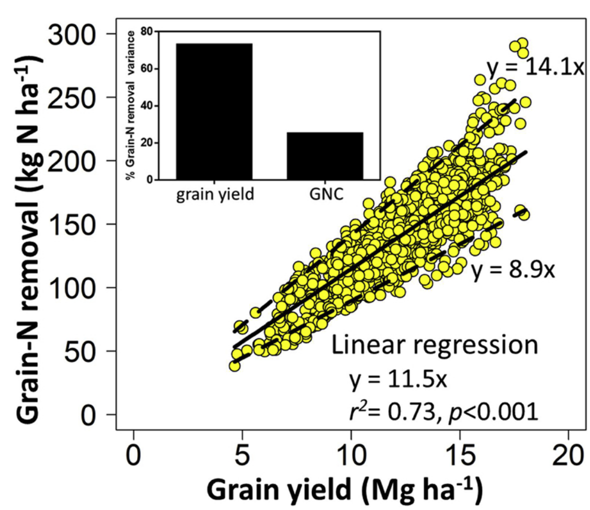

Fig. 2. Relationship between grain-nitrogen (N) removal and grain yield based on data collected from field-grown maize across the US North Central region. Slopes of the linear regression (solid line) and boundary functions fitted for the 5th and 95th percentiles are shown (dashed lines). Fitted regressions were forced through the origin. Grain yields were reported at 15.5% moisture content. Inset shows proportion of grain-N removal variation explained by grain yield and grain N concentration (GNC).

discern the degree to which hybrid explains variation in GNC relative to other factors, using a subset of data that contained 12 hybrids grown consistently across 3 sites and 2 years in IL. All hybrids were grown under rainfed conditions, with N fertilizer rate of 252 kg N ha^{-1}, and plant density of 7.9 plant m^{-2}. Relative maturity ranged between 109–114 d among hybrids and GNC was measured using near infrared. Likewise, previous studies have attributed differences in GNC to a 'dilution effect', suggesting a trade-off between GNC and grain yield (*e.g.*, Gupta et al., 1975; Dudley et al., 1977; Boone et al., 1984; Simmonds, 1995). To assess the degree to which GNC could be explained by grain yield, linear regression models between GNC and grain yield were fitted separately for the entire database, each study, and each study-site-year.

Regression tree analysis was used to quantify the influence of weather and management variables on GNC using the "rpart" package in R (Hothorn et al., 2006). Regression tree analysis is a non-parametric method which recursively partitions the data into successively smaller groups with binary splits based on a single continuous predictor variable (Breiman et al., 1984; Verbyla, 1987; Clark and Pregibon, 1992; Prasad et al., 2006b). Regression tree analysis produces a tree-diagram output, with branches determined by splitting rules and a series of terminal nodes that contain the mean response (*i.e.*, GNC) and the number of observations that fall within each terminal node. The procedure initially grew maximal trees and then used a cross-validation technique (*i.e.*, maxdepth) to prune the over-fitted tree to an optimal size (Therneau and Atkinson, 1997). A "caret" package in R was used to split the dataset into training (80%) and testing (20%) datasets. The training dataset was used to run the regression tree analysis, while the testing dataset was utilized to estimate the mean square error (MSE) between observed and predicted GNC (Supplementary Table S1). The regression tree analysis handled missing values in the explanatory factors (*na.rpart* function), excluding cases only if the response variable (*i.e.*, GNC) or all explanatory factors were missing. When missing values were encountered in considering a split, they were ignored and predictions are calculated from the non-missing values of that factor (Venables and Ripley, 2002). For the regression tree analysis, we excluded some variables due to high collinearity. For example, high correlation (Pearson $r = 0.87$, $P < 0.001$) was found between number of days in July with $T_{max} \geq 32$ °C and July T_{mean}, so we only included the latter variable (Supplementary Table S2). Likewise, incident solar radiation in July was correlated with water balance (Pearson $r = 0.31$; $P < 0.001$) and T_{mean} (Pearson $r = 0.44$, $P < 0.001$). Additionally, source of N applied was highly associated with geographical site (ammonium nitrate was only used in MN, while urea and urea ammonium nitrate were the dominant sources in other experiments); hence, we did not include it in the analysis. Initially, previous crop (*i.e.*, maize and soybean) was included as an explanatory factor and showed to influence GNC. However, in the regression tree, previous crop only differentiated between maize or soybean *versus* no previous crop reported, hence, it was excluded as an explanatory factor. After accounting for these issues, 10 variables remained as potential explanatory factors for variation in GNC (Tables 1 and 2). This same set of explanatory factors was used to generate a regression tree for grain yield to help differentiate drivers for GNC *versus* grain yield variation (Supplementary Fig. S1).

Table 2
Summary statistics for categorical factors used in the analysis.

Categorical variables	% observations
N application time (*n* = *589*)	
spring only	89
fall and spring	11
N split application (*n* = *1307*)	
yes	25
no	75
Tile drainage (*n* = *1030*)	
yes	30
no	70
Tillage method (*n* = *715*)	
conventional[a]	81
no-till	19

[a] Conventional tillage includes chisel plow, disk, field cultivator, strip till, and vertical till.

Relationships between GNC and weather and agronomic factors that were identified as the most important at explaining GNC variation by the regression tree were further explored using linear regression. These factors included July T_{mean} and N fertilizer rate. Mean GNC and standard error were calculated for different intervals of July T_{mean} and N fertilizer rate. Duncan's multiple range test was used to determine significant differences ($\alpha = 0.05$) between means.

We compared the grain-N removal prediction ability of the regression tree GNC estimates with a fixed 1.2% GNC value (as recommended by IPNI in absence of measured GNC) at two spatial levels: field and climate-soil domain (*i.e.*, TED). Agreement between observed and predicted grain-N removal was evaluated using the root mean square error (RMSE) and absolute mean error (ME). Regression analysis was used to explore biases in the relationship between predicted and observed grain-N removal. Frequency distributions were used to estimate the percentage of fields with differences in observed *versus* predicted grain-N removal $\geq |20|$ kg N ha^{-1}. At the TED scale, grain-N removal was estimated by averaging the values across all fields located within the same TED (Fig. 1). The objective of this evaluation was two-fold: (i) to discern any advantage of estimating GNC using a predictive model instead of using a fixed GNC value and (ii) to analyze the uncertainty in predicted grain-N removal at field as opposed to regional level.

3. Results

3.1. Variation in grain nitrogen concentration

The database included variation in GNC, weather, and management practices that is typical of conditions across producer fields in the US North Central region (Tables 1 and 2). The GNC ranged from 0.76 to

Table 1
Summary statistics for maize grain nitrogen concentration and continuous variables (N fertilizer rate, weather variables, and plant density) collected from maize experiments across the US North Central region. The 25th (P25) and 75th (P75) percentiles of the distributions are also shown.

Variables	*n*	Minimum	P25	Median	Mean	P75	Maximum
Grain N concentration (g kg^{-1})	1307	0.76	1.03	1.14	1.15	1.26	1.66
Continuous variables							
N fertilizer rate (kg N ha^{-1})	1307	45	134	196	186	224	381
Total water balance (mm)	1300						
July		−314	−227	−125	−120	0	0
August		−325	−220	−122	−115	0	25
Mean air temperature (ºC)	1300						
July		18.9	21.9	23.8	23.6	25.7	27.9
August		17.8	20.9	21.6	22.1	23.7	28.2
Plant density at harvest (m^{-2})	1096	4.4	7.4	8.0	7.9	8.2	11.9

1.66%, averaging 1.15% across all observations. Average GNC derived here was slightly, though statistically significant (*t*-test; $P < 0.001$), lower than the 1.2% reference reported by IPNI. On average, grain-N removal increased at a rate of 11.5 kg N per Mg of grain yield (Fig. 2), although there was substantial variation in grain-N removal at a given grain yield level due to variation in GNC. Slopes of the quantile regression in Fig. 2 indicate that GNC can vary from 0.89% to 1.41% for a given grain yield. Hence, using the recent (2013–2017) US average grain yield of 10.6 Mg ha^{-1} (https://www.nass.usda.gov), grain-N removal can vary from 94 to 150 kg N ha^{-1}, corresponding to a difference of 56 kg N ha^{-1} in the associated N balance. On the other hand, the proportion of variation in grain-N removal explained by grain yield was *ca.* three times larger than the variance accounted for by GNC (73 *versus* 25%) (Fig. 2, inset).

3.2. Environment versus hybrid influence on grain nitrogen concentration

At issue is the degree to which GNC is influenced by hybrid. An ANOVA, using a subset with a uniform set of hybrids grown across multiple site-years in IL, showed that hybrid influenced GNC more than it affected grain yield (% of sum of squares [%SS] = 32 *versus* 6%, respectively). The portion of variation explained by year, site, and their interaction (*i.e.*, environmental effects) on GNC was higher, but of same order of magnitude, compared with the variation explained by hybrid alone (%SS = 49 *versus* 32). Site effect on GNC was 4-fold larger than year effect, which may reflect the importance of site-specific average weather and/or soil properties (Table 3).

3.3. Relationship between grain yield and grain nitrogen concentration

If variation in GNC is associated with a 'N dilution' effect, one would *a priori* expect a strong negative relationship between GNC and grain yield. In constrast with this expectation, we found a statisitically significant, though weak, positive relationship between GNC and grain yield when the entire dataset was used ($p < 0.001$; $r^2 = 0.02$) (Fig. 3a). The linear regression analysis using the entire database may have been biased by differences in the environmental and/or management background across site-years. To account for this potential confounding effect, we fitted separate regressions to the data compiled from each study (Fig. 3b) and from each study-site-year (Fig. 3c), which indicated that there was a statistically significant negative relationship ($p < 0.001$) in only 11 and 3% of the cases, respectively. We concluded that, for our dataset, observed variation in GNC cannot be attributed to 'N dilution' effect due to yield. Hence, our subsequent analysis did not consider grain yield as an explanatory factor for variation in GNC.

Table 3
Analysis of variance (ANOVA) for the effects of year, site, hybrid, and their interactions on maize grain nitrogen concentration (GNC) and grain yield, in a factorial combination of 6 site-years by 12 commercial hybrids.

Variables	d.f.	F-value[a]		% SS (%)[b]	
		GNC (%)	Grain yield	GNC (%)	Grain yield
Year (Y)	1	40***	88***	7	23
Site (S)	2	95***	64***	32	34
Hybrid (H)	11	17***	2*	32	6
Y x S	2	30***	33***	10	17
Y x H	11	3***	1	5	4
S x H	22	2**	1	8	8
Y x S x H	22	2	1	6	8

[a] *F*-test significant at
* $P < 0.05$.
** $P < 0.01$.
*** $P < 0.001$.
[b] Proportion (in %) of total sum of squares (SS) excluding the error.

3.4. Environmental factors influencing variation in grain nitrogen concentration

The regression tree explained 35% of variation in maize GNC using five variables, including July and August T_{mean}, July and August total water balance, and N fertilizer rate (Fig. 4). July T_{mean} was the most important variable associated with GNC, with crops exposed under warm conditions during July ($T_{mean} \geq 22.5$ °C) exhibiting higher GNC in relation with their counterparts with lower T_{mean} (1.17 *versus* 1.09%). The influence of high air temperature during July on GNC was amplified in fields that were also exposed to unfavorable water balance (*i.e.,* water shortage) and high air temperature in August. In contrast, N fertilizer rate was the most important factor influencing GNC in fields exposed to lower July T_{mean} (< 22.5 °C). In these fields, highest GNC was observed with large N fertilizer input and unfavorable water balance, while fields with lowest GNC were associated with small N fertilizer inputs (Fig. 4). Fields with lowest GNC corresponded to those exposed to the same conditions as fields with highest GNC, but with lower T_{mean} during August (< 21.6 °C). Finally, the explanatory power of the regression tree for GNC was about one-half of that for grain yield ($R^2 = 0.35$ *versus* 0.65; Fig. 4, Supplementary Fig. S1) and different in relation to the driving variables.

We further investigated the relationships between GNC and two variables identified in the regression tree: July T_{mean} and N fertilizer rate (Fig. 5). GNC increased with increasing July T_{mean} and N fertilizer rates (Fig. 5a, b). Across the entire range of N fertilizer rates, GNC was higher in warmer environments; however, this difference was larger for small and moderate N fertilizer rates (Fig. 5c). At high N rates (300–400 kg N ha^{-1}), there was no significant difference in GNC between fields exposed to high *versus* low July T_{mean}.

3.5. Comparison of grain-N removal with fixed and modelled GNC

We evaluated two methods (regression tree's estimates *versus* fixed 1.2% GNC value) on their performance to reproduce the observed grain-N removal (Fig. 6). Predicted grain-N removal based on reported grain yield and GNC estimated from the regression tree had a slightly better fit to observed values compared with the approach based on a fixed value, with RMSE representing 12% *versus* 15% of the mean observed grain-N removal, respectively (Fig. 6a, b). Consistent with this finding, the percentage of site-years with large differences ($\geq |20|$ kg N ha^{-1}) between predicted and observed grain-N removal was smaller using regression tree *versus* fixed GNC values (25 *versus* 36% of total fields) (Fig. 6a, b, insets). However, both approaches underestimated grain-N removal in the upper range of observed values (> 200 kg N ha^{-1}), which was consistent with the statistically significant quadratic term revealed by our regression analysis ($P < 0.001$). Agreement between predicted and observed values at the TED level was improved compared to agreement of field-level data (average RMSE% = 9 *versus* 13%), with very little difference in accuracy between estimates based on the fixed GNC *versus* regression-tree (RMSE%: 10 *versus* 9% of observed mean) (Fig. 6c, d).

4. Discussion

The influence of environmental and management factors on maize GNC were assessed using data collected from multiple sites and years across the US North Central region to include field experiments that are representative of dominant management practices in producer fields. Average maize GNC calculated for the entire database was 1.15%, which was slightly lower than the commonly used GNC of 1.2%, and corresponds with a continuing decline in GNC over time (Welch, 1971; Boone et al., 1984; Duvick and Cassman, 1999; Ciampitti and Vyn, 2012). Overall, the regression tree explained 35% of variation in GNC across the US North Central region, with air temperature and water balance during July and August and N fertilizer rate identified as the

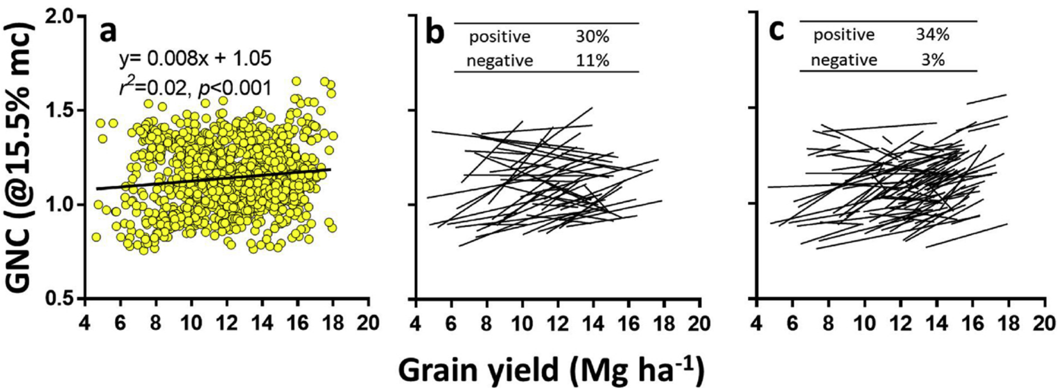

Fig. 3. Relationships between grain nitrogen concentration (GNC) and grain yield for the entire dataset (a), each study (b), and each study-site-year (c). Data points were removed and only the fitted linear regressions are shown in (b) and (c) and percentage of cases with statistically significant positive and negative relationships are shown ($p < 0.001$).

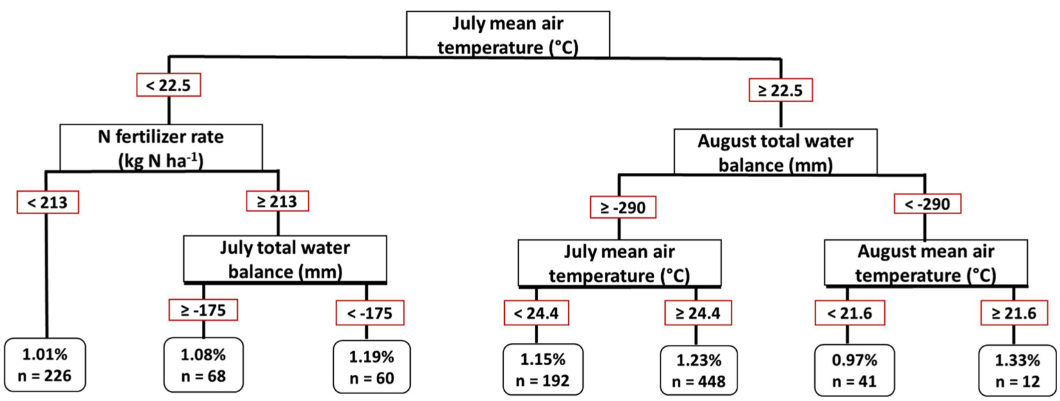

Fig. 4. Regression tree model showing sources of variation in grain nitrogen concentration (GNC) due to weather and management factors (overall $R^2 = 0.35$; MSE = 0.02%). Boxes are splitting nodes (SN), with bottom boxes representing terminal nodes (TN). Values within each TN indicate average GNC at a 15.5% moisture content basis and the number of observations (n) in each terminal node.

most important factors explaining variation in GNC. We recognize that part of the unexplained variation could be attributed to hybrid, which could account for *ca.* one third of GNC variation as indicated by our analysis using a subset of site-years where the same set of hybrids were grown. Nonetheless, accounting for hybrid effect for predictive purpose is very difficult given the large number of hybrids available in the market and their rapid turnover. Soil factors may also account for part of the unaccounted variation in GNC. Our ANOVA indicated a much larger influence of site rather than year on GNC, which could reflect differences in soil properties, although it is difficult to separate this effect from weather variation across sites. This finding highlights the importance of collecting *in situ* key soil and topography data (*e.g.*, available-water holding capacity, soil texture, landscape position, *etc.*) or, at least, reporting of exact experiment coordinates so that these attributes can be retrieved from existing databases such as SSURGO (www.websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx).

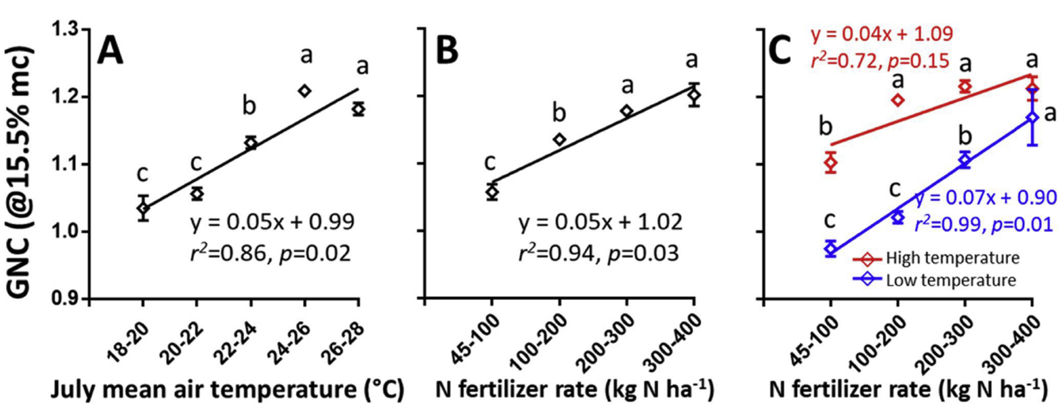

Fig. 5. Relationships between average grain nitrogen concentration (GNC, 15.5% moisture content basis) and July mean air temperature (A) and N fertilizer rate (B). Relationship between GNC and N fertilizer rate, for fields with contrasting July mean air temperature (greater or lower than 22.5 °C based on Fig. 4), is shown in (C). Fitted linear regressions and their parameters are shown. Each data point represents average GNC for fields that fall within each July mean air temperature and/or N fertilizer rate interval. Vertical bars indicate the standard error of the mean. Different letters indicate statistically significant differences (Duncan's test; alpha = 0.05).

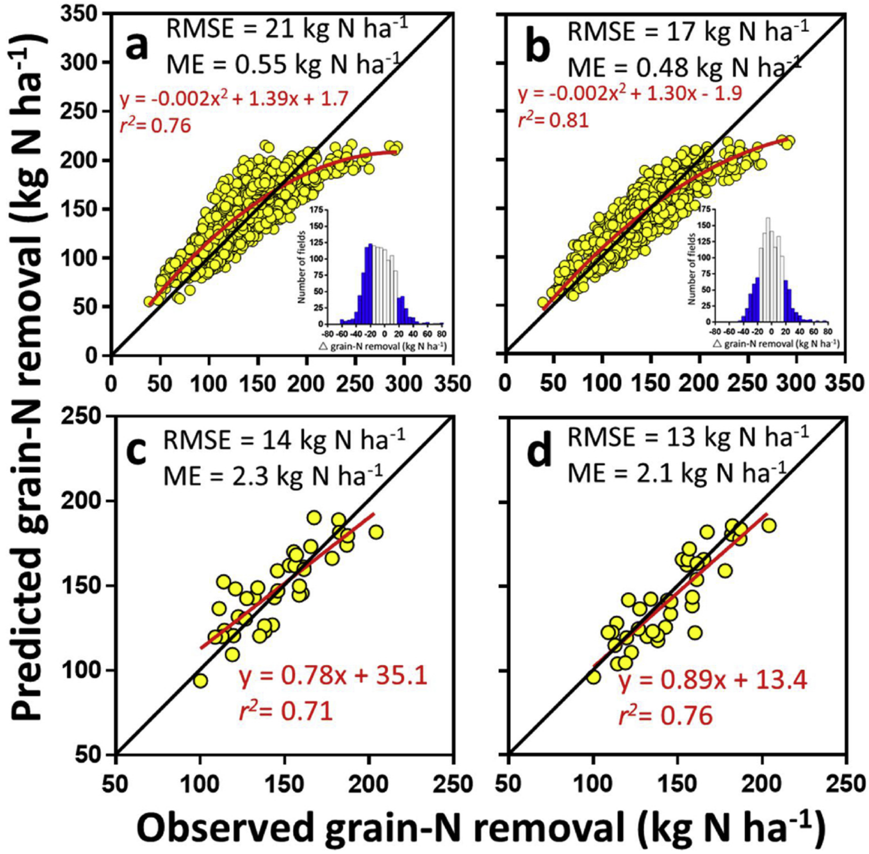

Fig. 6. Predicted *versus* observed grain-N removal in maize for each site-year-treatment case (a, b) and for climate-soil domains (c, d). Predicted grain-N removal was calculated based on a fixed (1.2%) grain nitrogen concentration (a, c) or based on concentration estimated from the regression tree model (b, d). Root mean square error (RMSE) and mean error (ME) are indicated and y = x (black) and quadratic or linear regression (red) lines are shown. Insets show frequency distributions for the difference between observed and predicted grain-N removal; fields with differences ≥ |20| kg N ha^{-1} are shown in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Unfortunately, soil parameters and/or field coordinates were not collected and/or missing for most of the observations in our database, so we could not include these factors in our evaluation.

Results from the regression tree analysis are consistent with current understanding of factors influencing GNC. In general, stressful weather conditions during July and August, such as high air temperature and unfavorable water balance (*i.e.,* water shortage), and high N fertilizer rates led towards high GNC, which is consistent with previous studies (Genter et al., 1956; Mayer et al., 2016). High temperature and unfavorable water balance during the kernel setting phase reduces kernel number (Hall et al., 1981; Otegui et al., 1995; Rattalino Edreira et al., 2011). Our study also suggested that unfavorable (favorable) weather conditions during August seem to amplify (ameliorate) the effect of stressful conditions during July. In relation to N supply, our analysis revealed an interactive effect of air temperature and N fertilizer rate on GNC, with largest differences in GNC between fields exposed to contrasting temperature in low N fertilizer rate conditions, which are consistent with published results for wheat (Altenbach et al., 2003).

Previous studies have reported that maize GNC tends to increase with decreasing grain yield as a result of 'N dilution' effect (Zuber et al., 1954; Simmonds, 1995; Uribelarrea et al., 2004). However, in the current study, GNC and yield were related weakly and inconsistently. Further, the fitted regression tree for grain yield was substantially different from the one for GNC (Fig. 4; Supplementary Fig. S1). A possible explanation for the discrepancy between our study and previous reports is that our database did not include extreme conditions such as severe drought, N omission plots or very high or low plant densities as in previous studies (Zuber et al., 1954; Lang et al., 1956) because these conditions are not common in producer fields. Instead, our objective was to understand GNC variation within the range of environment and management practices typically found in producer fields. Another explanation is that most studies used for our analysis included treatments with varying N fertilizer amounts which caused, in most cases, a simultaneous increase in grain yield and GNC with increasing N fertilizer input. In contrast, previous studies reporting a trade-off between GNC and grain yields for maize were based on experiments in which yield differences were a consequence of using different hybrids and/or plant densities across treatments, without changing N fertilizer amounts (*e.g.*, Gupta et al., 1975; Dudley et al., 1977; Boone et al., 1984; Simmonds, 1995). In other words, the trade-off between grain yield and GNC is not apparent when variation in yield is due to differences in N fertilizer input. In agreement with this hypothesis, a number of studies (Zuber et al., 1954; Chen and Vyn, 2017; DeBruin et al., 2017) reported decreasing GNC with increasing yield due to improved hybrids and/or higher plant density, but the same authors reported that *both* GNC and grain yield increased with increasing N fertilizer rate.

The predictive model developed for estimating GNC is more accurate, relative to the approach using a fixed GNC value, at estimating grain-N removal and N balance for individual site-years. Hence, the predictive model can help obtain more accurate estimates of grain-N removal and N balance in producer maize fields, in absence of GNC data, although this advantage needs to be weighed against the extra data needed (weather, N fertilizer) to use the model. The predictive model underestimated grain-N removal in the upper range of observed values (> 200 kg N ha^{-1}). An implication of this finding is that grain-N removal may be underestimated in high-yield environments that favor large N uptake. Indeed, 96% of the observations with grain-N removal > 200 kg N ha^{-1} corresponded to irrigated maize in NE—a production environment where producers routinely attain yields that correspond to 80–90% of their yield potential as determined by climate and current genetics (Grassini et al., 2011). Predictions of grain-N removal using both approaches were more accurate at climate-soil domain level compared with estimates for individual site-year cases. This suggests that comparisons for these parameters (*i.e.*, grain-N removal and N balance) among climate-soil domains using aggregated values are more reliable compared with assessments for individual fields. In addition, result from this study indicates that using the fixed GNC value of 1.2% would work reasonably well for estimating grain-N removal at climate-soil domain level. Hence, in absence of measured GNC data, the N balance approach would still provide reasonable estimates of potential N losses for major climate-soil domains where maize is grown in the US North Central region. In contrast, estimates for individual fields will be subjected to greater uncertainty and, ultimately, GNC should be measured for accurate quantification of N balance. New technologies, such as combines equipped with NIR to map protein at the same level of yield maps, may allow direct measurement of N-grain removal at field and intra-field scales in the future (Montes et al., 2006; https://grdc.com.au/resources-and-publications/grdc-update-papers/tab-content/grdc-update-papers/2017/07/on-the-go-protein-sensors-using-real-time-protein-data-for-more-profitable-marketing-aggregations-and-nitrogen-decisions). The methodology described in this paper for understanding sources of variation in GNC estimation could potentially be applied to other regions or crops depending upon availability of data on GNC and ancillary variables.

5. Conclusions

Variation in GNC causes uncertainty in estimates of grain-N removal and N balance. Our results identified N fertilizer rate and air temperature and water balance in July and August as the most important factors explaining variation in GNC. We did not find evidence of a negative correlation between GNC and grain yield. In absence of measured GNC data, the predictive model developed here can help refine estimates of grain-N removal and N balance for specific site-years, although its advantage needs to be balanced out against the extra data requirements. Estimates of N balance seem to be more accurate when aggregated to climate-soil domains compared with individual fields; in that case, using a fixed GNC value to estimate grain-N removal would work reasonably well.

Acknowledgments

We are grateful for the financial support received from the Environmental Defense Fund (EDF), Agriculture and Food Research Initiative (AFRI) of the United States Department of Agriculture, and Nebraska Corn Board (NCB). We also thank the University of Nebraska-Lincoln (UNL) Quantitative Life Sciences Initiative for their assistance with the regression-tree analysis. We also thank Drs Tony Vyn & Juan Pablo Burzaco (Purdue University), and Dr Ignacio Ciampitti (Kansas State University) for providing some data for the analysis and useful comments about the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fcr.2018.10.017.

References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements, FAO Irrigation and Drainage Paper 56. FAO, Rome, Italy.
Altenbach, S.B., DuPont, F.M., Kothari, K.M., Chan, R., Johnson, E.L., Lieu, D., 2003. Temperature, water and fertilizer influence the timing of key events during grain development in a US spring wheat. J. Cereal Sci. 37, 9–20.
Boone, L.V., Vasilas, B.L., Welch, L.F., 1984. The nitrogen content of corn grain as affected by hybrid, population, and location. Commun. Sol. Sci. Plant Anal. 15, 639–650.
Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Wadsworth, Belmont, CA.
Cantarero, M.G., Cirilo, A.G., Andrade, F.H., 1999. Night temperature at silking affects set in maize. Crop Sci. 39, 703–710.
Cassman, K.G., Dobermann, A., Walters, D.T., 2002. Agroecosystems, nitrogen-use efficiency, and nitrogen management. Ambio 31, 132–140.
Chen, K., Vyn, T.J., 2017. Post-silking factor consequences for N efficiency changes over 38 years of commercial maize hybrids. Front. Plant Sci. 8, 1–18.
Ciampitti, I.A., Vyn, T.J., 2012. Physiological perspectives of changes over time in maize yield dependency on nitrogen uptake and associated nitrogen efficiencies: a review. Field Crops Res. 133, 48–67.
Cicchino, M., Rattalino Edreira, J.I., Uribelarrea, M., Otegui, M.E., 2010. Heat stress in field-grown maize: response of physiological determinants of grain yield. Crop Sci. 50, 1438–1448.
Clark, L.A., Pregibon, D., 1992. Tree-based models. In: Chambers, J.M., Hastie, T.J. (Eds.), Statistical Models in S. Wadsworth, Pacific Grove (CA), pp. 377–419.
Correll, R., Butler, J., Spouncer, L., Wrigley, C., 1994. The relationship between grain-protein content of wheat and barley and temperatures during grain filling. Aust. J. Plant Physiol. 21, 869–873.
DeBruin, J.L., Schussler, J.R., Mo, H., Cooper, M., 2017. Grain yield and nitrogen accumulation in maize hybrids released during 1934 to 2013 in the US Midwest. Crop Sci. 57, 1–16.
Dudley, J.W., Lambert, R.J., de la Roche, I.A., 1977. Genetic analysis of crosses among corn strains divergently selected for percent oil and protein. Crop Sci. 17, 111–117.
Duvick, D.N., Cassman, K.G., 1999. Post-green revolution trends in yield potential of temperate maize in the north-central United States. Crop Sci. 39, 1622–1630.
Erisman, J.W., Galloway, J.N., Seitzinger, S., Bleeker, A., Dise, N.B., Petrescu, A.M.R., Leach, A.M., de Vries, W., 2013. Consequences of human modification of the global nitrogen cycle. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368, 1621.
Feil, B., Thiraporn, R., Geisler, G., Stamp, P., 1990. Genotype variation in grain nutrient concentration in tropical maize grown during a rainy and a dry season. Agronomie 10, 717–725.
Genter, C.F., Eheart, J.F., Linkous, W.N., 1956. Effect of location, hybrid, fertilizer, and rate of planting on the oil and protein content of corn grain. Agron. J. 48, 63–67.
Grassini, P., Thornburn, J., Burr, C., Cassman, K.G., 2011. High-yield irrigated maize in the Western U.S. Corn Belt: I. On-farm yield, yield potential, and impact of agronomic practices. Field Crops Res. 120, 144–152.
Grassini, P., Specht, J., Tollenaar, T., Ciampitti, I., Cassman, K.G., 2014. High-yield maize-soybean cropping systems in the U.S. Corn Belt. In: Sadras, V.O., Calderini, D.F. (Eds.), Crop Physiology- Applications for Genetic Improvement and Agronomy, 2nd edition. Elsevier, The Netherlands.
Gupta, D., Kovacs, I., Gaspar, L., 1975. Protein quality traits and their relationships with yield and yield components of opaque-2 and analogous normal hybrids and inbred lines. Theor. Appl. Genet. 45, 341–348.
Hall, A.J., Lemcoff, J.H., Trapani, N., 1981. Water stress before and during flowering in maize and its effects on yield, its components, and their determinants. Maydica 26, 19–38.
Hansen, P.M., Jørgensen, J.R., Thomsen, A., 2002. Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression. J. Agric. Sci. 139, 307–318.
Herrero, M.P., Johnson, R.R., 1980. High temperature stress and pollen viability in maize. Crop Sci. 20, 796–800.
Hothorn, T., Hornik, K., van de Wiel, M.A., Zeileis, A., 2006. A lego system for conditional inference. Am. Stat. 60, 257–263.
Jones, C.A., Kiniry, J.R., 1986. CERES-Maize: A Simulation Model of Maize Growth and Development. U.S.A., Texas A&M University Press, College Station, TX, pp. 194.
Koenker, R., Basset, G., 1978. Regression quantiles. Econometrica 46, 33–50.
Koenker, R., 2017. Quantreg: quantile regression. R package version 5.36. 2017. http://CRAN.R-project.org/package=quantreg.
Lang, A.L., Pendleton, J.W., Dungan, G.H., 1956. Influence of population and nitrogen levels on yield and protein and oil contents of nine hybrids. Agron. J. 48, 284–289.
Liang, B.C., MacKenzie, A.F., Zhang, T.Q., 1996. Grain yields and grain nitrogen concentration of corn as influenced by fertilizer nitrogen rate. J. Agron. Crop Sci. 177, 217–223.
Liu, H.L., Yang, J.Y., Drury, C.F., Reynolds, W.D., Tan, C.S., Bai, Y.L., He, P., Jin, J., Hoogenboom, G., 2010. Using the DSSAT-CERES-Maize model to simulate crop yield and nitrogen cycling in fields under long-term continuous maize production. Nutr. Cycl. Agroecosyst. 89, 313–328.
Lobell, D.B., Hammer, G.L., McLean, G., Messina, C., Roberts, M.J., Schlenker, W., 2013. The critical role of extreme heat for maize production in the United States. Nat. Clim. Change 3, 497–501.
Mayer, L.I., Rossini, M.A., Maddonni, G.A., 2016. Heat stress during grain filling modifies kernel protein composition in field-grown maize. Crop Sci. 56, 1890–1903.
McLellan, E.L., Cassman, K.G., Eagle, A.J., Woodbury, P.B., Sela, S., Tonitto, C., Marjerison, R.D., van Es, H.S., 2018. The nitrogen balancing act: tracking the environmental performance of food production. BioSci 68, 194–203.
Montes, J., Utz, H., Schipprack, W., Kusterer, B., Muminovic, J., Paul, C., Melchinger, A., 2006. Near-Infrared spectroscopy on combine harvesters to measure maize grain dry matter content and quality parameters. Plant Breed. 125, 591–595.
Mourtzinis, S., Rattalino Edreira, J.I., Conley, S.P., Grassini, P., 2017. From grid to field: assessing quality of gridded weather data for agricultural applications. Eur. J. Agron. 82, 163–172.
Otegui, M.E., Andrade, F.H., Suero, E.E., 1995. Growth, water use, and kernel abortion of maize subjected to drought at silking. Field Crops Res. 40, 87–94.
Prasad, P.V.V., Boote, K.J., Allen Jr, L.H., 2006a. Adverse high temperature effects on pollen viability, seed-set, seed yield and harvest index of grain-sorghum [Sorghum bicolor (L.) Moench] are more severe at elevated carbon dioxide due to high tissue temperature. Agric. For. Meteorol. 139, 237–251.
Prasad, A.M., Iverson, L.R., Liaw, A., 2006b. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199.
Rattalino Edreira, J.I., Budakli Carpici, E., Sammarro, D., Otegui, M.E., 2011. Heat stress effects around flowering on kernel set of temperate and tropical maize hybrids. Field Crops Res. 123, 62–73.
Rattalino Edreira, J.I., Cassman, K.G., Hochman, Z., van Ittersum, M.K., van Bussel, L.,

Claessens, L., Grassini, P., 2018. Beyond the plot: technology extrapolation domains for scaling out agronomic science. Environ. Res. Lett. 13, 054027.
Smith, G.P., Gooding, M.J., 1999. Models of wheat grain quality considering climate, cultivar and nitrogen effects. Agric. For. Meteorol. 94, 159–170.
Simmonds, N.W., 1995. The relation between yield and protein in cereal grain. J. Sci. Food Agric. 67, 309–315.
Therneau, T.M., Atkinson, E.J., 1997. An Introduction to Recursive Partitioning Using the RPART Routine. Technical Report 61, Mayo Clinic, Section of Statistics.
Uribelarrea, M., Below, F.E., Moose, S.P., 2004. Grain composition and productivity of maize hybrids derived from the Illinois protein strains in response to variable nitrogen supply. Crop Sci. 44, 1593–1600.
Van Wart, J., Grassini, P., Cassman, K.G., 2013. Impact of derived global weather data on simulated crop yields. Glob. Change Biol. Bioenergy 19, 3822–3834.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics With S, fourth edition. Springer-Verlag, New York.
Verbyla, D.L., 1987. Classification trees: a new discrimination tool. Can. J. Res. 17 (1150), 1152.
Viets, F.G.Jr., Domingo, C.E., 1948. Yields and nitrogen content of corn hybrids as affected by nitrogen supply. Soil Soc. Am. Proc. 73, 303–306.
Welch, L.F., 1971. More nutrients are added to soil than are hauled away in crops. Illinois Res. 14 (1), 3–4 No.
Yakoub, A., Lloveras, J., Biau, A., Lindquist, J.L., Lizaso, J.I., 2017. Testing and improving the maize models in DSSAT: development, growth, yield, and N uptake. Field Crops Res. 2012, 95–106.
Zhang, X., Davidson, E.A., Mauzerall, D.L., Searchinger, T.D., Dumas, P., Shen, Y., 2015. Managing nitrogen for sustainable development. Nature 528, 51–59.
Zhao, C.J., Liu, L.Y., Wang, J.H., Huangm, W.J., Song, X.Y., Li, C.J., 2005. Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress. Int. J. Appl. Earth Obs. Geoinf. 7, 1–9.
Zuber, M.S., Smith, G.E., Gehrke, C.W., 1954. Crude protein of corn grain and stover as influenced by different hybrids, plant population and N levels. Agron. J. 46, 257–261.