Field Crops Research 284 (2022) 108578

Contents lists available at ScienceDirect

Field Crops Research

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

Estimating maize harvest index and nitrogen concentrations in grain and residue using globally available data

Cameron I. Ludemann[a,*], Renske Hijbeek[a], Marloes P. van Loon[a], T. Scott Murrell[b], Achim Dobermann[c], Martin K. van Ittersum[a]

[a] Wageningen University & Research, Plant Production Systems, Wageningen, The Netherlands
[b] African Plant Nutrition Institute and Mohammed VI Polytechnic University, Ben Guerir, Morocco
[c] International Fertilizer Association, Paris, France

ARTICLE INFO

Keywords:
Nutrient
Budget
Nutrient removal
Nutrient use efficiency
Maize
Harvest index
Random forest
Mixed-effect model

ABSTRACT

Reliable estimates of crop nitrogen (N) uptake and offtake are critical in estimating N balances, N use efficiencies and potential losses to the environment. Calculation of crop N uptake and offtake requires estimates of yield of crop product (e.g. grain or beans) and crop residues (e.g. straw or stover) and the N concentration of both components. Yields of crop products are often reasonably well known, but those of crop residues are not. While the harvest index (HI) can be used to interpolate the quantity of crop residue from available data on crop product yields, harvest indices are known to vary across locations, as do N concentrations of residues and crop products. The increasing availability of crop data and advanced statistical and machine learning methods present us with an opportunity to move towards more locally relevant estimates of crop harvest index and N concentrations using more readily available data. The aim of this study was to investigate whether improved estimates of maize crop HI and N concentrations of crop products and crop residues can be based on crop data available at the global scale, such as crop yield, fertilizer application rates and estimates of yield potential. Experiments from 1487 different locations conducted across 31 countries were used to test various prediction models. Predictions from mixed-effects models and random forest machine learning models provided reasonable levels of prediction accuracy (R^2 of between 0.33 and 0.68), with the random forest method having greater accuracy. Although the mixed-effects prediction models had lower prediction accuracy than random forest, they did provide better interpretability. Selection of which method to use will depend on the objective of the user. Here, the random forest and mixed-effects methods were applied to N in maize, but could equally be applied to other crops and other nutrients, if data becomes available. This will enable obtaining more locally relevant estimates of crop nutrient offtake to improve estimates of nutrient balances and nutrient use efficiency at national, regional or global levels, as part of strategies towards more sustainable nutrient management.

1. Introduction

Reliable estimates of nutrient uptake and nutrient offtake (removal) are critical for estimating nutrient balances and nutrient use efficiencies at scales that may range from an individual field or farm to whole countries, regions and the world. Particularly for nitrogen (N), such nutrient budgeting approaches are constrained by numerous uncertainties about the underlying data (Zhang et al., 2021a). Considering that crop nutrient removal is the major nutrient output component in such input-output budgeting approaches, estimating it more precisely is at the core of improving the monitoring of N surpluses, N use efficiency (NUE) or other nutrient indicators, with a wide range of potential applications. For example internal nutrient use efficiency, calculated from nutrient uptake and crop yield, is an important factor in algorithms that recommend nutrient applications rates to farmers (Witt et al., 1999).

Nitrogen offtake is an important component of the N balance, as it provides an indication of how efficiently fertilizer and other N inputs are utilized by the crop (Cassman et al., 2002). Numerous global estimates

Abbreviations: AGY, above ground yield; CPN, crop product nitrogen concentration; CPY, crop product yield; CRN, crop residue nitrogen concentration; CRY, crop residue yield; HI, harvest index; N, nitrogen; P, phosphorus; K, potassium.

* Corresponding author.
E-mail address: cameron.ludemann@wur.nl (C.I. Ludemann).

https://doi.org/10.1016/j.fcr.2022.108578
Received 20 February 2022; Received in revised form 10 May 2022; Accepted 13 May 2022
Available online 30 May 2022

Table 1
Relevant variables in the maize dataset for this study.

Abbreviation	Description	Units
AGY	Above ground (biomass) yield	Mg DM ha^{-1}[a]
CPY	Crop product yield	Mg DM ha^{-1}
CRY	Crop residue yield	Mg DM ha^{-1}
HI	Harvest index (CPY/AGY)	Proportion
FN	Fertilizer nitrogen (N) application rate	kg N per m^2
FP	Fertilizer phosphorus (P) application rate	kg elemental P per m^2
FK	Fertilizer potassium (K) application rate	kg elemental K per m^2
CPN	Crop product N concentration	% (100 × kg N per kg DM^{-1})
CRN	Crop residue N concentration	% (100 × kg N per kg DM^{-1})
Yp	Non-water limited yield potential	Mg DM ha^{-1} (of CPY)
Yw	Water limited yield potential	Mg DM ha^{-1} (of CPY)
Ypot	Yield potential where Yp was used for irrigated sites and Yw was used for non-irrigated (rainfed) sites.	Mg DM ha^{-1} (of CPY)
RY	Relative yield gap closure (calculated specifically for each year) calculated as CPY divided by Ypot.	Proportion

[a] Megagrams (Mg) of dry matter (DM) per hectare (ha).

have been made for N balances in crop production, and these estimates vary widely at global to national scales (Zhang et al., 2021a). Most of these calculations have been made with single 'average' estimates of crop nutrient concentration applied across the world (Conant et al., 2013; Lassaletta et al., 2014; Zhang et al., 2021a). Crop product yield (CPY) for an area of interest is often reasonably well known from surveys, field measurements or official crop statistics, although yield estimates can vary a lot too, depending on the method chosen (Kosmowski et al., 2021). In contrast, crop residue yield (CRY) remaining in the field is rarely known for larger areas. Harvest index (HI) is the proportion of the above ground biomass that is harvested and marketable (CPY) and is a unitless decimal fraction ($0 < HI < 1$). While the HI can be used to interpolate the total quantity of CRY from available data on CPY, HI is known to vary across locations, as do N concentrations of crop residues (crop residue nitrogen concentration-CRN) and crop products (crop product nitrogen concentration-CPN). Another complication is that amounts of crop residue left in the field may vary widely, depending on harvest technologies, economic use options, and farmer preferences. Some general estimates of crop residue amounts have been made for other reasons, such as use of residues as biofuel (Lal, 2005) or accounting of nitrous oxide emissions from crop residues (www.fao.org/faostat/en/#data/GA/metadata). However, to our knowledge, there is no readily available global data base on crop residues amounts and their use in cropland.

The increasing availability of crop data and advanced statistical and machine learning methods presents us with the opportunity to move toward more locally relevant estimates of crop HI, CPN and CRN, which in turn can improve N budgets and use efficiency estimates at national to global scales. Some important sources of globally available data that may offer potential to improve estimates of crop N offtake include CPY data from the UN-FAO (FAOSTAT, 2020), fertilizer application rate data (Heffer et al., 2017; Ludemann et al., 2022), and crop yield potential data (van Ittersum et al., 2013). We hypothesized that these three variables are all potentially related to HI, CPN and CRN of maize.

In this study, we use maize as a case-study to assess how well HI, CPN and CRN can be predicted using a limited set of globally available variables (i.e. CPY, fertilizer application and potential yields). First, an elaboration is made on some of the key concepts used in this study as well as how the three key explanatory variables are expected to be related to HI, CPN and CRN (as hypotheses). The work in this paper focuses on the above metrics for N in maize. Subsequently, we aim to apply the same methodology to other crops and nutrients, hoping to establish a globally applicable set of nutrient removal prediction algorithms.

2. Concepts and hypotheses

2.1. Concepts

Nutrient offtake is defined as "nutrient removal from the soil system through the harvest of crops" (NAL USDA, 2021), whereas nutrient uptake refers to the total quantity of nutrients accumulated in above-ground biomass at the time of sampling. Note that, primarily for pragmatic reasons, this definition excludes nutrients accumulated in belowground biomass (roots). Nutrient uptake defined in this way is appropriate for annual crops, and sampling typically occurs at physiological maturity. For annual seed crops like cereals, nutrient uptake is determined by summing the nutrient accumulation across two above ground components: 1) the seed and 2) all non-seed organs, referred to commonly as the "residue" or "stover", including stems and leaves. For both components, nutrient accumulation is calculated as the product of

Table 2
Hypothesis 1, 2 and 3 of this study listed with the models used to test these hypotheses.

CPY models	FN models	Ypot models	RY models	CPN models
Hypothesis 1: That variation in harvest index (HI) of maize can be explained using crop product yield (CPY), fertilizer application rates of N (FN), phosphorus (FP), potassium (FK), yield potential (Ypot), and/or relative yield potential (RY)[a].				
1) HI∼ CPY	6) HI∼FN	9) HI∼Ypot	11) HI∼RY	
2) HI∼ CPY^2	7) HI∼FN+FP	10) HI∼Ypot + CPY		
3) HI∼CPY+FN	8) HI∼FN+FP+FK			
4) HI∼CPY+FN+FP				
5) HI∼CPY+FN+FP+FK				
Hypothesis 2: That variation in crop product nitrogen (CPN) concentration of maize can be explained using CPY, FN, FP, FK Ypot and/or RY[a].				
1) CPN∼ CPY	5) CPN∼FN	8) CPN∼Ypot	13) CPN∼RY	
2) CPN∼CPY+FN	6) CPN∼FN+FP	9) CPN∼ Ypot+ CPY	14) CPN∼RY+FN	
3) CPN∼CPY+FN+FP	7) CPN∼FN+FP+FK	10) CPN∼Ypot+FN	15) CPN∼RY+FN+FP	
4) CPN∼CPY+FN+FP+FK		11) CPN∼Ypot+FN+FP	16) CPN∼RY+FN+FP + FK	
		12) CPN∼Ypot+FN+FP+FK		
Hypothesis 3: That variation in crop residue nitrogen (CRN) concentration of maize can be explained using CPY, FN, FP, FK, Ypot, RY and CPN[a].				
1) CRN∼ CPY	4) CRN∼FN	5) CRN∼Ypot	9) CRN∼RY	12) CRN∼CPN
2) CRN∼CPY+FN		6) CRN∼Ypot +CPN	10) CRN∼RY + CPN	13) CRN∼CPN+ CPY
3) CRN∼CPY+FN +FP		7) CRN∼Ypot +FN	11) CRN∼RY +FN	14) CRN ∼CPN + CPY + FN
		8) CRN∼Ypot +FN +FP		15) CRN ∼CPN + CPY + FN +FP
				16) CRN ∼CPN + CPY + FN +FP +FK

[a] Abbreviations for variables are described in Table 1, and all these models included a 'region' random effect in a mixed-effects model. Not all combinations were modelled due to lack of data. Further explanation of how the models were used to test the hypotheses are included in the Material and Methods section.

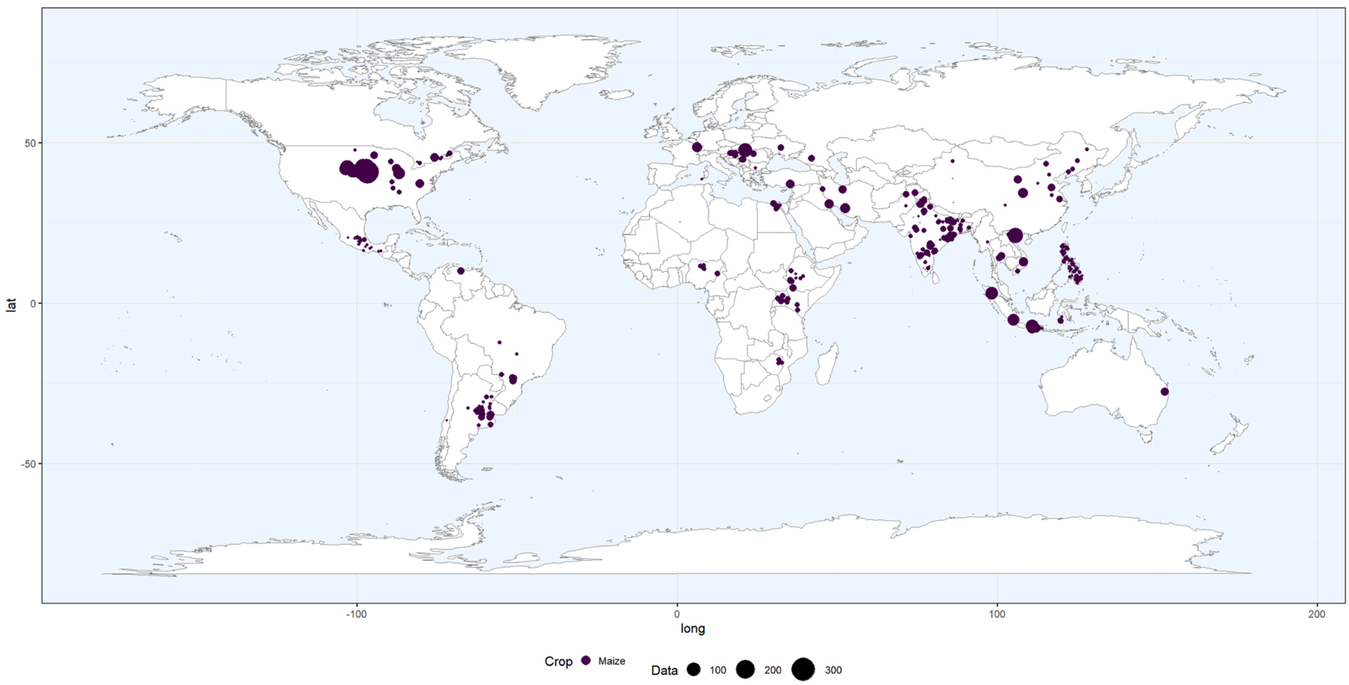

Fig. 1. A map showing the locations of trials from which data were used in the analysis of this study. The size of the symbols relate to the number of unique datum for each location. A unique datum represents a unique treatment-replicate combination.

Table 3
Mean values of important variables in dataset analyzed in this study* .

Region* *	Mean values for each variable* **									Trials per region****
	FN (kg ha^{-1})	FP (kg ha^{-1})	FK (kg ha^{-1})	CPY (Mg DM ha^{-1})	CRY (Mg DM ha^{-1})	Ypot (Mg DM ha^{-1})	HI (-)	CPN (%)	CRN (%)	
Africa	113	29	27	3.71	6.45	7.59	0.38	1.26	0.86	331
East Asia	165	54	77	6.38	6.46	10.63	0.48	1.32	0.65	417
Eastern Europe & Central Asia	30	0	0	6.13	NA	NA	NA	NA	NA	1
Latin America	116	28	38	8.29	8.84	9.54	0.48	1.34	0.45	76
North America	151	19	41	10.5	9.88	11.8	0.5	1.3	0.76	38
Oceania	174	0	0	5.08	8.68	NA	0.36	1.13	0.68	1
South Asia	129	32	47	4.76	7.91	8.3	0.39	1.47	0.75	605
West Asia	156	65	29	8.79	10.5	17.59	0.44	1.7	1.03	7
Western and Central Europe	107	34	81	8.53	NA	6.64	NA	1.19	NA	11
World	136	32	50	7.73	8.57	10.29	0.45	1.32	0.73	1487

*The mean values are representative across all treatments in the data provided and may not represent the mean values for the typical fertilizer application rates in each region. Standard deviations for these parameters are included in Appendix D.

* *Region based on IFA (2021).

* **where: CPY=crop product yield (megagrams dry matter (Mg DM) ha^{-1}), CRY=crop residue yield (Mg DM ha^{-1}), HI=harvest index (product yield as a proportion of above ground biomass), FN=fertilizer nitrogen (N) applied (kg N ha^{-1}), FP=fertilizer phosphorus applied (kg elemental P ha^{-1}), FK=fertilizer potassium applied (kg elemental K ha^{-1}), CPN=crop product N concentration % (100 × kg N kg^{-1} DM), CRN=crop residue N concentration % (100 × kg N kg^{-1} DM),Ypot= yield potential from Global Yield Gap Atlas (Mg DM crop product ha^{-1}).

* ** *Unique trial locations (based on GPS coordinates) were used as a proxy for estimating the number of trials per region.

the mass of accumulated dry matter (kg DM) and nutrient concentration expressed on a dry matter basis (g kg DM^{-1}). In this paper, we use CPY to denote the mass of dry matter accumulated in the marketable crop product, expressed on an area basis (Mg DM ha^{-1}) (Table 1). We define CRY as the mass of dry matter accumulated among all other above ground plant organs, expressed on an area basis (Mg DM ha^{-1}). Above ground yield (AGY) is the sum of CPY and CRY (Mg ha^{-1}).

Nutrient uptake, nutrient offtake, AGY, CPY, CRY, and HI data is scalable. They can be quantified for a single plant or summarized for a crop production region encompassing several countries. As these metrics are scaled, their sources of data become different. Published data at the sub-field to field scales usually come from scientific investigations. Each data set from this scale will often report several of the metrics. When compiled across many such studies, data on all metrics can be collected and summarized. Published data at sub-country to country scales often come from surveys conducted by governmental or industrial organizations and usually contain only CPY, fertilizer consumption, and/or fertilizer application rates. Such data are typically collected at this scale to inform decisions about markets. Thus, to calculate nutrient offtake and uptake at larger scales, nutrient concentration data assembled from smaller scale studies are combined with unrelated, survey data conducted at larger scales.

The following section will include rationale for including variables in models to test the three hypotheses (Table 2) in this study.

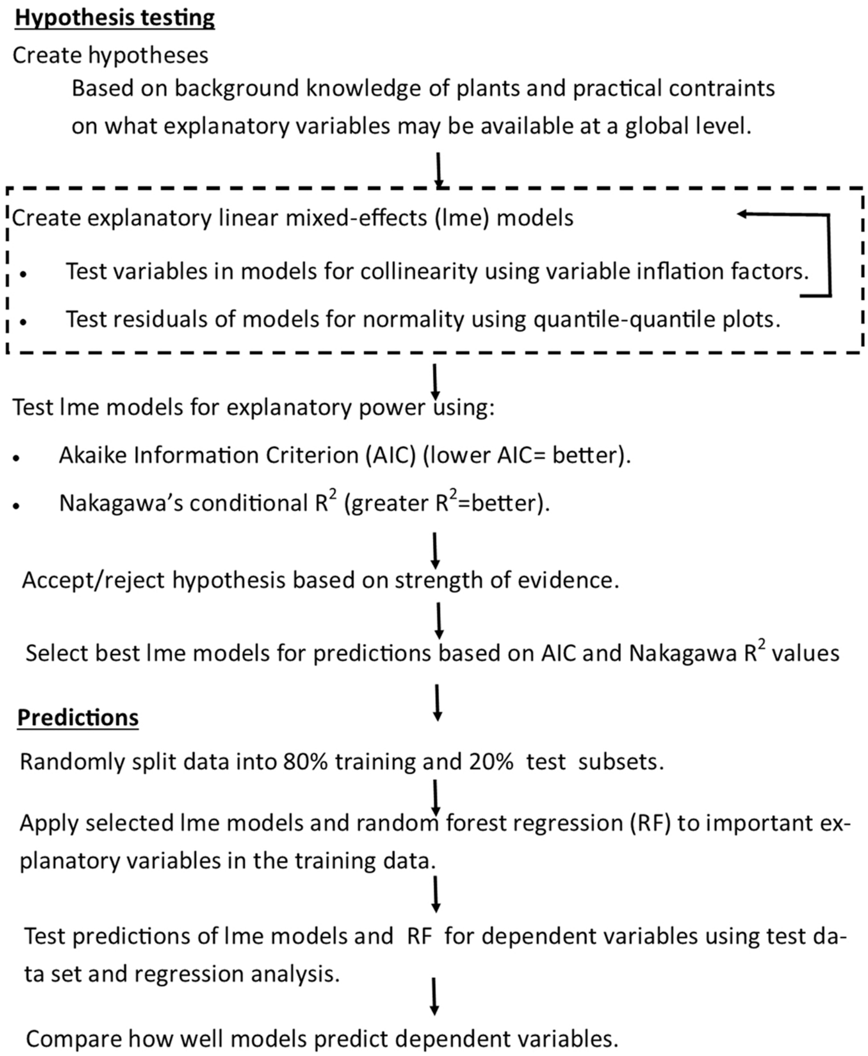

Fig. 2. Work flow used for investigating whether variation in crop product and crop residue nitrogen concentration and crop harvest index of maize can be explained using only a limited set of widely available variables.

2.2. *Hypothesis 1: harvest index*

2.2.1. *Crop product yield*

We hypothesize that HI will increase with CPY. In analyzing historical yield trends, Lorenz et al. (2010) pointed out that the relationship between CPY and HI was decreasing because some modern cultivars may be approaching a HI threshold by which further increases in grain yield will come from proportionate increases in biomass yield. They also speculated that maize in the USA may have already reached this threshold. Genetics plays a key part in variation in maize HI (Hay and Gilbert, 2001) and HI is a fairly stable trait, but variations in management practices (e.g. plant density, water supply, nutrient input) also affect HI (Evans, 1998; Trachsel et al., 2016). Unfortunately, there are no global aggregations of maize genetics and associated HI. However, CPY of maize differs by crop genetics, so CPY could be an alternative (widely available) explanatory variable for HI.

2.2.2. *N application*

A priori, the hypothesis is that there is an effect of N application on HI in maize, i.e. an increase in HI as a result of increased grain production relative to vegetative biomass (up to a maximum) due to higher N application. However, these effects may also be dwarfed by other factors. For instance, Hay and Gilbert (2001) showed no difference in HI for the same varieties of maize when no and adequate (200 kg N ha^{-1} $year^{-1}$) N fertilizer was applied, but a statistically significant difference of 0.1 (in HI) between 'Landrace' and 'Improved' varieties of maize when the same quantity of nutrients were applied to each variety. Furthermore, while Trachsel et al. (2016) showed a 0.03 greater HI in maize subjected to high levels of N fertilization compared with low nitrogen conditions, the difference in HI from N application paled in comparison with the 0.1–0.15 difference in HI between maize genetics.

2.2.3. *Yield potential*

Crop yield potential was defined by van Ittersum et al. (2013) as the *'yield of a crop cultivar when grown with water and nutrients non limiting and biotic stress effectively controlled*'. The yield potential estimated specifically for either rainfed or irrigated conditions will in this study be referred to as the 'Ypot' yield potential. Ypot values are estimated solely on crop growth determined by 'solar radiation, temperature, atmospheric CO_2 concentrations and genetic characteristics' (and water limitations in the case of rainfed or partially irrigated conditions) (www.yieldgap.org/glossary). Given how Ypot is estimated, it is therefore unsurprising that Ypot values are generally correlated to actual CPY values (van Ittersum et al., 2013). This offers an opportunity to use Ypot as an explanatory/predictive variable in the absence of, or in addition to, actual CPY data for predicting HI.

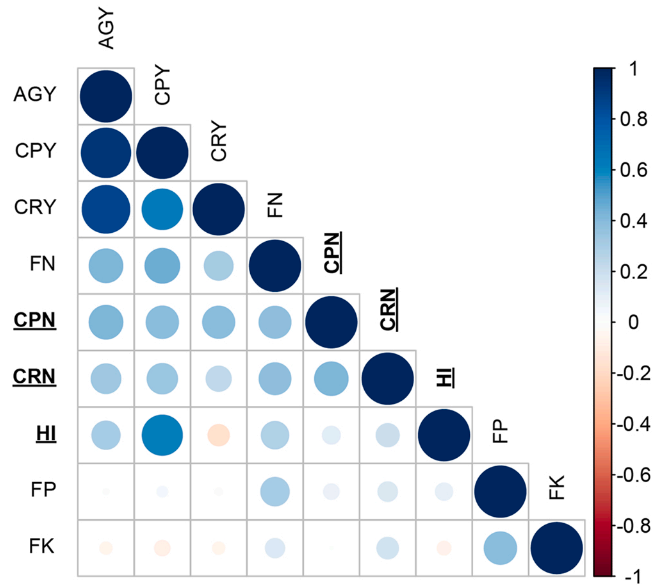

Fig. 3. Correlation (R^2) plot for a selection of variables from the maize dataset where variables are described in Table 1 (not all variables were included due to availability of data). Dependent variables are underlined.

Table 4
Akaike Information Criterion (AIC) and Nakagawa's conditional r squared (R^2) values for mixed effects models explaining observed variation in harvest index (HI). **Models in bold** were chosen as predictor models. Model variables are described in Table 1.

Model name	Data sub-selection*	Model equation ("Region" was used as a random factor in all equations)	AIC* *	R^2 * **
log_1	**CPY**	**HI~ a*log(CPY)+b**		**NA**
H1_1	**CPY**	**HI ~ CPY**		**0.54**
H1_2	**CPY**	**HI ~ CPY + CPY²**	**Best AIC**	**0.57**
H1_3	CPY	HI ~ CPY + FN		0.53
H1_4	CPY	HI ~ CPY + FN + FP		0.51
H1_5	CPY	HI ~ CPY + FN + FP + FK		0.45
H1_6	FN	HI ~ FN	Best AIC	0.31
H1_7	FN	HI ~ FN + FP		0.29
H1_8	FN	HI ~ FN + FP + FK		0.36
H1_9	Ypot	HI ~ Ypot		0.22
H1_10	**Ypot**	**HI ~ Ypot + CPY**	**Best AIC**	**0.46**
H1_11	RY	HI ~ RY	Best AIC	0.26

*Each data sub-selection included the same data for the variable listed in this column.
* *Akaike Information Criterion (AIC) is a measure of the relative quality of statistical models for a set of data whereby a more negative value indicates a greater relative quality compared with another model (Burnham and Anderson, 2002). This column indicates which model is best based on AIC for each sub-selection of the dataset, i.e. the sub-selection of the overall dataset used for the CPY, FN, Ypot and RY models.* **This is a conditional coefficient of determination for a linear mixed-effects model that is based on variance of the fixed and random effects (a value between 0 and 1) (Nakagawa et al., 2017).

2.3. Hypotheses 2 and 3: N concentration of crop products and crop residues

2.3.1. Crop product yield

CPY was shown to explain observed variation for CPN in maize (Ciampitti and Vyn, 2013). A meta-analysis by Liu et al. (2021) indicated that CPN decreased with increasing CPY of maize, rice and wheat.

Table 5
Akaike Information Criterion (AIC) and Nakagawa's conditional r squared (R^2) values for mixed-effects models explaining observed variation in crop product nitrogen concentration (CPN). **Models in bold** were chosen as predictor models. Model variables are described in Table 1.

Model name	Data sub-selection*	Model equation ("region" was used as a random factor in all equations)	AIC* *	R^2 * **
H2_1	CPY	CPN ~ CPY		0.26
H2_2	CPY	CPN ~ CPY + FN		0.37
H2_3	**CPY**	**CPN ~ CPY + FN+ FP**	**Best AIC**	**0.39**
H2_4	CPY	CPN ~ CPY + FN + FP + FK		0.38
H2_5	FN	CPN ~ FN		0.35
H2_6	**FN**	**CPN ~ FN + FP**	**Best AIC**	**0.36**
H2_7	FN	CPN ~ FN + FP + FK		0.35
H2_8	Ypot	CPN ~ Ypot		0.58
H2_9	Ypot	CPN ~ Ypot +CPY		0.63
H2_10	**Ypot**	**CPN ~ Ypot + FN* ** ***	**Best AIC**	0.67
H2_11	Ypot	CPN ~ Ypot + FN + FP		0.68
H2_12	Ypot	CPN ~ Ypot + FN + FP + FK		0.70
H2_13	RY	CPN ~ RY		0.58
H2_14	**RY**	**CPN ~ RY + FN**	**Best AIC**	**0.65**
H2_15	RY	CPN ~ RY + FN + FP		0.64
H2_16	RY	CPN ~ RY + FN + FP + FK		0.67

*Each data sub-selection included the same data for the variable listed in this column.
* *Akaike Information Criterion (AIC) is a measure of the relative quality of statistical models for a set of data whereby a more negative value indicates a greater relative quality compared with another model (Burnham and Anderson, 2002). This column indicates which model is best based on AIC for each sub-selection of the dataset, i.e. the sub-selection of the overall dataset used for the CPY, FN, Ypot and RY models.
* **This is a conditional coefficient of determination for a linear mixed-effects model that is based on variance of the fixed and random effects (a value between 0 and 1) (Nakagawa et al., 2017).
* ** *H2_10 was chosen instead of H2_11 (despite H2_11 having a better AIC) because it was shown that the coefficient for FP was non-significant (using 95% confidence intervals).

Likewise, analysis of maize data by Ciampitti and Vyn (2013) indicated a negative asymptotic relationship between maize nutrient concentration of numerous nutrients and mass of plant biomass.

Crop-nutrient models sometimes include ranges for maximum dilution and maximum accumulation of nutrients in the crop, such as used in the QUEFTS model for maize (Janssen et al., 1990) and its later applications (Setiyono et al., 2010). The maximum dilution of nutrients represents a situation where a particular nutrient is the 'grain yield limiting factor'. In this case the nutrient is diluted in the plant to its maximum extent and the grain yield is at its zenith given the amount of nutrient absorbed (Janssen et al., 1990). In contrast, the maximum accumulation of nutrients represents the situation where that particular nutrient is excessively available and the nutrient is maximally accumulated. In such case, grain yield is being limited or reduced by one or more growth limiting and reducing factors (apart from the nutrient concerned) (Janssen et al., 1990; van Ittersum et al., 2013).

The median slope of the maximal dilution and maximal accumulation of a nutrient are sometimes used to aid assumptions for crop nutrient balances at various spatial levels (Salvagiotti et al., 2021). For N in maize, Setiyono et al. (2010) estimated a value between 40 and 83 kg yield (kg N uptake)$^{-1}$, or as a corollary to these values, 0.025–0.012 kg N (kg DM)$^{-1}$. These relationships have been supported by multiple studies (Keulen and Heemst, 1982; Saidou et al., 2003; Xu et al., 2013; Shehu et al., 2019).

A direct relationship between CPY of maize and CPN or CRN has not been well defined. This can be partly explained by the fact that in many experiments, greater yield is achieved through greater application rates

Table 6
Akaike Information Criterion (AIC) and Nakagawa's conditional r squared (R^2) values for mixed effects models explaining observed variation in crop residue nitrogen concentration (CRN). **Models in bold** were chosen as predictor models. Model variables are described in Table 1.

Model name	Data sub-selection*	Model equation ("region" was used as a random factor in all equations)	AIC* *	R^2 * **
H3_1	CPY	CRN ~ CPY		0.30
H3_2	CPY	CRN ~ CPY + FN		0.40
H3_3	**CPY**	**CRN ~ CPY + FN+ FP**	**Best AIC**	**0.45**
H3_4	FN	CRN ~ FN	Best AIC	0.38
H3_5	Ypot	CRN ~ Ypot		0.56
H3_6	Ypot	CRN ~ Ypot + CPN		0.47
H3_7	**Ypot**	**CRN ~ Ypot + FN* ** ***		**0.58**
H3_8	Ypot	CRN ~ Ypot + FN + FP	Best AIC	0.58
H3_9	RY	CRN ~ RY		0.56
H3_10	RY	CRN ~ RY + CPN		0.48
H3_11	**RY**	**CRN ~ RY + FN**	**Best AIC**	**0.62**
H3_12	CPN	CRN ~ CPN		0.42
H3_13	CPN	CRN ~ CPN + CPY		0.44
H3_14	CPN	CRN ~ CPN + CPY + FN		0.48
H3_15	CPN	CRN ~ CPN + CPY + FN + FP		0.50
H3_16	**CPN**	**CRN ~ CPN + CPY + FN + FP + FK**	**Best AIC**	**0.30**

*Each data sub-selection included the same data for the variable listed in this column.
* *Akaike Information Criterion (AIC) is a measure of the relative quality of statistical models for a set of data whereby a more negative value indicates a greater relative quality compared with another model (Burnham and Anderson, 2002). This column indicates which model is best based on AIC for each sub-selection of the dataset, i.e. the sub-selection of the overall dataset used for the CPY, FN, Ypot and RY models.* **This is a conditional coefficient of determination for a linear mixed-effects model that is based on variance of the fixed and random effects (a value between 0 and 1) (Nakagawa et al., 2017).
* ** *H3_7 was chosen instead of H3_8 (despite H3_8 having a better AIC) because there was little to no advantage in terms of Nakagawa R^2 values for including the FP variable.

of nutrients, confounding the relationship between nutrient concentration and crop yield. This highlights the importance of assessing the potential of application rates of nutrients (especially N) to predict CPN and CRN, in combination with CPY.

2.3.2. N application

Application of nutrients has long been known to affect nutrient concentration of crop products and crop residues. For example, Dilz (1971) showed that N concentration of wheat grain and residue were closely (positively) related to increasing N application rate, albeit with lower absolute N concentration of the crop residue compared to the crop product. Numerous studies have highlighted positive relationships between the quantity of N applied and the CPN of maize (Correndo et al., 2021).

2.3.3. Yield potential

An alternative approach to using CPY values for improving estimates of CPN and CRN is to use estimates of yield potential for each location. The Ypot information may provide superior explanatory power compared with CPY alone because unlike CPY, Ypot is not confounded by application rates of nutrients and other agronomic factors such as differences in crop protection and plant density. It is known from N response and genetic gain trials that N concentration in harvested product decreases with increasing Ypot (due to 'nutrient dilution'), at a given N rate (Peng et al., 2022). In other words, N application rate relative to Ypot would govern N concentration. Hence combining both these predictors simultaneously could make sense. Alternatively, if actual yield is N limited, the ratio of actual yield to potential yield (CPY/Ypot, i.e. relative yield potential, RY) already incorporates both N rate (and Ypot); therefore this ratio might prove a suitable predictor for N concentration of crop components in models that leave out N application rate.

RY could therefore be used as an independent variable to estimate N concentration of crop products and residues. When CPY is below Ypot, this indicates one or more yield limitations exist. If N is the primary limitation, then CRN and possibly CPN will tend toward 'maximum dilution' as defined in the QUEFTS framework. Under this condition, CPY is linearly related to total N uptake. If N is not the primary limiting factor, then CPN and CRN are more likely to tend toward 'maximum accumulation.' The strength of the linear relationship between CPY and CRN and/or CPN when CPY< <Ypot thus depends on the degree to which N is the primary cause of the yield limitation.

3. Material and Methods

3.1. Data collection

Two main sources of data were used in this study. The first source resulted from a request for raw experimental field data sent out to researchers and organizations around the world. The second source resulted from a data search from the peer-reviewed literature. In situations where we found summary data from the literature and received raw data from the same experiments through the data request, only the raw data were used in the statistical analysis to avoid duplication.

3.1.1. Data from requests

Approximately 330 individuals were invited to contribute raw, replicated field experiment data to this study. Individuals were chosen based on the authors' contacts as well as through contact details available within published peer reviewed articles. Those who were asked for data were requested to share variables with the minimum and optional data requirements shown in Appendix A.

3.1.2. Data from peer reviewed literature

Data from a literature search were also included in the dataset for analysis. Terms used in Google Scholar (https://scholar.google.com/), CAB Abstracts (https://www.cabdirect.org/) and Ovid (https://ovidsp.ovid.com/) literature search platforms were as follows: *(nitrogen) AND (maize OR corn) AND ("nutrient concentration" OR "nutrient content")*. Some variations on these search terms were made so they were aligned to the formatting requirements of each search engine. Additionally, an email search alert was set up using these terms in Ovid and Google Scholar between the 1st of August 2020 and the 5th of October 2021. The list of articles was refined to include only articles from a peer-reviewed journal or an accepted university thesis, and which included replicated field experiment results for fertilizer application rate, grain and/or residue yield, grain and/or residue nitrogen concentration, and experimental information such as location, year of experiment and experimental design. In total, the queries from the multiple sources resulted in 91 articles (Appendix B). Data from the articles were manually converted into a standardized Excel format (Microsoft Office, Microsoft, Redmond, Washington, USA). Where data were not available in the publications as tables, data were extracted from figures using the GetData Graph Digitizer software (version 2.26, http://www.getdata-graph-digitizer.com/).

Experiments included in the database covered a wide range of countries with the USA, India, China, Indonesia, Philippines, Argentina, and Hungary being countries with the most data (Fig. 1). Over 86% of total world maize grain production (for the 2018 reporting year) is represented by the 31 countries included in the dataset for this study (FAOSTAT, 2020).

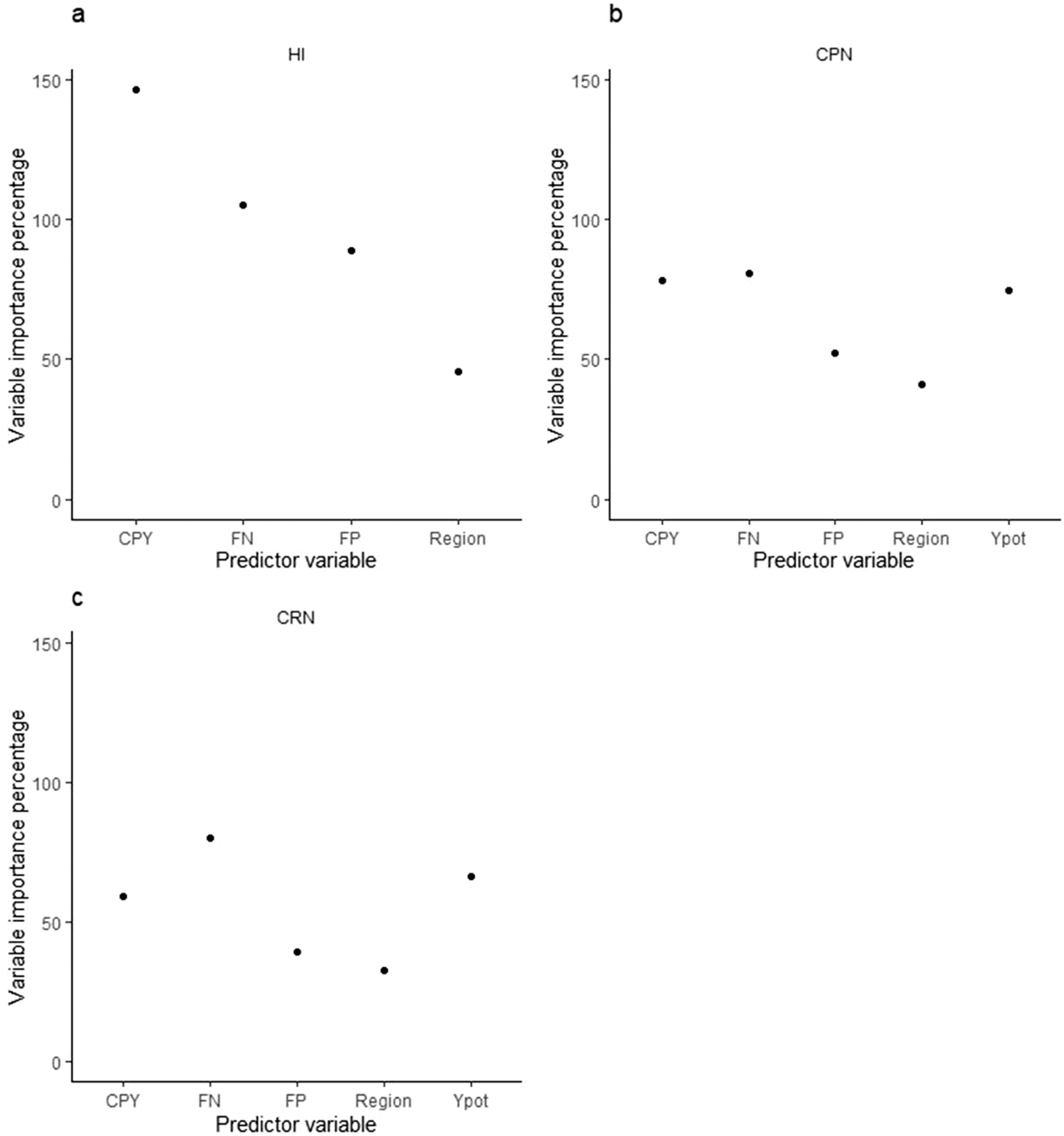

Fig. 4. Relative importance of predictor variables for accuracy of the random forest models for harvest index (HI-plot a), crop product nitrogen concentration (CPN-plot b), and crop residue nitrogen concentration (CRN-plot c). Relative importance of individual variables is defined as the percentage increase in mean square error of predictions when that variable is excluded from the prediction model.

3.2. Processing data before analysis

Before analysis, data were first assessed for outliers. Summary statistics and visualizations using the "ggplot2" package in R version 4.1.0 (R Core Team, 2020) (herein referred to as 'R') were used to determine obvious errors or outliers in the data. From this process, only a limited set of (HI) data were categorized as being outliers as follows. HI values above 0.65 were deemed biologically infeasible outliers and were excluded from analysis based on maximum HI values in a review by Unkovich et al. (2006).

Data collected from the literature were generally expressed as means across multiple replicates, whereas raw data from the data requests were values for each replicate. For a fair comparison of the data between the two data sources, the mean values were calculated with a weighting placed on the number of replicates and years of data from each source. For example, if a mean value for HI was available over a 2-year period from a field experiment with three replicates, a value of $3 \times 2 = 6$ replicates was used to weight that mean. This was to induce a greater weighting of values from longer runtime or more replicated trials, following Linquist et al. (2012). While there are other more complicated weighting methods, the alternatives often require estimates of variance in parameters to be available. In some cases the methods can overcome some missing data for variance of the key parameters. However, in most cases no estimates of variance were available in the studies included our analysis (80–90% of articles did not have estimates of variance, depending on which variable was examined). Given the large proportion

Table 7
Prediction accuracy of linear mixed-effects models based on the R^2 of predicted versus actual values for the dependent variable.

Model name	Model equation (where "region" was used as a random factor)	R^2 of predicted versus actual*	Mixed-effects model used for comparison with random forest?
log_1	**HI = a × log(CPY) +b**	**0.43**	**Yes**
H1_1	HI ~ CPY	0.31	
H1_2	**HI ~ CPY + CPY²**	**0.34**	**Yes**
H1_10	HI ~ Ypot + CPY	0.18	
H2_3	CPN ~ CPY + FN + FP	0.09	
H2_6	CPN ~ FN + FP	0.10	
H2_10	**CPN ~ Ypot + FN**	**0.24**	**Yes**
H2_14	CPN ~ RY+FN	0.21	
H3_3	CRN ~ CPY + FN + FP	0.22	
H3_7	**CRN ~ Ypot + FN**	**0.24**	**Yes**
H3_11	CRN ~ RY + FN	0.27	
H3_16	CRN ~ CPN + CPY + FN +FP+FK	0.20	

*See Appendices E to I for regression plots.

of data with no indication of variance, and for the purposes of making our method easier to reproduce we did not apply the alternative weighting methods to our data. In addition, when the same analysis was performed on 'unweighted' data it did not have a material effect on results (data not shown).

Furthermore, key variables were visually assessed for correlations using the "chart.Correlation" function in R. For hypothesis testing, the data points were converted into unit values that were of a similar order of magnitude. CPY and Ypot were therefore in units of megagrams (Mg) of crop product dry matter per hectare, and fertilizer application rates were in kilograms of elemental nutrient applied per square meter.

Grain yields on a fresh weight basis were converted to a dry matter basis using the fresh grain to grain dry matter conversion factor of 0.845 kg dry matter per kg (fresh) grain weight (McKevith, 2004). Some sources of data had grain and/or residue yield and nutrient uptake of grain and/or residue available, but did not include grain and/or residue nutrient concentration. In these cases the nutrient concentration was interpolated by dividing the uptake of the applicable nutrient for that plant component by the dry matter yield of that plant component. If crop product protein concentration data was available, but CPN was not available, the grain protein concentration of maize was divided by a factor of 6.25 (WHO and FAO, 2007) to get an estimate of the CPN.

Non-water limited yield potential (Yp) and water limited yield potential (Yw) values (van Ittersum et al., 2013) were included in the analysis using data from the Global Yield Gap Atlas (www.yieldgap.org) where available. In cases where the article indicated that the site had been irrigated the Yp was used to represent the Yield Potential (Ypot). In cases where articles did not indicate that irrigation had been applied to the site, the Yw was used to represent the Ypot value. The RY for each location were calculated by dividing the actual CPY by the Ypot.

Some sites did not have Ypot data available for maize in the Global Yield Gap Atlas. Sites without Ypot data were assigned (where possible) Ypot values predicted using random forest using growing degree days, aridity index, temperature seasonality, latitude, duration of growth from sowing to harvest, total seasonal precipitation, total available water capacity of the dominant soils in each area as predictor variables (Appendix C). These predictions were made using the "randomForest" package (version 4.6–14) in R.

For reference, a list of the relevant variables (and their abbreviations) in the combined dataset and the units is shown in Table 1.

3.3. Summary statistics

As shown in Table 3, the mean CPY across all countries included in the dataset (defined as the world mean values) was 7.73 Mg DM ha^{-1} with a coefficient of variation (CV%) of 48%; mean HI was 0.45 (CV%= 22%), mean CPN was 1.32% (CV% =23%), and mean CRN was 0.73% (CV%=39%). The region that had the greatest mean CPY was North America (10.5 Mg DM ha^{-1}, CV%=27%), with the Africa region having the lowest mean CPY (3.71 Mg DM ha^{-1}, CV% =47%). The region with the greatest CPN was West Asia (1.7% DM as N, CV% =14%), while the lowest mean CPN was observed in Oceania (1.13% DM as N, CV% = 12%). West Asia had the greatest CRN (1.03% DM as N, CV% =1%), while the lowest mean CRN was in Latin America (0.45% DM as N, CV% = 31%). However, it should be noted that the mean values are representative across all treatments in the data provided and may not represent averages for the typical farm management in each region.

3.4. Hypothesis and prediction accuracy testing

The main aim of this study was to investigate whether observed variations in maize HI, CPN and CRN can be explained using only a limited set of widely available variables. As shown in Fig. 2, this analysis was split into two forms of investigation:(1) hypothesis testing; (2) prediction accuracy.

For the hypothesis-testing aspect of this investigation, we used mixed-effects models. They differ from simple linear regression models in that they can include random factors as random effects (Burnham and Anderson, 2002). Mixed-effects models allow the user to examine a condition of interest whilst also accounting for variability within and across random factors (Brown, 2021). This method also handles missing data and unbalanced experimental designs well, making it a superior analytic tool over ANOVA when a wide range of random factors are present. The breadth of locations in our data set necessitated the use of region as a 'random' factor in our models to account for inter-regional variability and hence the need to use mixed-effects models.

Following, we compared the accuracy of the mixed-effects model predictions with predictions based on the machine learning random forest method. The comparison was made to assess whether there was congruity in conclusions between the two methods. Random forest predictions are robust in solving non-linear problems, can work with large data sets and have performed well in many machine learning applications (Jiang et al., 2021). In addition, simple linear regressions were also included in this analysis to assess the effect of excluding region as a random variable from the linear mixed-effects models with the greatest prediction accuracy.

3.4.1. Statistical methods for testing hypotheses (Fig. 2)

Three hypotheses were used to guide the testing of the explanatory power of different mixed-effect models, as listed in Table 2. The equations were developed based on background knowledge (Heinze et al., 2018) of what independent ('predictor') variables made biological sense in terms of their ability to explain variation in the dependent variable and to avoid collinearity. To explain observed variation in HI and CPN (hypothesis 1 and 2 respectively) the equations were defined as being CPY, fertilizer N application rate (FN), Ypot, and RY models to provide structure to the model testing. To explain observed variation in CRN (hypothesis 3), CPN models were used in addition to the CPY, FN, Ypot

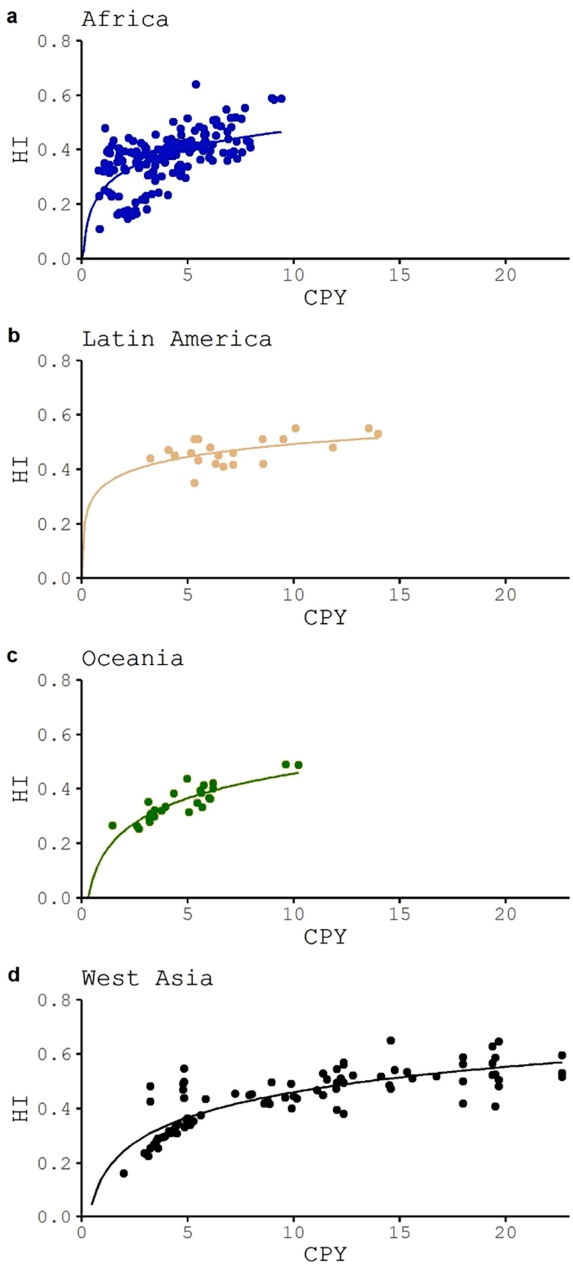

Fig. 5. Visualization of the 'log_1' prediction curves (where crop product yield-CPY- was log-transformed) estimated by region for harvest index (HI) in relation to CPY (in megagrams per hectare).

and RY models.

The CPY and FN models were chosen to test the expected relationships between CPY and FN with HI, CPN and CRN. The Ypot and RY models were tested based on the assumption that they may provide an improved predictor variable compared with the CPY predictor variable as explained in Section 2.3. While CPN is often not a widely available statistic at a country level, CPN was included as a predictor variable for CRN to assess how much better predictions of CRN could be made if CPN data were available.

The explanatory variables in each equation were tested for collinearity using the variable inflation factor (VIF) R. The list of equations for testing the three hypotheses was altered to ensure predictor variables had a VIF less than 10 (which means that there is very low collinearity) and that there were enough data available. The assumption of normality and homoscedasticity of residuals for each model was also tested visually using the normal quantile-quantile plot (NIST/SEMATECH, 2012). For the quantile-quantile plots we looked for linearity and for the residual plots we looked for lack of discernable patterns.

Interaction effects between explanatory variables were not included in the models to reduce the probability of introducing false inferences or type 1 errors (whereby significant effects are seen in some variables by chance) (Matuschek et al., 2017; Harrison et al., 2018).

The equations in Table 2 were included in linear mixed-effects models using the "lmer" function in R Base, assuming 'region' was a random effect.

The Akaike Information Criterion (AIC) (Akaike, 1998) and the Nakagawa conditional coefficient of determination (R^2) (Nakagawa et al., 2017) were applied to the linear mixed-effects models using Base R and the "performance" package in R respectively (Lüdecke et al., 2020). The AIC and Nakagawa conditional R^2 equations were chosen, because traditional methods of estimating standard errors of regressions and coefficient of determinations for simple linear regression models

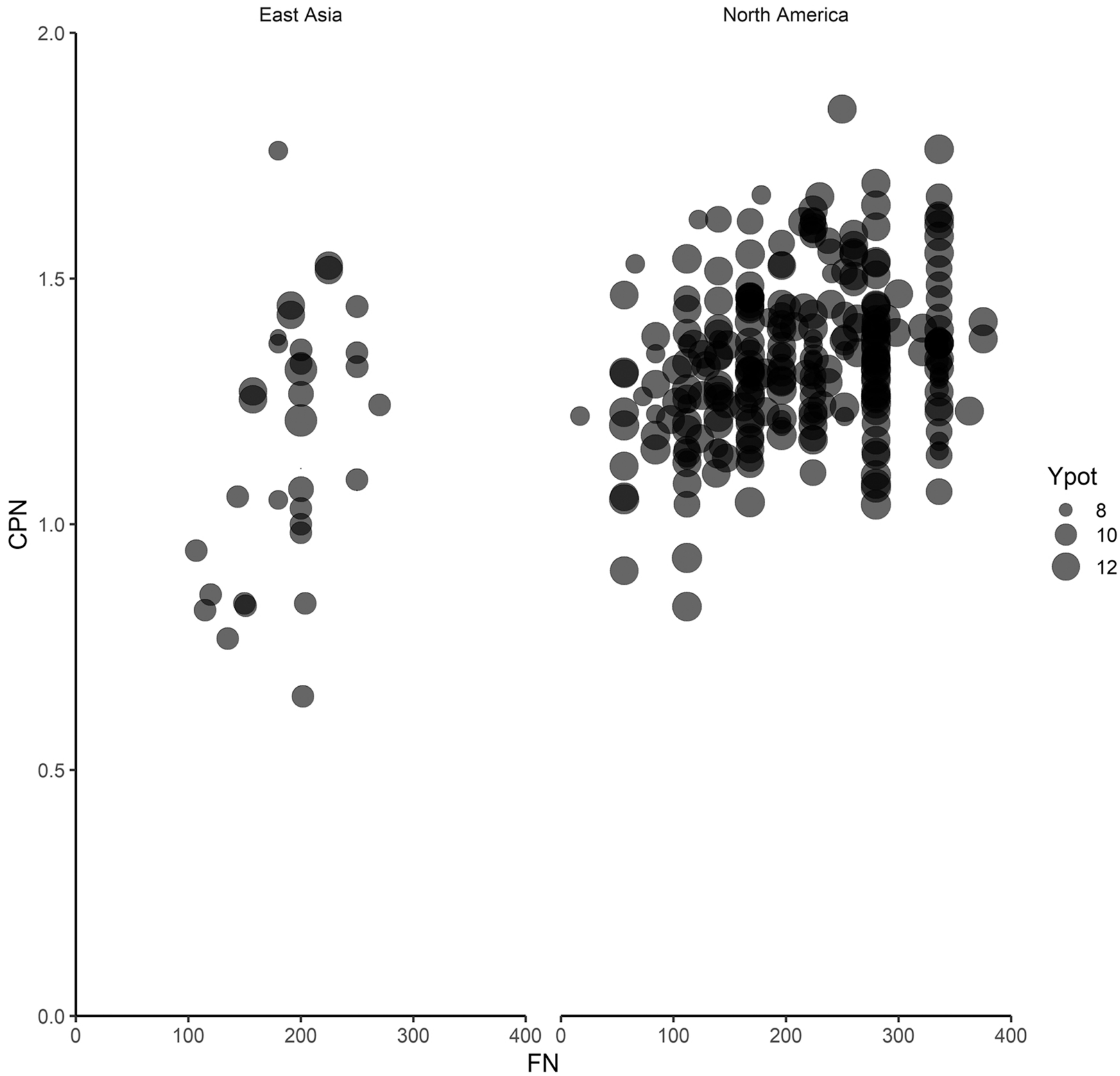

Fig. 6. Crop product nitrogen concentration (CPN as a % of dry matter) in relation to yield potential (Ypot in Mg dry matter per ha) and fertilizer nitrogen (N) application rate (FN, in kg N per hectare), for East Asia and North America.

cannot be applied to linear mixed-effects models (Burnham and Anderson, 2002; Nakagawa et al., 2017). The AIC is a value 'to select a parsimonious (most economical) approximating model for the observed data' (Burnham and Anderson, 2002 p.157) and is a relative value whereby lower numbers denote a more preferred model. Comparisons of AIC between different models that used different data is not possible. Therefore, when AIC were presented in results tables the 'best' model (based on AIC) was indicated for within each 'group' of models that used the same data. These included the CPY, FN, Ypot, RY and CPN (for CRN as the dependent variable only) groups of models (Table 2). The AIC values were estimated using the maximum likelihood framework (Pinheiro and Bates, 2000; Madden et al., 2016). However, use of the maximum likelihood framework in mixed-effects models is known to overestimate variance components. To overcome this, when the mixed-effects models were used to estimate variance components (or to perform predictions), the restricted maximum likelihood (REML) framework was used to avoid the biases that would otherwise occur when using the maximum likelihood framework (Pinheiro and Bates, 2000).

The conditional Nakagawa R^2 values were used as the best approximation of the coefficient of determination to determine how well the model predicts the observed outcomes. The conditional Nakagawa R^2 is estimated based on the variance of the fixed and random effects (Nakagawa et al., 2017). More specifically, the 'performance' package we used to estimate the conditional Nakagawa R2 values in this study 'iteratively removes predictors of interest from the model and monitors the change in the variance of the linear predictor. The difference to the full model gives a measure of the amount of variance explained uniquely by a particular predictor or a set of predictors' (Stoffel et al., 2021).

Linear mixed-effects models were subjectively selected as predictor models based on how parsimonious they were (based on AIC value) as well as how well they explained variation in the dependent variable (based on the conditional Nakagawa R^2 values).

3.4.2. Methods for testing prediction accuracy of models

The data were randomly split into two subsets with 80% of data apportioned to a 'training' dataset, and the remaining 20% apportioned to a 'testing' dataset. For predictions using linear mixed-effects models, the lme4 package (version 1.1–27.1) of R was applied to the models listed in Table 2 (using the 80% of data set aside for training the model) to create predictions for the 20% of the data that was set aside for testing. For predictions using random forest the "randomForest" package (version 4.6–14) in R was applied to the 80% training dataset with the 'trained' random forest model then applied to the test dataset. Predictions were compared with the actual (measured) values using linear regression analysis to determine the coefficient of determination (R^2). In

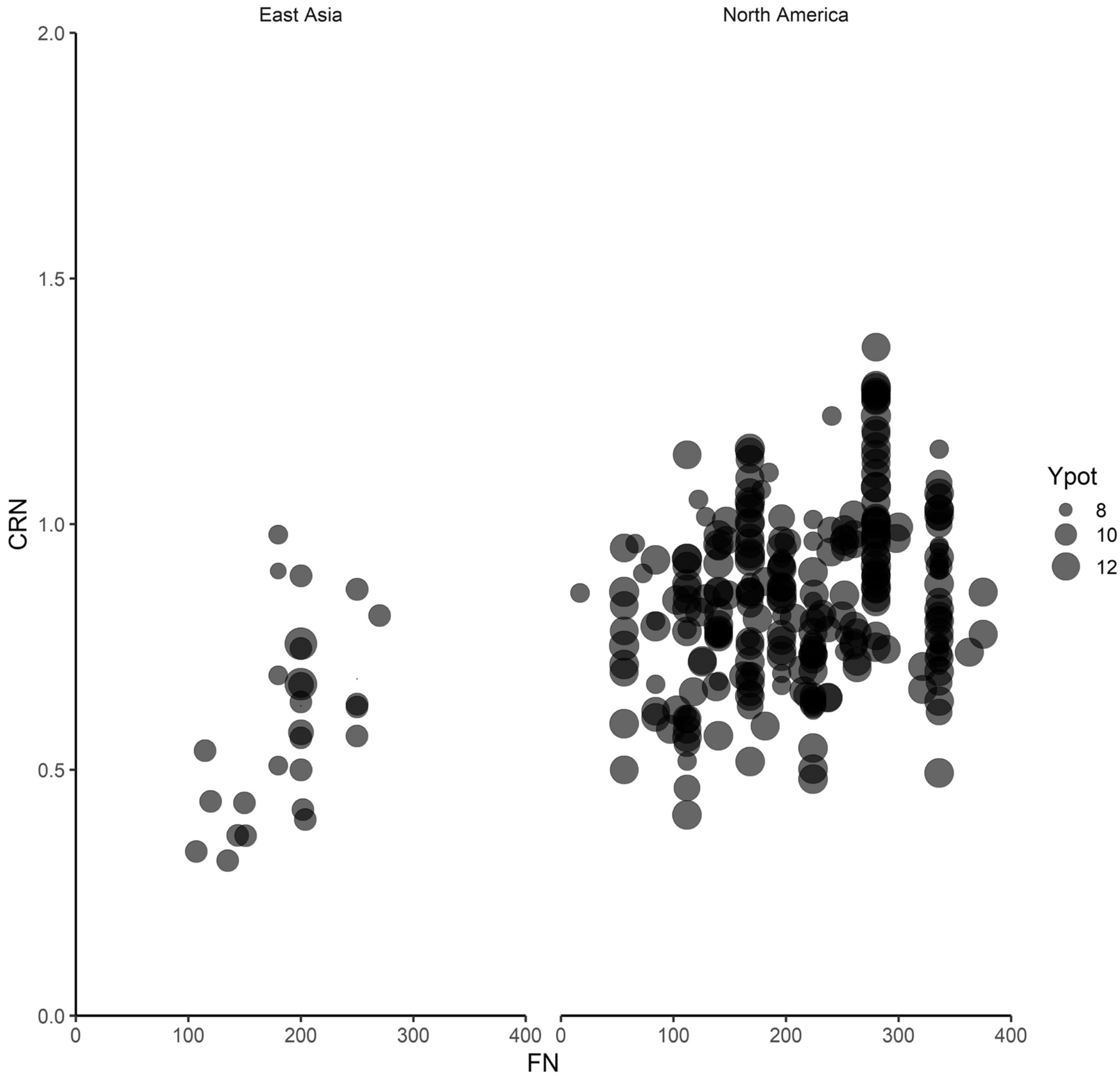

Fig. 7. Crop residue nitrogen concentration (CRN as a % of dry matter) in relation to yield potential (Ypot in Mg dry matter per ha) and fertilizer nitrogen (N) application rate (FN, in kg N per hectare), for East Asia and North America.

addition, CPY was log-transformed to predict HI using the formula (HI = a*log(CPY)+b) with region as a random variable. The 'a' represented the factor to multiply against the log of CPY and 'b' represented a value to vary the height in plateau of the HI curve. This model was included in the analysis as it was deemed more biologically realistic to assume a logarithmic (near) plateau in HI at a certain point as compared to a linear increase (assumed in model 1 for HI), or a quadratic relationship (assumed in model 2 for HI).

As with the linear mixed-effects model, the number of variables included in the random forest model was limited to those one could reasonable expect to be widely available for the selected crops at country levels (region, Ypot, CPY, FN, and FP) and which were shown to contribute most explanatory power in the linear mixed-effects model (hence FK was not included for any dependent variable and Ypot was not included for the HI dependent variable). In addition to regression analysis, the random forest method had its predictor variables assessed visually using the "varImpPlot" function in R Base to determine the relative importance of the different variables in the predictions. Relative importance was defined as the percentage increase in mean square error of predictions when that variable was excluded from the prediction model.

4. Results

4.1. Correlations

There was strong correlation between CPY and AGY (0.92) and CRY and AGY (0.86), and a 0.61 correlation between CPY and HI (Fig. 3). Appendix E includes the two dimensional linear regressions between many of the key variables to illustrate the linearity (or otherwise) of relationship between variables.

4.2. Observed variations in the key variables and their explanation

4.2.1. Harvest index

CPY was the best determinant of HI, when used in a (quadratic) linear mixed-effects model, or when used with Ypot (Table 4).

In each table the 'best' model (based on AIC) is indicated for each sub-selection of the dataset. For HI, the Nakagawa conditional R^2 values (referred herein to as R^2 values) were up to 0.57 across the range of models. Three linear mixed-effects HI models were chosen for use as a predictor model (see models in bold in Table 4) based on how parsimonious the models were (based on AIC) and their R^2 values. The model which had a log-transformation of CPY was chosen, despite the fact that R^2 values cannot be estimated for log-transformed models. It is

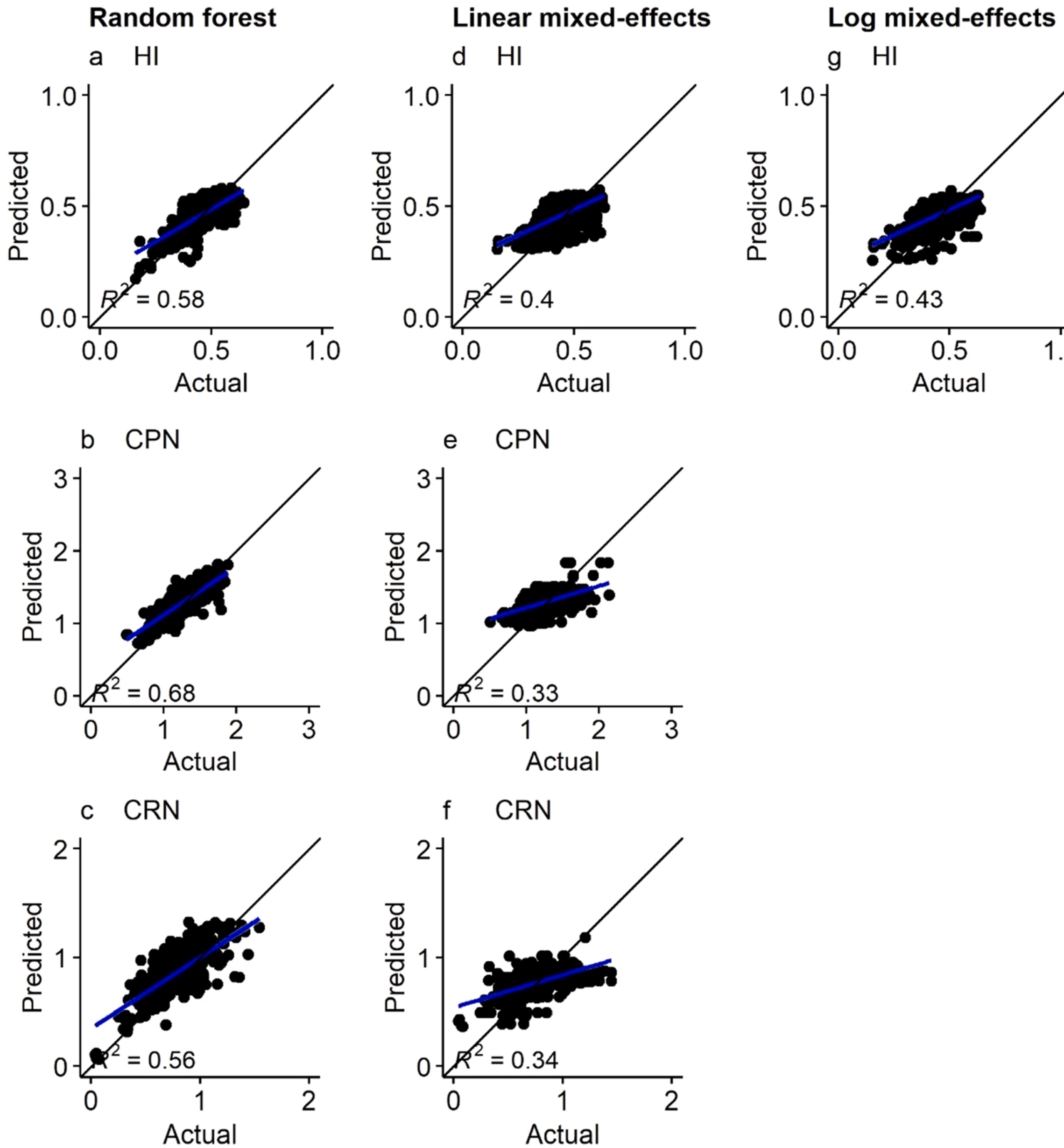

Fig. 8. Linear regression of the predicted versus actual, harvest index (HI), crop product nitrogen percentage (CPN) and crop residue nitrogen percentage (CRN) for maize based on random forest prediction (left-most column), linear mixed-effects models (second to left-most column), and log-transformed mixed-effects regression (right-most column). Linear mixed-effects model H1_2 was chosen for HI, model H2_10 was chosen for CPN, and model H3_7 was chosen for CRN.

important to note that the addition of fertilizer application rate information as an independent variable did not improve the AIC or R^2 values for the CPY models, indicating fertilizer application rate information was not good for inclusion in a prediction model of HI.

4.2.2. Crop product nitrogen concentration

FN was consistently a good variable to explain variation in CPN when included in the model on its own or in a model in combination with the CPY, FP, Ypot or RY explanatory variables (Table 5).

For CPN, the R^2 values were up to 0.70 across the range of models. Four CPN models were chosen for use as a predictor model (see models in bold in Table 5) based on how parsimonious the model was and their R^2 values. A common theme of the chosen models was that they all had FN as an independent variable.

4.2.3. Crop residue nitrogen concentration

Similarly to CPN, FN improved the power of the linear mixed-effects models to explain observed variation in CRN (Table 6). The R^2 values of the linear mixed-effects models were up to 0.62. Four CRN models were chosen for use as a predictor model (see models in bold in Table 6) based on how parsimonious the model was and their R^2 values. CPN was shown to be an important explanatory/predictor variable for CRN. The linear mixed-effects models that included CPN had some of the greatest R^2 values. Our study therefore shows that CPN could be a useful variable for improving estimates of CRN.

4.3. Importance of variables in random forest models

Fig. 4 illustrates the relative importance of the variables used in the random forest models. For HI, it shows that CPY is the most important variable for predictions followed by FN, FP, then region. For CPN, Ypot, FN and CPY were seen to be nearly equally important variables for predictions followed by FP and region. For CRN, FN was the most important variable for predictions followed by Ypot, CPY, FP and region.

4.4. Prediction accuracy of linear mixed-effects models and random forest models

Prediction accuracies of the best mixed-effects models selected from Tables 4–6 are shown in Table 7 (with coefficients for the log-transformed model shown in Appendix F). The mixed-effects model that had a log function applied to CPY had the greatest prediction accuracy R^2 of 0.43 for HI (Table 7, Appendix G and H). The model with the greatest prediction accuracy for CPN was model H2_10 with an R^2 of 0.24 (Table 7, Appendix I). The next best model was model H2_14 with an R^2 of 0.21. Model H3_11 had the greatest prediction accuracy for predicting CRN (R^2 =0.27) followed by model H3_7 with a R^2 of 0.24 (Table 7 and Appendix J). The log_1, H1_2, H2_10, and H3_7 models were chosen for comparison with the random forest model because of their high prediction accuracies and how parsimonious they were relative to the other models.

When region was not included in the linear models (using simple linear regression) the prediction accuracies of those models were less than the equivalent linear mixed-effects models (Appendix K) and did not affect the overall conclusions of this study.

Eqs. 1 and 2 include the coefficients to predict CPN and CRN (based on linear mixed effects models H2_10 and H3_7). In Eqs. 1 and 2, CPN and CRN are the crop product and crop residue N concentrations respectively (% of DM), Ypot is the yield potential of crop product in Mg dry matter per hectare and FN is the quantity of fertilizer N applied in kg N per m^2. The value 'a' in Eq. 1, was 1.27 for Africa, 1.15 for East Asia, 1.22 for Latin America, 1.31 for North America, 1.64 for South Asia and 1.89 for West Asia. The value 'a' in Eq. 2, was 0.82 for Africa, 0.78 for East Asia, 1.0 for North America, and 1.34 for South Asia.

$$\text{CPN} \sim a - 0.013 \times \text{Ypot} + 9{\cdot}56 \times \text{FN} \tag{1}$$

$$\text{CRN} \sim a - 0.029 \times \text{Ypot} + 8{\cdot}46 \times \text{FN} \tag{2}$$

Fig. 5 provides a visualization of prediction curves for the mixed-effects models (that had a log function applied to CPY) for estimating HI by region. It shows how the log_1 model curve represents the HI data well in most of the regions.

Fig. 6 shows the observed relations between CPN, FN and Ypot. With more N application, CPN increases. At the same time, greater Ypot at a similar N application leads to a lower CPN.

Similar patterns for CPN shown in Fig. 6 are evident for CRN in Fig. 7. As with CPN, the is a positive relationship between FN and CRN and data points with greater Ypot tended to have lower CRN at the same level of FN.

Fig. 8 compares the prediction accuracies of the 'best' random forest, mixed-effects models. It shows that prediction accuracies of the random forest models were greater than the mixed-effects models. The mixed-effects model (that had a log function applied to CPY) had a similar prediction accuracy for HI (R^2 =0.43) as the best (quadratic) linear mixed-effects model (R^2 =0.40). However, it is more biologically reasonable to expect a plateau in HI (predicted using a log-transformed model) compared with an initial increase then subsequent decrease in HI (at greater CPY) as predicted by the quadratic model.

5. Discussion

A major objective of this study was to enable more locally relevant estimation of crop nutrient offtake to improve the overall accuracy of nutrient balances at a national, regional or global level and thus provide better guidance to more sustainable nutrient management. Our method used data from the literature and raw data files to analyze results from all treatments at the same time. In this way our study could be classified as multiple-treatment (network) meta-analysis, as has been developed in medical science, and is becoming more common in agricultural fields of research (Madden et al., 2016).

Results of our analysis indicate that crop HI, CPN and CRN can be estimated for maize using a limited set of independent variables which are widely available, including CPY, FN and Ypot. The prediction methods for these variables used in this study can be used to create nutrient balances estimated with more locally relevant CRY, CPN and CRN.

Our data analysis was based on both hypothesis testing and testing prediction accuracy of either mixed-effects or random forest models. Regardless of the method used, it is promising that in most cases the methods seemed to place similar levels of importance on the selected variables that are relatively widely available at a global level. This provides supporting evidence that predictions of HI, CPN and CRN can be made with only a limited set of predictor variables. In turn, this suggests that crop nutrient removal can be estimated much better, using an estimation of crop residues (often absent) and differentiated concentrations of nutrients in crop product and crop residues.

Greater prediction accuracy was found using random forest models compared with the linear mixed-effects models. Other research has highlighted the same relative accuracy advantage of random forest compared with regression models (Philibert et al., 2013; Jeong et al., 2016). However, it is important to note that the nature of the random forest method makes it impossible to produce an equation that shows how those predictions were made (unlike equations shown for linear mixed-effects models). Furthermore, random forest cannot precisely analyze the effects of different independent variables on the dependent variables, or tell whether the variable had a positive or negative effect (Philibert et al., 2013). This can make random forest more difficult to interpret, given its algorithm consists of a large number of decision trees that in many cases may not be described mechanistically (Jeong et al., 2016). On the other hand, from a practical and applied perspective, the random forest method has proven its ability to handle larger datasets and to improve predictions over time with a wide range of variables, and as more data become available (Jiang et al., 2021). If one is only interested in prediction accuracy, and the flexibility of adding new data or other variables to improve predictions, then random forest offers a feasible solution. However, if one wants to better understand the underlying causative factors that contribute to the predictions then the linear mixed-effects model could be the preferred method, despite its lower prediction accuracy.

Although results of this study are based on maize, the methods described in this study can equally be applied to other crops and nutrients, subject to availability of data. Fortunately, data for the predictive variables we employed are likely to become more widely available in future. This is because there is increasing interest in developing and sharing open databases that relate to crop nutrients (https://www.precisioncropnutrition.net/data/). A standardized and open crop nutrient dataset will allow users to apply the statistical and machine learning methods described in this study to the open sources of data. This is in contrast to using the mean HI and N concentration values from Table 3 in this study. Use of mean values shown in Table 3 for (say) country scale N balances has the disadvantage that values will only be reflective of the conditions from the included trials. This means they may not reflect actual farmer practice in each region. The mixed-effects and random forest prediction methods described in this study when applied to the open datasets will allow estimates of HI and N concentration of maize to vary based on region and by what were seen to be important predictor variables such as CPY for HI, and Ypot and FN for N concentration of crop products and crop residues. Application of values (for predictor variables) that are representative of farmers practice in each location should provide more locally relevant estimates compared with the mean values from the field experiments. For instance, Ypot values should represent the availability of water at each location with water-limited yield potential (Yw) values used for rainfed locations, and non-water limited (irrigated) yield potential (Yp) values used for irrigated locations. As we expand the open dataset these estimates will be improved. The authors' are therefore interested in receiving data from any crop. This is because contributions of data will help improve the accuracy and granularity of the predictor variables which in turn will improve estimates of HI, CPN and CRN. This would provide more locally relevant estimates of crop nutrient offtake to improve the accuracy of nutrient balances at a global level and provide opportunities for more sustainable nutrient management.

Further discussion on the relationships for HI, CPN and CRN is included in the following sections.

5.1. Harvest index

CPY gave the best explanatory and predictive power to describe HI, based on both the testing of hypothesis and model prediction accuracy. Fertilizer nutrient application rates did not contribute to the explanatory power of models for HI. This supports previous research indicating fertilizer N application rates had a negligible effect on maize HI (Hay and Gilbert, 2001). It is interesting to note that there was congruence in the relative importance of CPY in predicting HI when using random forest and the linear mixed-effects models.

To account for a plateau in HI at greater CPY, our analysis indicated that it was better to apply a log transformation of CPY in the prediction equation compared with use of a standard linear mixed-effects model. While a quadratic equation was shown to have a similar prediction accuracy as a model where CPY was log-transformed, it is more realistic to assume there will be a plateau in HI than a quadratic curve for HI, because a quadratic curve assumes HI will decrease after it peaks which is unlikely. The mixed-effects model which had CPY log-transformed is therefore the recommended model for predicting HI.

Our work is the first of its kind to explicitly show a relationship between yield potential - as defined by van Ittersum et al. (2013) - and HI. This is logical considering CPY is a component of HI and genetic progress of CPY (and hence potential yield) in crops has been strongly related to HI (Evans, 1998). In contrast to Ypot, the relationship between CPY and HI has been analyzed more often in the literature (Lorenz et al., 2010). However, our study, does not support the finding made by Lorenz et al. (2010) that HI did not change or did not follow the trend in CPY. Our analysis indicates that accounting for an increasing trend in HI for maize with greater CPY is a reasonable assumption to make when predicting HI at a region level. It must be noted that Lorenz et al. (2010) had a more limited range of data (less than six studies) with many data coming from one country only (i.e. USA). This may explain some of the differences in conclusions.

5.2. Crop product N concentrations

Among all variables tested, FN had the largest explanatory and predictive power for CPN. This makes sense considering the availability of N (through FN) allows the plant to better express its genetic potential by accumulating N in the crop product. The positive relationship between FN and CPN in maize found in this study aligns with work by Dilz (1971) and is supported by numerous studies as summarized by Correndo et al. (2021). In contrast to expectations, FP and FK did not contribute significantly to the explanatory or predictive power for CPN. A larger CPY was weakly correlated with a higher CPN in the present study. This confirms findings by Tenorio et al. (2019) who showed that on balance more studies (at a ratio of 10:1) had a positive relationship between CPY and CPN. In our study however, there is a confounding effect of greater nutrient applications for the data points that had greater maize yield, or differences in management practices.

Our study showed that a larger Ypot leads to lower CPN at similar N application rates. This observation aligns well with recent findings from the long-term Broadbalk experiment as reported by van Grinsven et al. (2022) in their supplementary note 9, using attainable yield instead of potential yield.

5.3. Crop residue N concentration

Like for CPN, the availability of N will facilitate full genetic expression of the maize trait in terms of N in the residue component. However, the relationship between FN and CRN is confounded by the fact that maize has a propensity to shift N from the residue components to the crop products as it reaches physiological maturity (Ciampitti et al., 2013). Despite this potential confounding effect, FN was seen to be a relatively important predictor variable for the random forest and linear mixed effect models for CRN. This highlights the relative importance of collecting fertilizer application rate data at more localized levels to improve predictions of crop nutrient concentrations.

6. Conclusions

Reliable estimates of crop N uptake and offtake are critical in estimating N balances, N use efficiencies and potential losses. Nitrogen offtake is an important component of N balances, as it provides an indication of NUE and the quantity of N that must be replenished. Our study aimed to develop a general concept and approach for obtaining more locally relevant estimates of crop nutrient offtake to improve the accuracy of nutrient balances up to the global level. Crop HI, CPN and CRN are important variables in the estimation of nutrient balances. However, these variables are often given single and default estimates per country or region.

The present study highlights the potential for predicting HI (and thus CRY), CPN and CRN for maize with only a limited set of globally available variables, namely CPY, FN and Ypot. Predictions from both the mixed-effects model and random forest method were shown to provide reasonable levels of accuracy with the random forest method having the greater accuracy. Random forest can also handle larger datasets and can improve predictions over time with the addition of a wide range of variables, and as more data become available. However, prediction accuracy should not be the only metric used to decide what prediction model to use. An advantage of using the linear mixed-effects models for predictions is the fact that is easier to interpret. Unlike random forest, linear mixed-effects models can provide tangible equations which make explicit the positive or negative correlations between the predictor and dependent variables.

The methods used in this study were only applied to maize, but could equally be applied to other crops and other nutrients which we now plan to do, focusing first on macronutrients in maize, wheat, rice and soybean. Anyone interested in contributing data from any crop to the open database described in this study may contact the corresponding author of this article. This will enable significantly improved estimates of crop nutrient offtake, nutrient balances and nutrient use efficiency at national to global levels and thus provide a sounder basis for guiding more sustainable nutrient management.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests. Cameron Ludemann reports financial support was provided by International Fertilizer Association. Achim Dobermann reports a relationship with International Fertilizer Association that includes: consulting or advisory and employment.

Data Availability

The dataset used in this study is available under the Creative Commons Attribution 4.0 (CC-BY) International license and can be accessed via https://doi.org/10.5061/dryad.j3tx95xhc.

Acknowledgements

This work was supported by the International Fertilizer Association (IFA). The authors wish to acknowledge all those who have contributed raw field experiment data for maize included in this study. Those include Fernando Garcia (Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata), Rachel Fields and Timothy Lewis (AngloAmerican), Tek Saptoka (CIMMYT) and Maryam Rahimi Jahangirlou. The authors also wish to acknowledge those who helped converting data from the literature into a standardized format including, Joost Krooshof, Bert Rijk, Tobias Bader and Victoria Miles-Hildago.

Appendix A. List of requested variables of replicated field experiments

See Appendix Table A1 here.

Table A1
List of requested variables of replicated field experiments.

Required/ Optional	Variable
Required	Name of organization responsible for trial data
Required	Crop common name
Required	Year of sowing
Required	Location of trial (incl GPS coordinates and/or country/state/province and nearest town)
Required	Yield of crop products and/or yield of crop residues and/or harvest index
Required	Nutrient concentration of crop products and/or crop residues
Required	Fertilizer application rates
Optional	Rainfall and irrigation information
Optional	Sub-species and/or cultivar information
Optional	Sowing date of trial
Optional	Soil texture as name or as percentages of silt, sand and clay
Optional	Soil organic matter with name of soil test method used.
Optional	Soil pH with name of soil test method used
Optional	Any other applicable soil test results
Optional	Information about where these data have been published if applicable, such as name of lead author/year published/name of journal/ digital objective identifier or website of article.

Appendix B. Sources of data

See Appendix Table B1 here.

Table B1
Sources of data from the literature or from raw data requests used in this study.

#	Source	Country/IFA Region	Publication or organization from which data came from	Contact who shared raw data
1	Literature	Argentina/Latin America	Parco et al. (2020)	
2	Literature	Argentina/Latin America	Carciochi et al. (2019)	
3	Literature	Argentina/Latin America	Martínez-Cuesta et al. (2020)	
4	Literature	Australia/Oceania	Dang et al. (2021)	
5	Literature	Bangladesh/South Asia	Islam et al. (2018)	
6	Literature	Bangladesh/South Asia	Ferdous et al. (2020)	
7	Literature	Brazil/Latin America	Gavilanes et al. (2020)	
8	Literature	Brazil/Latin America	Ferreira et al. (2014)	
9	Literature	Brazil/Latin America	Fuentes et al. (2018)	
10	Literature	Bulgaria/Western and Central Europe	Nenova et al. (2019)	
11	Literature	Canada/North America	Sanders (2020)	
12	Literature	Canada/North America	Zhang et al. (1993)	
13	Literature	Canada/North America	Ma and Biswas (2015)	
14	Literature	Canada/North America	Gagnon et al. (2020)	
15	Literature	Chile/Latin America	Hirzel et al. (2020)	
16	Literature	China/East Asia	Su et al. (2020)	
17	Literature	China/East Asia	Cheng et al. (2020)	
18	Literature	China/East Asia	Chen et al. (2016)	
19	Literature	China/East Asia	Shi et al. (2021)	
20	Literature	China/East Asia	Li et al. (2021)	
21	Literature	China/East Asia	Yang et al. (2021)	
22	Literature	China/East Asia	Zhang et al. (2021b)	
23	Literature	China/East Asia	Ji et al. (2021)	
24	Literature	China/East Asia	Kang et al. (2021)	
25	Literature	China/East Asia	Zhang et al. (2020)	
26	Literature	China/East Asia	Jiang et al. (2018)	
27	Literature	China/East Asia	Cao et al. (2021)	
28	Literature	China/East Asia	Qiang et al. (2020)	
29	Literature	Côte d'Ivoire/Africa	Van Reuler and Janssen (1996)	
30	Literature	Egypt/Africa	Emam and Osman (2020)	

(continued on next page)

Table B1 (*continued*)

#	Source	Country/IFA Region	Publication or organization from which data came from	Contact who shared raw data
31	Literature	Egypt/Africa	Abbas et al. (2021)	
32	Literature	Egypt/Africa	Kandil et al. (2020)	
33	Literature	Egypt/Africa	El-Sayed et al. (2021)	
34	Literature	Egypt/Africa	Fekry Ali (2020)	
35	Literature	Ethiopia/Africa	Negash et al. (2021)	
36	Literature	Ethiopia/Africa	Tadesse and Kim (2014)	
37	Literature	Ethiopia/Africa	Tadesse and Sultan (2021)	
38	Literature	Ethiopia/Africa	Anbessa et al. (2022)	
39	Literature	Hungary/Western and Central Europe	Berecz and Debreczeni (2000)	
40	Literature	Hungary/Western and Central Europe	Horváth et al. (2021)	
41	Literature	Hungary/Western and Central Europe	Széles et al. (2019)	
42	Literature	India/South Asia	Ray et al. (2020)	
43	Literature	India/South Asia	Ghosh et al. (2020)	
44	Literature	India/South Asia	Madagoudra et al. (2021)	
45	Literature	India/South Asia	Thakur et al. (2020)	
46	Literature	India/South Asia	Singh et al. (2021)	
47	Literature	India/South Asia	Bhimireddy et al. (2018)	
48	Literature	India/South Asia	Sowmya et al. (2021)	
49	Literature	India/South Asia	Kuntoji et al. (2021)	
50	Literature	India/South Asia	Dawson and Maseeh (2021)	
51	Literature	India/South Asia	Singh et al. (2020)	
52	Literature	India/South Asia	Nsanzabaganwa et al. (2014)	
53	Literature	India/South Asia	Walia and Kumar (2021)	
54	Literature	India/South Asia	Fayaz et al. (2021)	
55	Literature	India/South Asia	Mishra and Monalisa (2020)	
56	Literature	India/South Asia	Bhattacharjee et al. (2020)	
57	Literature	India/South Asia	Sharma et al. (2021)	
58	Literature	Iran/West Asia	Alijani et al. (2021)	
59	Literature	Iran/West Asia	Khalafi et al. (2021)	
60	Literature	Iraq/West Asia	Bakr et al. (2020)	
61	Literature	Iraq/West Asia	Ramadhan (2021)	
62	Literature	Kenya/Africa	Pasley et al. (2019)	
63	Literature	Kenya/Africa	Njoroge (2019)	
64	Literature	Nigeria/Africa	Abdullahi and Bello (2020)	
65	Literature	Pakistan/South Asia	Ahmed et al. (2020)	
66	Literature	Pakistan/South Asia	Azeem et al. (2021)	
67	Literature	Pakistan/South Asia	Ilyas et al. (2021)	
68	Literature	Pakistan/South Asia	Mussarat et al. (2021)	
69	Literature	Pakistan/South Asia	Ullah et al. (2021)	
70	Literature	Romania/Western and Central Europe	Barșon et al. (2021)	
71	Literature	Russia/Eastern Europe and Central Asia	Bagrintseva and Ivashenenko (2021)	
72	Literature	Serbia/Western and Central Europe	Latkovic et al. (2020)	
73	Literature	Serbia/Western and Central Europe	Tamindžić et al. (2021)	
74	Literature	Serbia/Western and Central Europe	Simić et al. (2020)	
75	Literature	Thailand/East Asia	Huq (1987)	
76	Literature	Thailand/East Asia	Feil et al. (2005)	
77	Literature	Turkey/West Asia	Ortas (2018)	
78	Literature	Turkey/West Asia	Ibrikci et al. (1998)	
79	Literature	Uganda/Africa	Kaizzi et al. (2012)	
80	Literature	United States/North America	Jung et al. (1972)	
81	Literature	United States/North America	Da Cunha Leme Filho et al. (2020)	
82	Literature	United States/North America	Lutz Jr et al. (1974)	
83	Literature	United States/North America	Miao et al. (2006)	
84	Literature	United States/North America	Mueller et al. (2019)	
85	Literature	United States/North America	Tsai et al. (1992)	
86	Literature	United States/North America	Perry and Olson (1975)	
87	Literature	United States/North America	Woli et al. (2018)	
88	Literature	United States/North America	Adeyemi et al. (2020)	
89	Literature	United States/North America	Cannon et al. (2021)	
90	Literature	Venezuela/Latin America	Barrios and Basso (2018)	
91	Literature	Zimbabwe/Africa	Manzeke-Kangara et al. (2021)	
92	Data request	Argentina/Latin America	Salvagiotti et al. (2017)	Fernando Garcia (Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata)
93	Data request	Brazil, China, India, USA (Latin America, East Asia, South Asia, North America).	AngloAmerican	Rachel Fields/Timothy Lewis (AngloAmerican)

(*continued on next page*)

Table B1 (*continued*)

#	Source	Country/IFA Region	Publication or organization from which data came from	Contact who shared raw data
94	Data request	USA, Vietnam, China, Ethiopia, Nigeria, India, Philippines, Indonesia (North America, East Asia, East Asia, Africa, Africa, South Asia, East Asia, East Asia).	International Fertilizer Association Wortmann et al. (2009) Setiyono et al. (2010)	Achim Dobermann (International Fertilizer Association)
95	Data request	Iran/West Asia	Jahangirlou et al. (2021)	Maryam Rahimi Jahangirlou
96	Data request	Mexico/Latin America	NA	Tek Saptoka (CIMMYT)

Appendix C. Random forest predictions for yield potential

The methodology for predicting yield potential for some locations not already included in the Global Yield Gap Atlas (GYGA, https://www.yieldgap.org/) were as follows:

Selection countries and climate zones (CZs).

1. Identify countries to be included. Threshold of > 50,000 ha national harvested area (using SPAM map) for all cereals and water-regime combinations separately.

2. Identify CZs to be included from the selected countries for each of the cereal and water-regime combinations. First, select CZs with > 1% national harvested area. Rank those CZs in the country by harvested area and select CZs until 50% of the national harvested area is reached.

Variables to explain variability in non-water limited yield potential (Yp), and water limited yield potential (Yw) .

- Climate zone variables (classes) obtained from GYGA
  - o Growing degree days (10 classes), Aridity index (10 classes), Temperature seasonality (3 classes)
- Latitude. The absolute latitude of the centroid of the climate zone was included as a measure of radiation intercepted. In case the climate zone is present in multiple disjunct locations across the same country the average value of the centroids was taken.
- Total growth duration (https://ipad.fas.usda.gov/ogamaps/cropcalendar.aspx).
- Additional variables to explain variability in Yw
  - o Total seasonal precipitation (NASA-power).
  - o Soil variables of dominant soil type in that CZ (ISRIC wise database, https://data.isric.org/geonetwork/srv/eng/catalog.search#/metadata/d9eca770-29a4- 4d95-bf93-f32e1ab419c3)
  - o Total available water capacity for the default soil depth of 100 cm for deep soil units, to 10 cm for shallow Lithosols, and to 30 cm for Rankers and Rendzinas (TAWC, 9 classes)

Regression.

We used a random forest regression model. Analysis was done in R using the randomForest function and for cross-validation the rf.crossValidation function (from the randomForest and the rfUtilities package, respectively) was used. For the random forest regression and the cross-validation the proportion of data withheld was 0.10 (default value), 99 cross validations (default value), and the number of trees to grow was set to 500.

Additional information for estimation of Yp.

In GYGA for rainfed crops when estimating Yw, at the same time Yp is also estimated (for the growing season and cultivar of rainfed crops). Normally this is only used for the purpose of obtaining the degree of water limitation for that specific climate zone. Here, we made use of this estimated Yp. For the present study it means that we included some CZs with irrigated cereals, while they do not feature in GYGA as their areas were considered too small (instead only rainfed cereals were considered).

Appendix D. Standard deviations of important variables

See Appendix Table D1 here.

Table D1
Standard deviation values of important variables in dataset analyzed in this study* .

Region* *	Mean values for each variable* **								
	FN (kg ha^{-1})	FP (kg ha^{-1})	FK (kg ha^{-1})	CPY (Mg DM ha^{-1})	CRY (Mg DM ha^{-1})	Ypot (Mg DM ha^{-1})	HI (-)	CPN (%)	CRN (%)
Africa	113	27	50	1.76	2.89	4.37	0.08	0.36	0.37
East Asia	108	52	65	2.91	2.57	1.34	0.09	0.38	0.28
Eastern Europe & Central Asia	30	0	0	0.85	NA	NA	NA	NA	NA
Latin America	77	17	25	2.49	2.52	0.65	0.06	0.22	0.14
North America	102	17	44	2.87	2.52	1.07	0.07	0.21	0.23
Oceania	82	0	0	2.14	1.75	NA	0.07	0.13	0.14
South Asia	81	24	49	2.61	3.18	0.68	0.09	0.36	0.36
West Asia	130	43	52	5.6	2.09	0.86	0.11	0.24	0.01
Western and Central Europe	76	27	73	2.7	NA	0	NA	0.27	NA

*The standard deviation values are representative across all treatments in the data provided and may not represent the standard deviation values for the typical fertilizer application rates in each region. * *Region based on IFA (2021). * **where: CPY=crop product yield (megagrams dry matter (Mg DM) ha^{-1}), CRY=crop residue yield (Mg DM ha^{-1}), HI=harvest index (product yield as a proportion of above ground biomass), FN=fertilizer nitrogen (N) applied (kg N ha^{-1}), FP=fertilizer phosphorus applied (kg elemental P ha^{-1}), FK=fertilizer potassium applied (kg elemental K ha^{-1}), CPN=crop product N concentration % (100 × kg N kg^{-1} DM), CRN=crop residue N concentration % (100 × kg N kg^{-1} DM),Ypot= yield potential from Global Yield Gap Atlas (Mg DM crop product ha^{-1}). * ** *Unique trial locations (based on GPS coordinates) were used as a proxy for estimating the number of trials per region.

Appendix E. Correlation matrix

See Appendix Fig. E1 here.

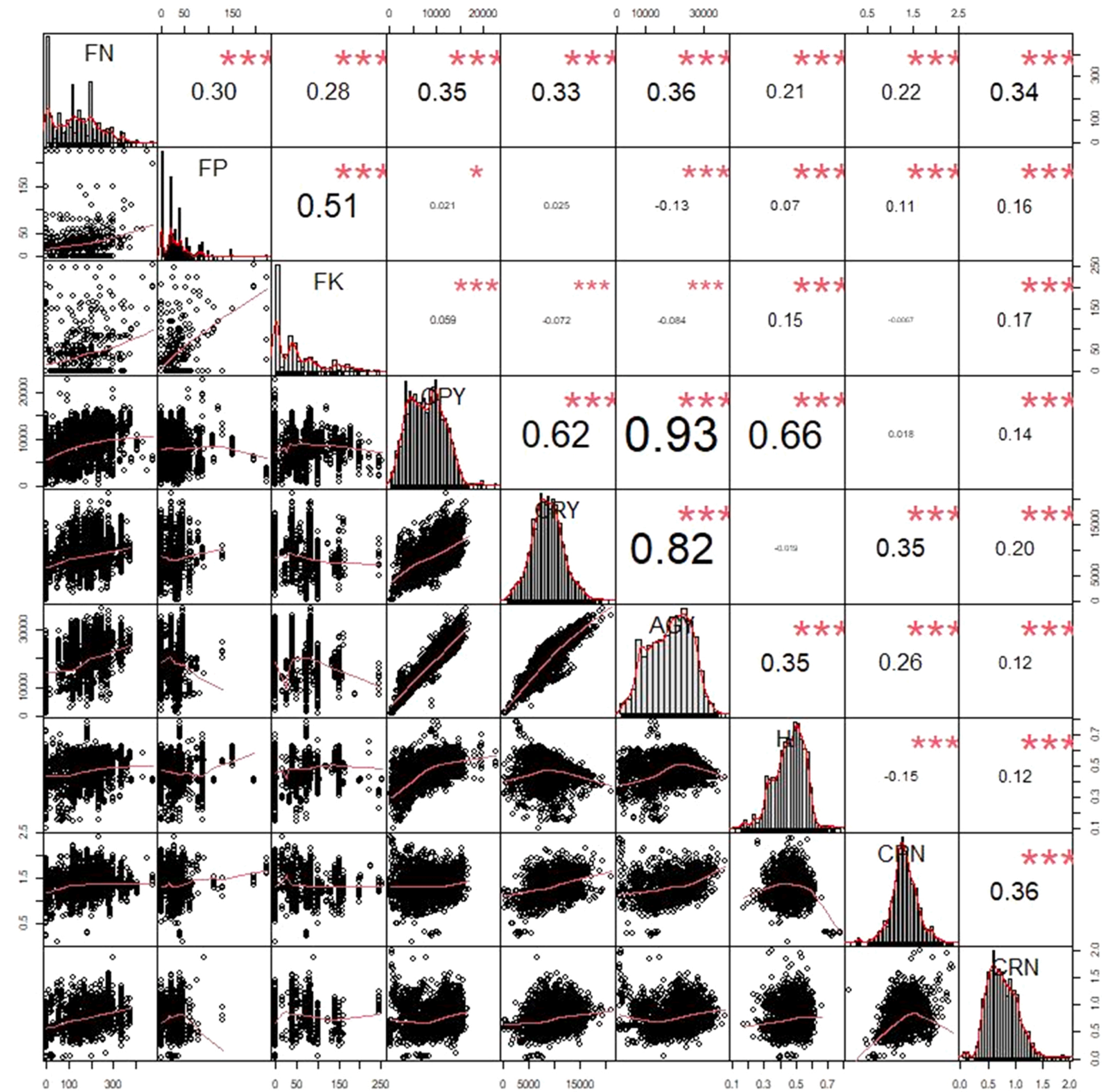

Fig. E1. Correlation matrix for selected variables where variable definitions are included in Table 1. On the diagonal, this figure shows the distribution of each variable as histograms, below the diagonal are the bivariate scatter plots with fitted (density) lines in red displayed. Above the diagonal are the correlation values with the significance level displayed with the following symbols denoting p-values of between: 0–0.001 = ** *, 0.001–0.01 = ** , 0.01–0.05 = *, 0.05–0.1 = ".", and 0.1–1 = " ".

Appendix F. Coefficients used in the final selected models whereby Crop Product Yield (CPY) is in units of megagrams (Mg) dry matter (DM) per hectare, Harvest Index (HI) is a proportion, CPY as a proportion of yield potential (relative yield-RY) is in Mg DM per hectare, fertilizer nitrogen application rates (FN) are in kg nitrogen per square meter, and fertilizer phosphorus application rates (FP) are in kg elemental phosphorus per square meter

See Table F1 here.

Table F1
Coefficients used in the prediction equation (a and b) for harvest index (HI) whereby the equation had CPY log-transformed i.e. HI = a × log(CPY) +b.

Region	a	b
Africa	0.09	0.25
East Asia	0.04	0.41
Eastern Europe and Central Asia	NA	NA
Latin America	0.07	0.34
North America	0.10	0.26
Oceania	0.13	0.16
South Asia	0.11	0.22
West Asia	0.13	0.15
Western and Central Europe	NA	NA
Across all regions*	0.07	0.33

*These coefficients were used for the prediction equation when regional variation was not accounted for.

Appendix G. Regression for actual and predicted harvest index (HI) of maize using the model where CPY was log-transformed

See Appendix Fig. G1 here.

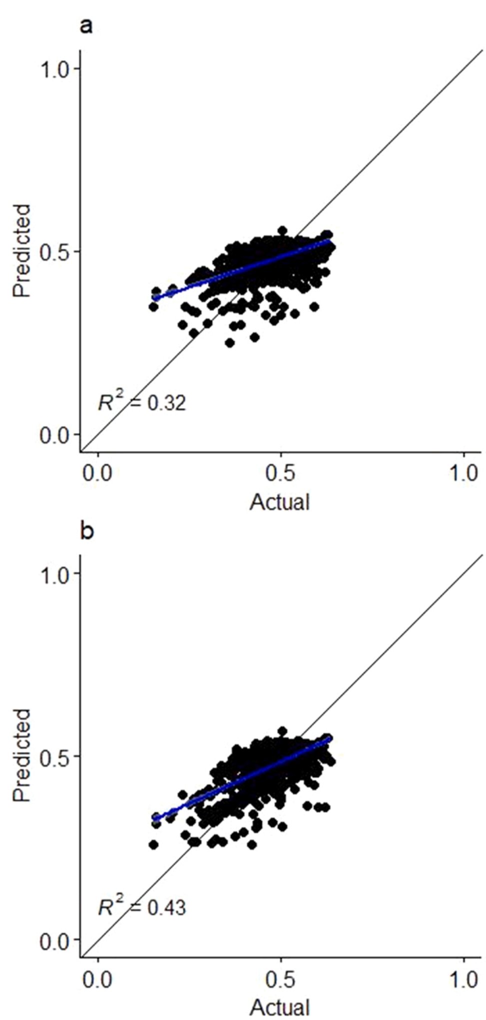

Fig. G1. Regression for actual and predicted harvest index (HI) of maize where CPY was log-transformed using the HI = a × log(CPY) +b model (log_1). Plot G.1a is shown for the regression model where regional effect was not accounted for, and plot G.1b is shown for the regression model where regional effect was accounted for.

Appendix H. Regression for actual and predicted harvest index (HI) of maize using linear mixed-effects models

See Appendix Fig. H1 here.

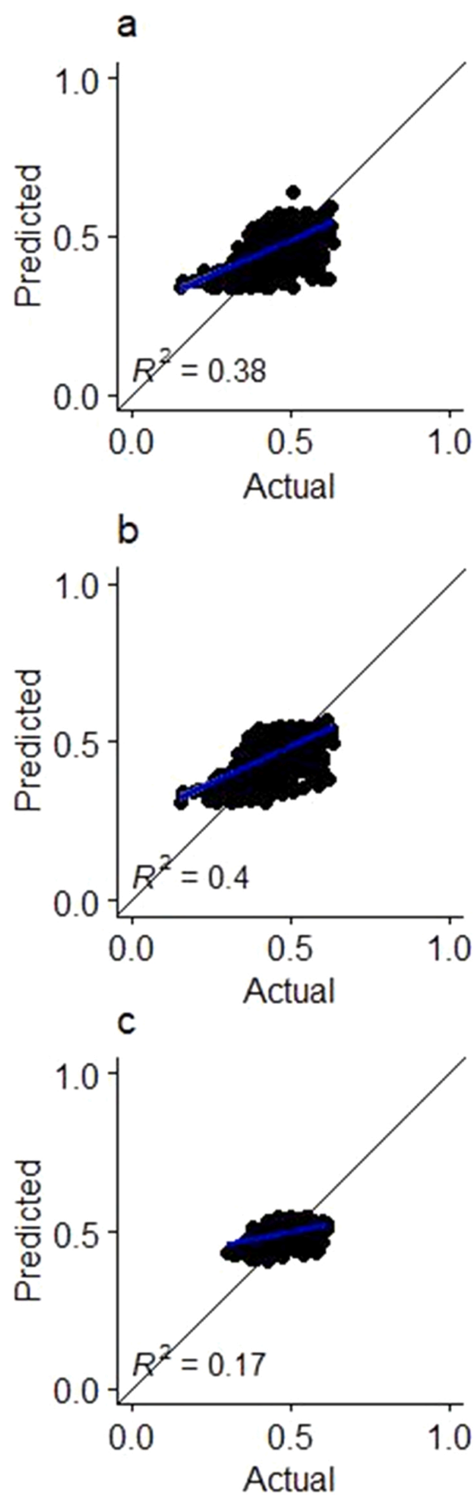

Fig. H1. Regression for actual and predicted harvest index (HI) of maize using the H1_1 (plot H.1a), H1_2 (plot H.1b), H1_10 (plot H.1c) linear mixed-effects models.

Appendix I. Regression for actual and predicted crop product nitrogen (N) concentration (CPN) of maize using the linear mixed-effects models

See Apendix Fig. I1 here.

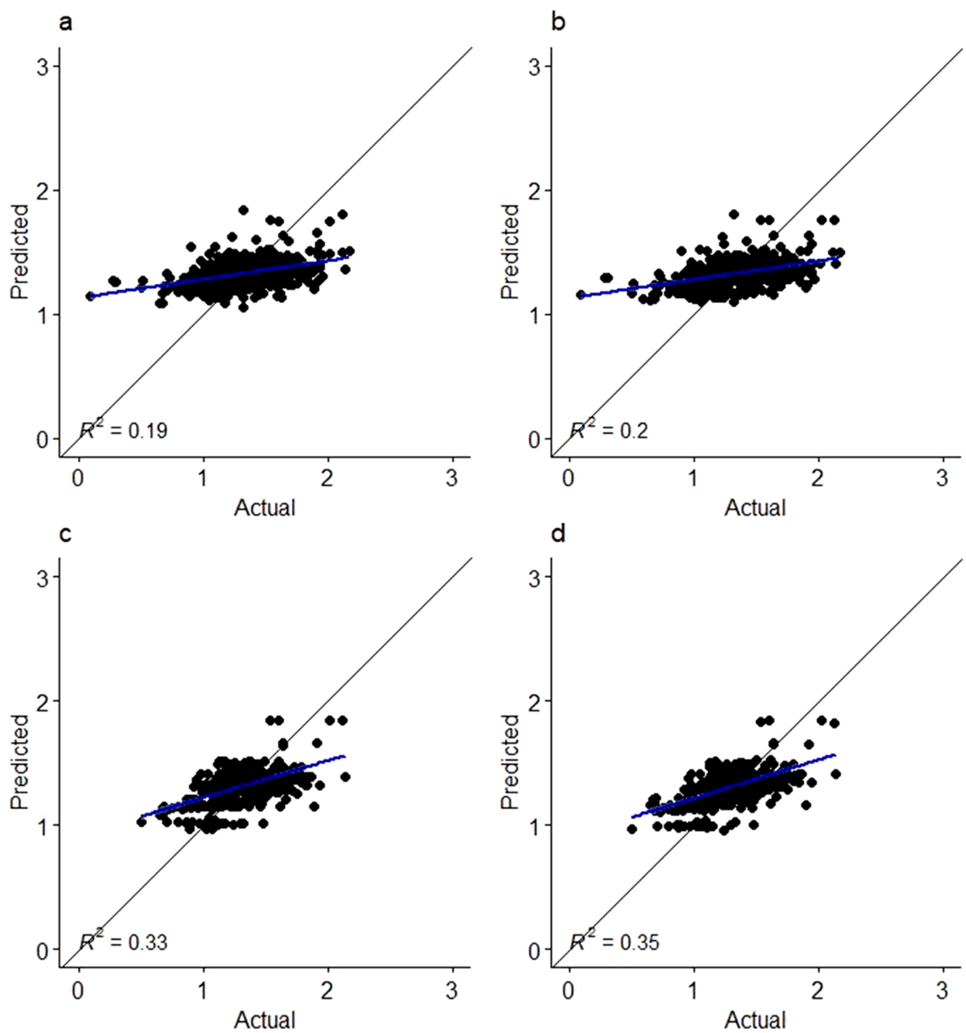

Fig. I1. Regression for actual and predicted crop product nitrogen (N) concentration (CPN) of maize using the H2_3 (plot I.1a), H2_6 (plot I.1b), H2_10 (plot I.1c), and H2_14 (plot I.1d) linear mixed-effects models.

Appendix J. Regression for actual and predicted crop residue nitrogen (N) concentration (CRN) of maize using the linear mixed-effects models

See Appendix Fig. J1 here.

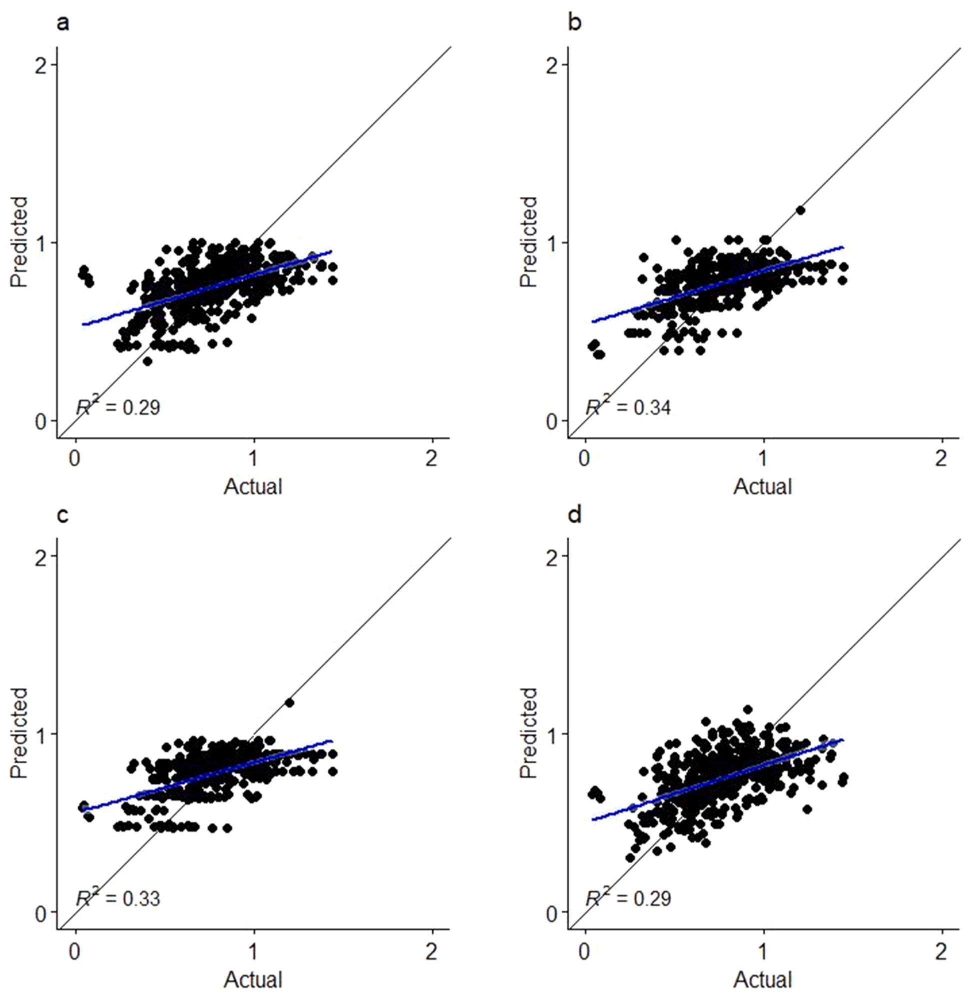

Fig. J1. Regression for actual and predicted crop residue nitrogen (N) concentration (CRN) of maize using the H3_3 (plot J.1a), H3_7 (plot J.1b), H3_11 (plot J.1c) and H3_16 (plot J.1d) linear mixed-effects models.

Appendix K. Regression for actual and predicted harvest index (HI), crop product nitrogen (N) concentration (CPN) and crop residue nitrogen (N) concentration (CRN) of maize using simple regression models

See Appendix Table K1 here.

Table K1
Prediction accuracy of simple linear regression models.

Model name	Model equation*	R^2 of predicted versus actual values
log_1 *	HI ~ a × log(CPY) +b	0.36
H1_1	HI ~ CPY	0.37
H1_2 *	HI ~ CPY + CPY^2	0.39
H1_10	HI ~ Ypot + CPY	0.28
H2_3	CPN ~ CPY + FN + FP	0.09
H2_6	CPN ~ FN + FP	0.07
H2_10	CPN ~ Ypot + FN	0.22
H2_14	CPN ~ RY+FN	0.29
H3_3	CRN ~ CPY + FN + FP	0.15
H3_7	CRN ~ Ypot + FN	0.21
H3_11	CRN ~ RY + FN	0.29
H3_16	CRN ~ CPN + CPY + FN +FP+FK	0.19

*Where CPY=crop product yield, Ypot = yield potential, FN=fertilizer nitrogen application, FP= fertilizer phosphorus application, and RY is relative yield expressed as CPY/Ypot.

References

Abbas, M., Abdel-Lattif, H., Shahba, M., 2021. Ameliorative effects of calcium sprays on yield and grain nutritional composition of maize (Zea mays L.) cultivars under drought stress. Agriculture 11, 285.

Abdullahi, M., Bello, A., 2020. Effects of different levels of phosphorous and zinc fertilizers on the yield and nutrient uptake of maize (Zea mays L.) on luvisols in Northern Guinea Savannah region of Nigeria. Asian J. Soil Sci. Plant Nutr. 6, 1–7.

Adeyemi, O., Keshavarz-Afshar, R., Jahanzad, E., Battaglia, M.L., Luo, Y., Sadeghpour, A., 2020. Effect of wheat cover crop and split nitrogen application on corn yield and nitrogen use efficiency. Agronomy 10, 1–10.

Ahmed, N., Habib, U., Younis, U., Irshad, I., Danish, S., Rahi, A.A., Munir, T.M., 2020. Growth, chlorophyll content and productivity responses of maize to magnesium sulphate application in calcareous soil. Open Agric. 5, 792–800.

Akaike, H. (Ed.), 1998. Information Theory and an Extension of the Maximum Likelihood Principle, 267–. Springer, New York, USA, p. 281.

Alijani, K., Bahrani, M.J., Kazemeini, S.A., Yasrebi, J., 2021. Soil and sweet corn quality responses to tillage, residue, and nitrogen management in Southern Iran. Int. J. Plant Prod.

Anbessa, B., Abera, G., Kassas, S., 2022. Effect of blended fertilizer on yield, nutrient uptake and economy of maize (Zea mays L.) in Assosa District, Western Ethiopia. Asian J. Plant Sci. 21, 130–138.

Azeem, K., Naz, F., Jalal, A., Galindo, F., Teixeira Filho, M., Khalil, F., 2021. Humic acid and nitrogen dose application in corn crop under alkaline soil conditions. Rev. Bras. De. Eng. Agríc. e Ambient. 25, 657.

Bagrintseva, V.N., Ivashenenko, I.N., 2021. Evaluating the response of maize hybrids to nitrogen fertilizer. Russ. Agric. Sci. 47, 17–20.

Bakr, R., Ahmed, G., Mam, Rasul, G., 2020. Influence of foliar application of some micronutrients on maize kernel yield and nutrient content. J. Macromol. Sci. Part A Pure Appl. Chem. 22, 78–88.

Barrios, M., Basso, C., 2018. Efecto de la fertilización nitrogenada sobre componenntes del renidimento y calidad nutricional del grnao de seis híbridos de maíz. Bioagro 30, 39–48.

Barșon, G., Șopterean, L., Suciu, L.A., Crișan, I., Duda, M.M., 2021. Evaluation of agronomic performance of maize (Zea mays L.) under a fertilization gradient in Transylvanian plain. Agriculture 11.

Berecz, K., Debreczeni, K., 2000. Effect of nutrient supply and site characteristics on NPK yield of different maize hybrids. Acta Agron. Hung. 48, 51–63.

Bhattacharjee, S., Bhale, V.M., Kumar, P., Kumar, S., Kumar, D., Bairwa, B., 2020. Exceeding potassium and zinc application rate above farmers practiced rate improves dry matter partitioning, photosynthetic attributes, yield and quality of maize (Zea mays L.) under dryland condition. Curr. J. Appl. Sci. Technol. 39, 35–44.

P. Bhimireddy, P., Mallareddy, M., Subbaiah, G., Sekhar, K.C., Reddy, D.V.V., Babu, P.R., 2018. Perform. no- maize drip-Fert. a Double Crop. Syst. semi Arid Telangana State India, p. *62*.

Brown, V.A., 2021. An introduction to linear mixed-effects modeling in R. Adv. Methods Pract. Psychol. Sci. 4, 2515245920960351.

Burnham, K.P., Anderson, D.R., 2002. In: New York, U.S.A. (Ed.), Model Selection and Multimodel Inference, second ed. Springer, p. 512.

Cannon, K., McClure, M., Yin, X., Sams, C., 2021. Corn and soybean response to sulfur fertilizer in West Tennessee. *Crop*. Forage Turfgrass Manag., e20092

Cao, D., Lan, Y., Chen, W., Yang, X., Wang, D., Ge, S., Yang, J., Wang, Q., 2021. Successive applications of fertilizers blended with biochar in the soil improve the availability of phosphorus and productivity of maize (Zea mays L.). Eur. J. Agron. 130.

Carciochi, W.D., Calvo, N.I.R., Wyngaard, N., Divito, G.A., Eyherabide, M., Echeverría, H.E., 2019. Prognosis and diagnosis of sulfur status in maize by plant analysis. Eur. J. Agron. 108, 1–10.

Cassman, K.G., Dobermann, A., Walters, D.T., 2002. Agroecosystems, nitrogen-use efficiency, and nitrogen management. AMBIO: A J. Hum. Environ. 31 (132–140), 139.

Chen, Q., Mu, X., Chen, F., Yuan, L., Mi, G., 2016. Dynamic change of mineral nutrient content in different plant organs during the grain filling stage in maize grown under contrasting nitrogen supply. Eur. J. Agron. 80, 137–153.

Cheng, Y., Wang, H.-Q., Liu, P., Dong, S.-T., Zhang, J.-W., Zhao, B., Ren, B.-Z., 2020. Nitrogen placement at sowing affects root growth, grain yield formation, N use efficiency in maize. Plant Soil 1–19.

Ciampitti, I., Vyn, T., 2013. Maize nutrient accumulation and partitioning in response to plant density and nitrogen rate: II. Calcium, magnesium, and micronutrients. Agron. J. 105, 1645.

Ciampitti, I.A., Camberato, J.J., Murrell, S.T., Vyn, T.J., 2013. Maize nutrient accumulation and partitioning in response to plant density and nitrogen rate: I. Macronutrients. Agron. J. 105, 783–795.

Conant, R.T., Berdanier, A.B., Grace, P.R., 2013. Patterns and trends in nitrogen use and nitrogen recovery efficiency in world agriculture. Glob. Biogeochem. Cycles 27, 558–566.

Correndo, A.A., Fernandez, J.A., Vara Prasad, P.V., Ciampitti, I.A., 2021. Do water and nitrogen management practices impact grain quality in maize? Agronomy 11, 1851.

Da Cunha Leme Filho, J.F., Ortiz, B.V., Balkcom, K.S., Damianidis, D., Knappenberger, T. J., Dougherty, M., 2020. Evaluation of two irrigation scheduling methods and nitrogen rates on corn production in Alabama. Int. J. Agron.

Dang, Y.P., Martinez, C., Smith, D., Rowlings, D., Grace, P., Bell, M., 2021. Maize production and nitrous oxide emissions from enhanced efficiency nitrogen fertilizers. Nutr. Cycl. Agroecosyst.

Dawson, J., Maseeh, Y., 2021. Effect of nitrogen and sulphur levels on growth and yield of maize (Zea mays L.). Pharm. Innov. 10, 241–245.

Dilz, K., 1971. Effect of Chlormequat on the Growth of Cereals. 2. 1965–1966 trials. Effects of Time of Chlormequat Application, Level of Nitrogen, and Split Application of Nitrogen on Resistance to Lodging and on Yield of Winter Wheat. Den Haag, The Netherlands. ⟨https://agris.fao.org/agris-search/search.do?recordID=US201302310402⟩.

El-Sayed, E.A., El-Sobky, Abdo, A.I., 2021. Efficacy of using biochar, phosphorous and nitrogen fertilizers for improving maize yield and nitrogen use efficiencies under alkali clay soil. J. Plant Nutr. 44, 467–485.

Emam, S.M., Osman, E.A., 2020. Integrated application of organic, bio and mineral fertilizers on nutrients uptake and productivity of Zea mays L. under semi-arid condition. J. Plant Nutr. 1–13.

Evans, L.T., 1998. Feeding the Ten Billion: Plants and Population Growth. Cambridge University Press, Cambridge, UK.

FAOSTAT. 2020. FAOSTAT. Retrieved 18 May 2020 from: ⟨http://www.fao.org/faostat/en/#country⟩.

Fayaz, S., Kanth, R., Ahmad, T., Anwar, M., Alie, B., Ahmed, Z., Mehdi, Z., Maqbool, D.S., Rasool, K., Cortes, J., Santiago, 2021. Effect of Precision Nitrogen Management through LCC on Nutrient Content and Uptake of Maize (Zea mays L.) under Temperate Conditions of Kashmir. Int. J. Plant Soil Sci. 33.

Feil, B., Moser, S., Jampatong, S., Stamp, P., 2005. Mineral composition of the grains of tropical maize varieties as affected by pre-anthesis drought and rate of nitrogen fertilization. Crop Sci. 45, 516–523.

Fekry Ali, M.E., 2020. Response of maize plants to different nitrogen rates and zinc application methods. Plant Arch. 20, 9144–9150.

Ferdous, Z., Ullah, H., Datta, A., Attia, A., Rakshit, A., Hossain, Molla, S., 2020. Application of biogas slurry in combination with chemical fertilizer enhances grain yield and profitability of maize (Zea mays L.). Commun. Soil Sci. Plant Anal. 1–10.

Ferreira, C.F., Motta, A.C., Barbosa, J.Z., dos Santos, N.Z., Prior, S., Gabardo, J., 2014. Maize (Zea mays L) cultivars nutrients concentration in leaves and stalks. Maydica 59, 7.

Fuentes, L.F.G., Souza, L.C.F.D., Serra, A.P., Rech, J., Vitorino, A.C.T., 2018. Corn agronomic traits and recovery of nitrogen from fertilizer during crop season and off-season. Pesqui. Agropecuária Bras. 53, 1158–1166.

Gagnon, B., Ziadi, N., Bélanger, G., Parent, G., 2020. Validation and use of critical phosphorus concentration in maize. Eur. J. Agron. 120, 126147.

Gavilanes, F.Z., Andrade, D.S., Zucareli, C., Horácio, E.H., Yunes, J.S., Barbosa, A.P., Alves, L.A.R., Cruzatty, L.G., Maddela, N.R., de Fátima Guimarães, M., 2020. Co-inoculation of Anabaena cylindrica with Azospirillum brasilense increases grain yield of maize hybrids. Rhizosphere 15, 100224.

Ghosh, D., Brahmachari, K., Brestic, M., Ondrisik, P., Hossain, A., Skalicky, M., Sarkar, S., Moulick, D., Dinda, N.K., Das, A., 2020. Integrated weed and nutrient management improve yield, nutrient uptake and economics of maize in the rice-maize cropping system of Eastern India. Agronomy 10, 1906.

Harrison, X.A., Donaldson, L., Correa-Cano, M.E., Evans, J., Fisher, D.N., Goodwin, C.E. D., Robinson, B.S., Hodgson, D.J., Inger, R., 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ 6 e4794-e4794.

Hay, R.K.M., Gilbert, R.A., 2001. Variation in the harvest index of tropical maize. Ann. Appl. Biol. 138, 103–109.

Heffer, P., Gruere, A., Roberts, T., 2017. Assessment of Fertiliser use by Crop at the Global Level. International Fertilizer Association, Institute IPN, Paris, France, p. 19. ⟨https://www.ifastat.org/plant-nutrition⟩.
Heinze, G., Wallisch, C., Dunkler, D., 2018. Variable selection – a review and recommendations for the practicing statistician. Biometr. J. 60, 431–449.
Hirzel, J., Undurraga, P., León, L., Panichini, M., Carrasco, J., González, J., Matus, I., 2020. Maize grain production, plant nutrient concentration and soil chemical properties in response to different residue levels from two previous crops. Acta Agric. Scand. Sect. B—Soil Plant Sci. 70, 285–293.
Horváth, É., András, T., Fejér, P., Széles, A., 2021. Effect of different N doses on maize yield and quality. Acta Agrar. Debr. 97–101.
Huq, S., 1987. Yield and protein quality of maize grain as affected by fertilizer application. Bangladesh J. Agric. 12, 169–179.
Ibrikci, H., Ulger, A.C., Cakir, B., Buyuk, G., Guzel, N., 1998. Modeling approach to nitrogen uptake by field-grown corn. J. Plant Nutr. 21, 1943–1954.
IFA. 2021. *IFA regional classification of countries- IFASTAT*. Retrieved 29 October 2021 from: ⟨https://api.ifastat.org/reports/download/12457⟩.
Ilyas, M., Arif, M., Akhtar, K., Riaz, M., Wang, H., 2021. Diverse feedstock's biochars as supplementary K fertilizer improves maize productivity, soil organic C and KUE under semiarid climate. Soil Tillage Res. 211, 105015.
Islam, S., Timsina, J., Salim, M., Majumdar, K., Gathala, M.K., 2018. Potassium supplying capacity of diverse soils and K-use efficiency of maize in South Asia. Agronomy 8, 121.
Jahangirlou, M.R., Akbari, G.A., Alahdadi, I., Soufizadeh, S., Parsons, D., 2021. Grain quality of maize cultivars as a function of planting dates, irrigation and nitrogen stress: a case study from semiarid conditions of Iran. Agriculture 11, 11.
Janssen, B.H., Guiking, F.C.T., van der Eijk, D., Smaling, E.M.A., Wolf, J., van Reuler, H., 1990. A system for quantitative evaluation of the fertility of tropical soils (QUEFTS). Geoderma 46, 299–318.
Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.-M., Gerber, J.S., Reddy, V.R., Kim, S.-H., 2016. Random forests for global and regional crop yield predictions. PLoS One 11, e0156571.
Ji, P.T., Li, X.L., Peng, Y.J., Zhang, Y.C., Tao, P.J., 2021. Effect of polyaspartic acid and different dosages of controlled-release fertilizers on nitrogen uptake, utilization, and yield of maize cultivars. Bioengineered 12, 527–539.
Jiang, W., Liu, X., Wang, Y., Zhang, Y., Qi, W., 2018. Responses to potassium application and economic optimum K rate of maize under different soil indigenous K supply. Sustainability 10.
Jiang, Z., Yang, S., Chen, X., Pang, Q., Xu, Y., Qi, S., Yu, W., Dai, H., 2021. Controlled release urea improves rice production and reduces environmental pollution: a research based on meta-analysis and machine learning. Environ. Sci. Pollut. Res.
Jung, J.P.E., Peterson, L.A., Schrader, L.E., 1972. Response of irrigated corn to time, rate, and source of applied N on sandy soils1. Agron. J. 64, 668–670.
Kaizzi, K.C., Byalebeka, J., Semalulu, O., Alou, I., Zimwanguyizza, W., Nansamba, A., Musinguzi, P., Ebanyat, P., Hyuha, T., Wortmann, C.S., 2012. Maize response to fertilizer and nitrogen use efficiency in Uganda. Agron. J. 104, 73–82.
Kandil, E.E., Abdelsalam, N.R., Mansour, M.A., Ali, H.M., Siddiqui, M.H., 2020. Potentials of organic manure and potassium forms on maize (Zea mays L.) growth and production. Sci. Rep. 10, 1–11.
Kang, L., Zhang, G., Chu, G., 2021. Split delivering phosphorus via fertigation to a calcareous soil increased P availability and maize yield (Zea mays L.) by reducing P fixation. J. Soils Sediment. 21, 2287–2300.
Keulen, H.V., Heemst, H.D.J.V., 1982. Crop Response to the Supply of Macronutrients. Pudoc, Centre for Agricultural Publishing and Documentation, Wageningen.
Khalafi, A., Mohsenifar, K., Gholami, A., Barzegari, M., 2021. Corn (Zea mays L.) Growth, Yield and Nutritional Properties Affected by Fertilization Methods and Micronutrient Use. Int. J. Plant Prod.
Kosmowski, F., Chamberlin, J., Ayalew, H., Sida, T., Abay, K., Craufurd, P., 2021. How accurate are yield estimates from crop cuts? Evidence from smallholder maize farms in Ethiopia. Food Policy 102, 102122.
Kuntoji, A., Subbarayappa, C.T., Sathish, A., Ramamurthy, V., Mallesha, B.C., 2021. Effect of different levels of nitrogen and zinc application on growth, yield and quality of maize in rural and peri-urban gradients of southern transect of Bengaluru. J. Pharmacogn. Phytochem. 10, 1562–1568.
Lal, R., 2005. World crop residues production and implications of its use as a biofuel. Environ. Int. 31, 575–584.
Lassaletta, L., Billen, G., Grizzetti, B., Juliette, A., Garnier, J., 2014. 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. Environ. Res. Lett. 105011, 105011.
Latkovic, D., Maksimovic, J., Dinic, Z., Pivic, R., Stanojkovic, A., Stanojkovic-Sebic, A., 2020. Case study upon foliar application of biofertilizers affecting microbial biomass and enzyme activity in soil and yield related properties of maize and wheat grains. Biology 9, 452.
Li, G.-H., Cheng, Q., Li, L., Lu, D.-L., Lu, W.-P., 2021. N, P and K use efficiency and maize yield responses to fertilization modes and densities. J. Integr. Agric. 20, 78–86.
Linquist, B.A., Adviento-Borbe, M.A., Pittelkow, C.M., van Kessel, C., van Groenigen, K. J., 2012. Fertilizer management practices and greenhouse gas emissions from rice systems: a quantitative review and analysis. Field Crop. Res. 135, 10–21.
Liu, L., Sadras, V.O., Xu, J., Hu, C., Yang, X., Zhang, S., 2021. Chapter five - genetic improvement of crop yield, grain protein and nitrogen use efficiency of wheat, rice and maize in China. In: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 203–252.
Lorenz, A.J., Gustafson, T.J., Coors, J.G., Leon, N., 2010. Breeding maize for a bioeconomy: a literature survey examining harvest index and stover yield and their relationship to grain yield. Crop Sci. 50, 1–12.
Lüdecke D., Makowski D., Waggoner P., Patil I. 2020. *performance: Assessment of Regression Models Performance.* Retrieved 7 May 2021 from: ⟨https://cran.r-project.org/web/packages/performance/index.html⟩.
Ludemann, C.I., Gruere, A., Heffer, P., Dobermann, A., 2022. Global data on fertilizer use by crop and by country. DRYAD. https://doi.org/10.5061/dryad.2rbnzs7qh.
Lutz Jr, J., Jones, G., Hale, E., 1974. Effects of irrigation and fertilization on yield and potassium content of corn and on available potassium contents of a Davidson soil. Commun. Soil Sci. Plant Anal. 5, 67–79.
Ma, B.L., Biswas, D.K., 2015. Field-level comparison of nitrogen rates and application methods on maize yield, grain quality and nitrogen use efficiency in a humid environment. J. Plant Nutr. 39, 727–741.
Madagoudra, Y., Chaithanya, J., Terin, P., R L, 2021. Effect of rate and time of nitrogen application on nutrient content and uptake in summer maize (Zea mays L.) and soil nutrient status after harvest. Int. J. Econ. Plants 8, 57–61.
Madden, L.V., Piepho, H.-P., Paul, P.A., 2016. Statistical models and methods for network meta-analysis. Phytopathology® 106, 792–806.
Manzeke-Kangara, M.G., Mtambanengwe, F., Watts, M.J., Broadley, M.R., Lark, R.M., Mapfumo, P., 2021. Can nitrogen fertilizer management improve grain iron concentration of agro-biofortified crops in Zimbabwe? Agronomy 11, 124.
Martínez-Cuesta, N., Carciochi, W., Sainz-Rozas, H., Salvagiotti, F., Colazo, J.C., Wyngaard, N., Eyherabide, M., Ferraris, G., Barbieri, P., 2020. Effect of zinc application strategies on maize grain yield and zinc concentration in mollisols. J. Plant Nutr. 1–12.
Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., Bates, D., 2017. Balancing Type I error and power in linear mixed models. J. Mem. Lang. 94, 305–315.
McKevith, B., 2004. Nutritional aspects of cereals. Nutr. Bull. 29, 111–142.
Miao, Y., Mulla, D.J., Robert, P.C., Hernandez, J.A., 2006. Within-field variation in corn yield and grain quality responses to nitrogen fertilization and hybrid selection. Agron. J. 98, 129–140.
Mishra, S., Monalisa, S., 2020. Physiological & biochemical response of maize (Zea mays L.) to low grade rock phosphate under acidic soils of Odisha. Int. J. Chem. Stud. 8, 2030–2035.
Mueller, S.M., Messina, C.D., Vyn, T.J., 2019. Simultaneous gains in grain yield and nitrogen efficiency over 70 years of maize genetic improvement. Sci. Rep. 9, 9095.
Mussarat, M., Ali, H., Muhammad, D., Ahmad Mian, I., Khan, S., Adnan, M., Fahad, S., Wahid, F., Dawar, K., Ali, S., Zia, A., Ahmad, M., Khan, S., Ali Shah, W., Romman, M., Parvez, R., Siddiqui M, H., Khan, A., Wang, D., Jiang, X., 2021. Comparing the phosphorus use efficiency of pre-treated (organically) rock phosphate with soluble P fertilizers in maize under calcareous soils. PeerJ *9*.
Nakagawa, S., Johnson, P.C.D., Schielzeth, H., 2017. The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. J. R. Soc. Interface 14, 20170213.
NAL USDA. 2021. *NAL Agricultural Theaurus and Glossary*. Retrieved 7 December 2021 from: ⟨http://lod.nal.usda.gov/nalt/966⟩.
Negash, H., Wondimu, W., Abraham, A., 2021. Maize yield improvement by optimal rate and timing of nitrogen fertilizer application in Southwest Ethiopia. Agric. Biol. Res. 37.
Nenova, L., Benkova, M., Simeonova, T., Atanassova, I., 2019. Nitrogen, phosphorus and potassium content in maize dry biomass under the effect of different levels of mineral fertilization. Agric. Sci. Technol. 1313–8820, 11.
NIST/SEMATECH. 2012. *e-Handbook of Statistical Methods*. Retrieved 9 December 2020 from: ⟨https://www.itl.nist.gov/div898/handbook/eda/section3/qqplot.htm⟩.
Njoroge KS. 2019. *Feed the crop, not the soil! Explaining variability in maize yield responses to nutrient applications in smallholder farms of western Kenya*. Unpublished thesis, Wageningen University, Wageningen, The Netherlands, 189 p.
Nsanzabaganwa, E., Das, T.K., Rana, D.S., Kumar, S.N., 2014. Nitrogen and phosphorus effects on winter maize in an irrigated agroecosystem in western Indo-Gangetic plains of India. Maydica 59, 9.
Ortas, I., 2018. Influence of potassium and magnesium fertilizer application on the yield and nutrient accumulation of maize genotypes under field conditions. J. Plant Nutr. 41, 330–339.
Parco, M., Ciampitti, I.A., D'Andrea, K.E., Maddonni, G.Á., 2020. Prolificacy and nitrogen internal efficiency in maize crops. Field Crop. Res. 256, 1–13.
Pasley, H.R., Cairns, J.E., Camberato, J.J., Vyn, T.J., 2019. Nitrogen fertilizer rate increases plant uptake and soil availability of essential nutrients in continuous maize production in Kenya and Zimbabwe. Nutr. Cycl. Agroecosyst. 115, 373–389.
Peng, C., Zhang, Z., Li, Y., Zhang, Y., Dong, H., Fang, Y., Han, L., Xu, W., Hu, L., 2022. Genetic improvement analysis of nitrogen uptake, utilization, translocation, and distribution in Chinese wheat in Henan Province. Field Crop. Res. 277, 108406.
Perry, L.J., Olson, R.A., 1975. Yield and quality of corn and grain sorghum grain and residues as influenced by N fertilization. Agron. J. 67, 816–818.
Philibert, A., Loyce, C., Makowski, D., 2013. Prediction of N_2O emission from local information with random forest. Environ. Pollut. 177, 156–163.
Pinheiro, J., Bates, D.M., 2000. Mixed-Effects Models in S and S-PLUS. Springer-Verlag, New York, USA.
Qiang, S., Sun, X., Zhang, Y., Zhao, H., Fan, J., Zhang, F., Sun, M., Gao, Z., 2020. Deep placement of mixed controlled-release and conventional urea improves grain yield, nitrogen use efficiency of rainfed spring maize. Arch. Agron. Soil Sci. 1–11.
R Core Team. 2020. *R: a language and environment for statistical computing* ⟨http://www.R-project.org⟩.
Ramadhan, M.N., 2021. Yield and yield components of maize and soil physical properties as affected by tillage practices and organic mulching. Saudi J. Biol. Sci.
Ray, K., Banerjee, H., Dutta, S., Sarkar, S., Murrell, T.S., Singh, V.K., Majumdar, K., 2020. Macronutrient management effects on nutrient accumulation, partitioning, remobilization, and yield of hybrid maize cultivars. Front. Plant Sci. 11, 1–19.

Saidou, A., Janssen, B.H., Temminghoff, E.J.M., 2003. Effects of soil properties, mulch and NPK fertilizer on maize yields and nutrient budgets on ferralitic soils in southern Benin. Agric., Ecosyst. Environ. 100, 265–273.

Salvagiotti, F., Prystupa, P., Ferraris, G., Couretot, L., Magnano, L., Dignani, D., Gutiérrez, Bgutiérrez-Boemoem, F.H., 2017. N:P:S stoichiometry in grains and physiological attributes associated with grain yield in maize as affected by phosphorus and sulfur nutrition. Field Crops Res. 203, 128–138.

Salvagiotti, F., Magnano, L., Ortez, O., Enrico, J., Barraco, M., Barbagelata, P., Condori, A., Di Mauro, G., Manlla, A., Rotundo, J., Garcia, F.O., Ferrari, M., Gudelj, V., Ciampitti, I., 2021. Estimating nitrogen, phosphorus, potassium, and sulfur uptake and requirement in soybean. Eur. J. Agron. 127, 126289.

Sanders A. 2020. Assessing sulfur fertilizer response in Ontario's corn, soybean, and winter wheat crops. Unpublished thesis.

Setiyono, T., Walters, D.T., Cassman, K., Witt, C., Dobermann, A., 2010. Estimating maize nutrient uptake requirements. Field Crop. Res. 118, 158–168.

Sharma, M., Parmar, D.K., Sharma, S.K., Kumar, P., Kumar, P., 2021. Influence of zinc nutrition on zinc availability, uptake and crop yield of rainfed maize-pea sequence in Northwestern Himalaya. J. Plant Nutr. 44, 2069–2080.

Shehu, B.M., Lawan, B.A., Jibrin, J.M., Kamara, A.Y., Mohammed, I.B., Rurinda, J., Zingore, S., Craufurd, P., Vanlauwe, B., Adam, A.M., Merckx, R., 2019. Balanced nutrient requirements for maize in the Northern Nigerian Savanna: Parameterization and validation of QUEFTS model. Field Crop. Res. 241, 107585.

Shi, Z., Liu, D., Liu, M., Hafeez, M.B., Wen, P., Wang, X., Wang, R., Zhang, X., Li, J., 2021. Optimized fertilizer recommendation method for nitrate residue control in a wheat–maize double cropping system in dryland farming. Field Crop. Res. 271, 108258.

Simić, M., Dragičević, V., Mladenović Drinić, S., Vukadinović, J., Kresović, B., Tabaković, M., Brankov, M., 2020. The contribution of soil tillage and nitrogen rate to the quality of maize grain. Agronomy 10, 1–14.

Singh, J., Partap, R., Singh, A., Kumar, N., Krity, 2021. Effect of nitrogen and zinc on growth and yield of maize (Zea mays L.). Int. J. Bio-Resour. Stress Manag. 12, 179–185.

Singh, S.P., Dutta, S.K., Jha, S., Prasad, S.S., Chaudhary, S.K., Sahi, V., Majumdar, K., 2020. Nutrient management in calcareous soil improves rice–maize sustainable yield index, performance indicators. J. Plant Nutr. 1–16.

Sowmya, S.J., Singh, S., Singh, V., 2021. Performance of plant growth regulators and micronutrient on growth and yield of maize (Zea mays L.). Asian J. Microbiol. Biotechnol. Environ. Sci. 23, 175–181.

Stoffel, M.A., Nakagawa, S., Schielzeth, H., 2021. partR2: Partitioning R2 in generalized linear mixed models. PeerJ 9, 1–14.

Su, W., Ahmad, S., Ahmad, I., Han, Q., 2020. Nitrogen fertilization affects maize grain yield through regulating nitrogen uptake, radiation and water use efficiency, photosynthesis and root distribution. PeerJ 8, 1–21.

Széles, A., Nagy, J., Ratonyi, T., Harsányi, E., 2019. Effect of differential fertilisation treatments on maize hybrid quality and performance under environmental stress condition in Hungary. Maydica 62, 11–14.

Tadesse, A., Kim, H.K., 2014. Yield related traits and yield of quality protein maize (Zea mays L.) affected by nitrogen levels to achieve maximum yield in the Central Rift Valley of Ethiopia. J. Biol. Agric. Healthc. 5, 139–148.

Tadesse, A., Sultan, M., 2021. Evaluation of NPSB fertilizer levels on yield and yield components of open pollinated maize (Zea mays L.) varieties under irrigated condition in South Omo Zone. Am. J. Biosci. 9, 86–94.

Tamindžić, G., Ignjatov, M., Milošević, D., Nikolić, Z., Kostić Kravljanac, L., Jovičić, D., Dolijanović, Ž., Savić, J., 2021. Seed priming with zinc improves field performance of maize hybrids grown on calcareous chernozem. Ital. J. Agron. 16, 1–7.

Tenorio, F.A.M., Eagle, A.J., McLellan, E.L., Cassman, K.G., Howard, R., Below, F.E., Clay, D.E., Coulter, J.A., Geyer, A.B., Joos, D.K., Lauer, J.G., Licht, M.A., Lindsey, A. J., Maharjan, B., Pittelkow, C.M., Thomison, P.R., Wortmann, C.S., Sadras, V.O., Grassini, P., 2019. Assessing variation in maize grain nitrogen concentration and its implications for estimating nitrogen balance in the US North Central region. Field Crop. Res. 240, 185–193.

Thakur, A., Sharma, R., Sankhyan, N., Kumar, R., 2020. Maize grain quality as influenced by 46 years' continuous application of fertilizers, farmyard manure (FYM), and lime in an alfisol of North-western Himalayas. Commun. Soil Sci. Plant Anal. 1–12.

Trachsel, S., San Vicente, F.M., Suarez, E.A., Rodriguez, C.S., Atlin, G.N., 2016. Effects of planting density and nitrogen fertilization level on grain yield and harvest index in seven modern tropical maize hybrids (Zea mays L.). J. Agric. Sci. 154, 689–704.

Tsai, C.Y., Dweikat, I., Huber, D.M., Warren, H.L., 1992. Interrelationship of nitrogen nutrition with maize (Zea mays) grain yield, nitrogen use efficiency and grain quality. Food Agric. 58, 1–8.

Ullah, I., Dawar, K., Tariq, M., Sharif, M., Fahad, S., Adnan, M., Ilahi, H., Nawaz, T., Alam, M., Ullah, A., Arif, M., 2021. Gibberellic acid and urease inhibitor optimize nitrogen uptake and yield of maize at varying nitrogen levels under changing climate. Environ. Sci. Pollut. Res. Int.

Unkovich, M., Baldock, J., Forbes, M., 2006. A review of biological yield and harvest index in Australian field crops. Advances in Agronomy. Elsevier Academic Press, pp. 174–219.

van Grinsven, H.J.M., Ebanyat, P., Glendining, M., Gu, B., Hijbeek, R., Lam, S.K., Lassaletta, L., Mueller, N.D., Pacheco, F.S., Quemada, M., Bruulsema, T.W., Jacobsen, B.H., ten Berge, H.F.M., 2022. Establishing long-term nitrogen response of global cereals to assess sustainable fertilizer rates. Nat. Food.

van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap analysis with local to global relevance—a review. Field Crop. Res. 143, 4–17.

Van Reuler, H., Janssen, B., 1996. Optimum NPK management over extended cropping periods in south-west Cote d'Ivoire. NJAS Wagening. J. Life Sci. 44, 263–277.

Walia, S., Kumar, R., 2021. Elucidating the yield and quality response of Tagetes minuta L. intercropped with Zea mays L. under different spacing in the western Himalayas. Ind. Crop. Prod. 171, 113850.

WHO, FAO, 2007. Codex Alimentarius, first ed. World Health Organisation, Food and Agriculture Organization of the United Nations, Rome, Italy, p. 243.

Witt, C., Dobermann, A., Abdulrachman, S., Gines, H.C., Guanghuo, W., Nagarajan, R., Satawatananont, S., Thuc Son, T., Sy Tan, P., Van Tiem, L., Simbahan, G.C., Olk, D. C., 1999. Internal nutrient efficiencies of irrigated lowland rice in tropical and subtropical Asia. Field Crop. Res. 63, 113–138.

Woli, K.P., Sawyer, J.E., Boyer, M.J., Abendroth, L.J., Elmore, R.W., 2018. Corn era hybrid macronutrient and dry matter accumulation in plant components. Agron. J. 110, 1648–1658.

Wortmann, C.S., Dobermann, A.R., Ferguson, R.B., Hergert, G.W., Shapiro, C.A., Tarkalson, D.D., Walters, D.T., 2009. High-yielding corn response to applied phosphorus, potassium, and sulfur in Nebraska. Agron. J. 101, 546–555.

Xu, X., He, P., Pampolino, M.F., Chuan, L., Johnston, A.M., Qiu, S., Zhao, S., Zhou, W., 2013. Nutrient requirements for maize in China based on QUEFTS analysis. Field Crop. Res. 150, 115–125.

Yang, C., Du, W., Zhang, L., Dong, Z., 2021. Effects of sheep manure combined with chemical fertilizers on maize yield and quality and spatial and temporal distribution of soil inorganic nitrogen. Complexity 2021, 4330666.

Zhang, F., Mackenzie, A.F., Smith, D.L., 1993. Corn yield and shifts among corn quality constituents following application of different nitrogen fertilizer sources at several times during corn development. J. Plant Nutr. 16, 1317–1337.

Zhang, M., Geng, Y., Cao, G., Zou, X., Qi, X., Stephano, M.F., 2020. Effect of magnesium fertilizer combined with straw return on grain yield and nitrogen use efficiency. Agron. J.

Zhang, X., Zou, T., Lassaletta, L., Mueller, N.D., Tubiello, F., Lisk, M.D., Lu, C., Conant, R. T., Dorich, C.D., Gerber, J., Tian, H., Bruulsema, T., McClellan-Maaz, T., Nishina, K., Leon, B., Bodirsky, L.B., Popp, A., Bouwman, L., Beusen, A., Chang, J., Havlík, P., Leclère, D., Canadell, J.G., Jackson, R.B., Billen, G., Heffer, P., Wanner, N., Zhang, W., Davidson, E.A., 2021a. Quantification of global and national nitrogen budgets for crop production. Nat. Food 1–14.

Zhang, X., Wang, R., Dong, Z., Zhang, P., Jia, Z., Han, Q., 2021b. Nutritional quality degradation: a potential risk due to nutrient dilution effects in film-mulched maize. Agric. Water Manag. 257, 107133.