Estimating soybean harvest index and nitrogen concentrations of grain and residue using globally available data

Emil Löwik

MSc Thesis PPS

March 2023

WAGENINGEN UNIVERSITY & RESEARCH

Estimating soybean harvest index and nitrogen concentrations of grain and residue using globally available data

MSc Thesis PPS

Name: Emil Löwik

Student number: 1125303

Study: MSc Plant Sciences (Specialisation C – Natural Resource Management)

Supervisors: Renske Hijbeek & Cameron Ludemann

Chair group: Plant Production Systems

Course code: PPS80436

Date: March 2023

Examiner:

Citation: Löwik, E., 2023, Estimating soybean harvest index and nitrogen concentrations of grain and residue using globally available data, MSc Thesis Wageningen University, p. 49

Contact: office.pp@wur.nl for access to data and scripts used for analysis

TABLE OF CONTENTS

Acknowledgements

First and foremost, I would like to thank my supervisors - Renske Hijbeek and Cameron Ludemann - for their help and support throughout this master's thesis project. I would also like to thank Tobias Bader for all the work he put in standardising soybean data from peer reviewed articles to supplement the dataset used in this study. Lastly, I would like to say that I truly enjoyed working on my master's thesis at the Plant Production Systems (PPS) chair group. PPS makes a sincere effort to facilitate master students during their thesis and gives them the opportunity to tag along as a member of the chair group. Thanks to everyone involved.

Key words and abbreviations

HI: Harvest index, proportion of aboveground biomass as grain yield

Grain N: Grain nitrogen concentration, in % of grain dry matter

Residue N: Residue nitrogen concentration, in % of residue dry matter

GY: Dry matter grain yield, in Mg ha^{-1}

RY: Dry matter residue yield, in Mg ha^{-1}

FN: Fertiliser nitrogen, total elemental nitrogen fertiliser applied in kg ha^{-1}

FP: Fertiliser phosphorus, total elemental phosphorus fertiliser applied in kg ha^{-1}

FK: Fertiliser potassium, total elemental potassium fertiliser applied in kg ha^{-1}

NDFA%: Nitrogen derived from atmosphere, expressed as a % of total nitrogen in aboveground plant biomass

NDFAkg: Nitrogen derived from atmosphere, expressed in kg N ha^{-1} in aboveground plant biomass

SoilP: Soil elemental phosphorus budget, cumulative soil phosphorus budget calculated with data from FAOSTAT (2022) in kg P m^{-2}

SoilK: Soil elemental potassium budget, cumulative soil potassium budget calculated with data from FAOSTAT (2022) in kg K m^{-2}

Abstract

Managing nutrients to sustain yields and reduce environmental pressure is one of the largest challenges agricultural systems currently face. Accurately quantifying the removal of nitrogen through crops can aid decision-making regarding nutrient management. Harvest index is an important variable in accurately quantifying crop nutrient removal, as it allows for the distinction to be made between grain and residue. In addition, biological nitrogen fixation through leguminous crops has gained renewed interest as a nitrogen input because it can decrease the necessity of nitrogenous fertilisers to sustain crop yields. This study aimed to investigate whether variability in soybean (*Glycine max*) nitrogen concentrations of grain and residue, harvest index, and biological nitrogen fixation could be explained using variables in globally available data.

Using a large database of 82 field experiments, harvest index and grain and residue nitrogen concentrations were predicted, comparing accuracy of linear mixed-effects models and random forest regression. Linear mixed-effects models were trained with 80% of the data to explain variation in response variables. Explanatory power of linear mixed-effects models was determined based on Nakagawa's R^2 and Akaike's Information Criterion. The other 20% of the data was used to validate predictions of response variables by the best linear mixed-effects models and random forest regression.

Across all experiments, mean harvest index for soybean was 0.38 (standard deviation (SD) = 0.09), mean nitrogen content of grain and residue was 5.9% (SD = 0.9%) and 1.1% (SD = 0.6%) and mean NDFA was 139 kg N ha^{-1} (SD = 93 kg N ha^{-1}) or 56% (SD = 19%). Region showed to be an important explanatory variable for all response variables. In addition, fertiliser application rates and cumulative soil phosphorus and potassium budgets also contributed to the explanatory power of many linear mixed-effects models. Higher nitrogen fertiliser application rates were consistent with lower harvest indices and higher nitrogen concentrations in both grain and residue. Prediction with random forest regression was more accurate compared to prediction with linear mixed-effects models across all response variables. Accurate predictions were made for harvest index and grain nitrogen concentration, with R^2 values of 0.83 and 0.95. Lack of data was a limitation in explaining variability and predicting residue nitrogen concentration and biological nitrogen fixation. Expanding the dataset with response variable data and explanatory variables, such as yield potential and variety information, could further improve the understanding and prediction of nitrogen flows in soybean cultivation.

1. Introduction

1.1 Intensification of Agriculture

Agriculture in the past century has primarily focused on realising high yields. In Europe, after the Second World War, a number of policies were implemented to incentivise farmers to produce more, to increase food self-sufficiency and guarantee affordable food. This led to a rapid intensification of agriculture in Europe, with significant increases in yield as a result (Emmerson et al., 2016). Innovations in many fields of agriculture helped farmers realise higher yields, from new agricultural machinery and equipment to developments in synthetic fertiliser and crop protection products. A critical turning point being the invention of the Haber-Bosch process in 1908, which enabled the synthesis of ammonia as a nitrogen fertiliser (Iannetta et al., 2016; Zhang et al., 2015).

Fertilisers, both synthetic and organic, have played an important role in realising yield increases over the past decades and are still vital to sustain production levels today. The percentage of crop yields attributable to commercial fertilisers is estimated to be as high as 60% in the USA and England (Stewart et al., 2005). However, more recently the use of fertilisers is put in a different perspective due to increased environmental concerns. Several of these environmental concerns are directly related to nitrogen pollution. Nitrogen is the most abundant nutrient in many fertilisers and is prone to volatilisation and leaching. Nitrogen pollution has negative effects on human health, contributes to climate change and threatens both aquatic and terrestrial ecosystems (Schulte-Uebbing et al., 2022; Van Egmond et al., 2002). In many agricultural systems, nutrient management is suboptimal. Nitrogen is often oversupplied as synthetic fertilisers are readily available and organic manures are accessible at low costs or even from an income stream to arable farmers due to a local, intensive livestock sector (Bos et al., 2017; Silva et al., 2021). Carefully managing nutrients to sustain yields and reduce environmental pressure is one of the largest challenges these agricultural systems currently face (West et al., 2014).

In contrast to a nitrogen surplus, there are many agricultural systems which lack sufficient nitrogen inputs, and face nitrogen stress or even scarcity. These agricultural systems are predominantly found in Sub-Saharan Africa, Central and South America and Southeast Asia (Liu et al., 2010; Schulte-Uebbing et al., 2022). This is illustrated in Figure 1, which shows exceedance of critical nitrogen surplus, where green areas indicate potential to increase nitrogen inputs without exceeding environmental limits. As a result, there are significant yield gaps in some of these areas. Moreover, it is predicted that many of these areas will experience rapid population growth in the near future (United Nations, 2015). In this context, agricultural intensification is not so much a problem but a necessity (Van Ittersum et al., 2016). To improve global food security and environmental sustainability, some regions will require an increase in nutrient inputs whilst others may benefit from a reduction of nutrient inputs (Schulte-Uebbing et al., 2022; West et al., 2014).

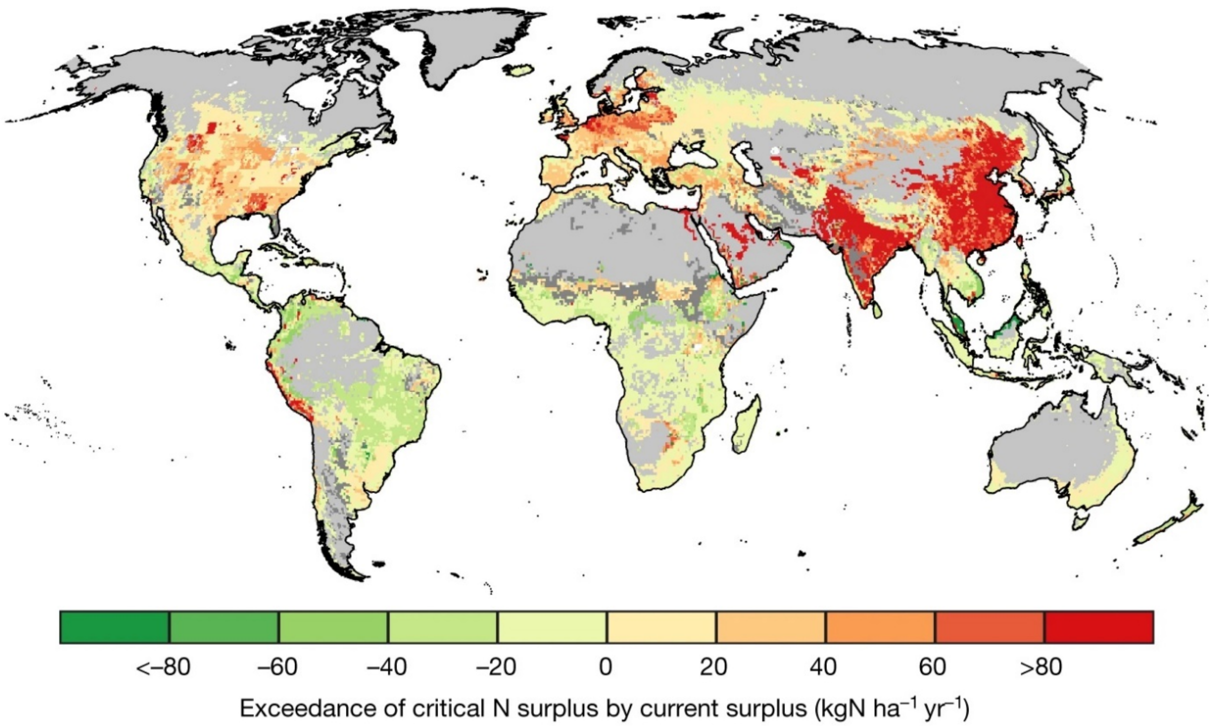

Figure 1: Spatial variation in global exceedance of nitrogen thresholds. Reductions of agricultural nitrogen surplus to respect deposition, surface water and groundwater thresholds simultaneously. Positive values (red) indicate needed reductions and negative values (green) indicate possible increases within thresholds. Reprinted from Schulte-Uebbing et al. (2022).

1.2 Nitrogen Budgets

Improved agronomic management can help reduce nutrient losses from croplands and increase yields in low input systems. Applying fertiliser at the right rate, source, time, and place can significantly reduce nitrogen losses whilst sustaining yields (Ma et al., 2022; Roberts, 2010). Accurate quantifications of nitrogen flows are important to optimise nutrient management. Nitrogen budgets, also known as nitrogen balances, are the difference between nitrogen inputs and outputs in a system. Nitrogen budgets can act as a tool to quantify and reduce nitrogen surpluses, which in turn can have economic benefits for farmers due to lower feed and fertiliser costs (Langeveld et al., 2007c). A complete nitrogen budget includes all the nitrogen inputs and outputs for a system. However, due to data limitations, partial nitrogen budgets can be calculated as an indicator for a specific nitrogen flow. Moreover, nitrogen budgets give insight on possible nitrogen stress or scarcity. These budgets can help identify areas in which yield gaps may be present due to nitrogen stress (West et al., 2014). Knowledge on the removal of nitrogen through crops can aid decision-making regarding nutrient management. Accurately quantifying the flow of nitrogen between various boundaries helps farmers, policy makers, and scientists work towards global food security and reduce the environmental effects of agriculture (Zhang et al., 2021).

An important component in nitrogen budget calculations is the nitrogen in aboveground plant biomass. Aboveground plant biomass captures the main nitrogen output from a field. Aboveground plant biomass can be further categorised into two separate components: crop products and residues. For an accurate nitrogen budget calculation, it is relevant to distinguish between these two components. Many crop products, such as grain or fruits, contain higher nitrogen concentrations compared to residues such as straw, stalks, and leaves (Nijhof, 1987). In addition, crop residues are sometimes left on the field and can be a nitrogen input, whilst crop products are taken off the field and are considered a nitrogen output. Widely available data on crop residue is lacking. Often, crop product yields are measured out of economic interest, but residue yields are not. For instance, country-level crop product yields are accessible in open-source databases, such as *FAOSTAT* (2022), but data on residue yields are not included within this database. To be able to distinguish between crop products and residues in nitrogen budget calculations, harvest index acts as an important conversion factor to calculate residue yield from crop product yield data. Together with the nitrogen concentration, the nitrogen component for residues can be calculated and included in a nitrogen budget calculation. Harvest index thereby allows for the distinction to be made between crop product and residue, resulting in a more accurate quantification of nitrogen budgets (Ludemann et al., 2022).

1.3 Biological Nitrogen Fixation

Biological nitrogen fixation (BNF) through leguminous crops has gained renewed interest as a nitrogen input because it can decrease the necessity of nitrogenous fertilisers to sustain crop yields (Iannetta et al., 2016). BNF occurs naturally in the root nodules of leguminous plants by nitrogen-fixing bacteria such as Rhizobia, a genus of soil bacteria that can reduce atmospheric nitrogen to reactive forms of nitrogen. For this reason, BNF through legumes, on itself or in combination with fertilisers, may be implemented as a sustainable farming practice (Iannetta et al., 2016; Jensen & Hauggaard-Nielsen, 2003). A study by Sainju et al. (2016) suggests that crop rotations with legumes have potential for reducing nitrogen surpluses and losses in intense farming systems whilst they might facilitate sustainable intensification of African smallholder farms (Vanlauwe et al., 2019).

1.4 Research Objective

This thesis built upon previous research which developed a methodology to create predictive models for maize harvest index and nitrogen concentrations of grain and residue using globally available data (Ludemann et al., 2022). The present study aimed to investigate whether variability in nitrogen concentrations of soybean (*Glycine max*) grain and residue as well as harvest index could be explained using variables in globally available data. Soybean was chosen as its production has been steadily increasing in the past decades and it is the only legume amongst the world's major staple crops (FAOSTAT, 2022). Harvest index and nitrogen concentrations of grain and residue as response variables are relevant for more accurate quantification of (partial) nitrogen budgets. In addition, this study investigated whether these variables could explain the variability in biological nitrogen fixation of soybean. Biological nitrogen fixation response variables were Nitrogen Derived from Atmosphere (NDFA%), expressed as a percentage of total nitrogen and as an absolute value in kg N ha^{-1} in aboveground plant biomass (NDFAkg). Lastly, linear mixed-effects models and random forest regression were used to predict soybean grain and residue nitrogen concentrations, harvest index, NDFAkg and NDFA%. To achieve this main objective, the following three research questions (RQ) were formulated.

1.5 Research Questions

1. What is the variability observed in soybean harvest index, nitrogen concentrations of grain and residue, NDFAkg and NDFA%?
2. What is the explanatory power of variables influencing soybean harvest index, nitrogen concentrations of grain and residue, NDFAkg and NDFA%?
3. To what extent can linear mixed-effects models and random forest regression predict soybean harvest index and nitrogen concentrations of grain and residue?

2. Material and Methods

2.1 Data Collection

The research questions were answered by analysing data from two main sources, including data from peer reviewed literature and from an open-source database. Data from peer reviewed articles all came from replicated field trials. Minimum required information for selection of data from peer reviewed literature is included in Table 1.

Table 1
List of required information and data variables within the peer reviewed articles for inclusion in this analysis

Variable
Name of author and organisation responsible for field trial
Year of sowing
Trial location
Number of replicates
Yield of soybean grain and/or residue and/or harvest index
Soybean grain and/or residue nitrogen concentration
Fertiliser application rates

Peer reviewed literature

An existing dataset from peer reviewed articles formed the basis to the dataset used in this project. The existing dataset included summary statistics from peer reviewed articles on the major staple crops, maize, wheat, soybean, and rice (Ludemann et al., 2023). For the purpose of this research, the dataset was filtered to data from soybean. Following, data from peer reviewed literature was added, under the premise that it met the requirements listed in Table 1.

Open-source data

The existing dataset was supplemented with data from the open-source database FAOSTAT. Nutrient budgets at country level were used to calculate accumulative soil phosphorus and potassium levels, which in turn were used as variables in linear mixed-effects modelling and random forest regression.

Data standardisation and cleaning

Data from peer reviewed articles were standardised and gathered in a single dataset. This dataset had uniform units per variable and fertiliser application rates were standardised to elemental fertiliser application. Before further analysis, the variables grain N, residue N, NDFAkg, NDFA%, harvest index and grain yield were checked for outliers. Data points under the 1st quartile (Q1) by 1.5 times the interquartile range (IQR) and above the 3rd quartile (Q3) by 1.5 times the IQR were deemed outliers. Statistical outliers were found and removed from the dataset for grain N (N = 19), residue N (N = 22), harvest index (N = 1), NDFAkg (N = 2), NDFA% (N = 1) and grain yield (N = 98). Moreover, data from trials with extreme fertiliser application rates were also excluded from the dataset. Maximum elemental nitrogen, phosphorus and potassium fertiliser application rates included in the data were 250 kg N ha^{-1}, 100 kg P ha^{-1} and 100 kg K ha^{-1} respectively. Fertiliser data were excluded for all fertiliser application rate variables; FN (N = 98), FP (N = 6) and FK (N = 103).

After outliers were removed, the analysis included data from 82 studies (Table 9, Appendix 4) and 31 different countries, with experimental years ranging between 1967 and 2020. The dataset included 1706 datapoints across all response variables for soybean, with sample sizes for the respective response variables ranging from 110 to 619 data points (Table 7, Appendix 1). This dataset was used to determine the variability of response variables (RQ1). Hereafter, data was split randomly in two subsets: a training and testing subset. The training subset included 80% of the data and was used to train linear mixed-effects models (RQ2) and random forest regression. The testing subset included 20% of the data and was used to validate predictions of response variables using both linear mixed-effects models and random forest regression (RQ3). Due to the lack of data for NDFAkg and NDFA% within the dataset, no attempt was made to predict these response variables. The full dataset was used for explaining variation in NDFAkg and NDFA% with linear mixed-effects models.

2.2 Statistical Analysis

All statistical analyses, modelling, prediction, as well as creating figures and graphs were performed using the programme R (R Core Team, 2022) following methodology described by Ludemann et al. (2022).

Data Variability

To answer the first research question on the variability of soybean grain N, residue N, harvest index, NDFAkg and NDFA% within the dataset, the range and distribution of data were plotted and analysed using summary statistics and the “ggplot2” package in R. Variables for linear mixed-effects modelling were chosen visually using the “chart.Correlation” function in R (Figure 12, Appendix 3) and available knowledge on biological processes. These first analyses were performed on the complete dataset, after outliers were removed.

Linear mixed-effects models

To identify which variables best explained the observed variability within the dataset, linear mixed-effects models were made for each response variable using the “lme4’ package in R. UN sub-region in which the data was gathered was used as a random effect in all linear mixed-effects models. Nakagawa’s conditional R^2 was used to evaluate how well the linear mixed-effects models explain variation in the response variables. Akaike’s Information Criterion was used to determine whether adding certain variables contributes to the explanatory power of the model. These indicators were determined by using the “performance” package in R. Data were weighed based on the years of data and the number of replicates from each mean value. Explanation of variation using linear mixed-effects models was performed on the training subset. Variable inflation factors were used to ensure no unacceptably high correlations between explanatory variables in linear mixed-effects models were present.

2.3 Prediction

Linear mixed-effects models

For each of the response variables grain N, residue N and harvest index, three linear mixed-effects models were subjectively chosen as predictor models. This was done based on the highest Nakagawa’s Conditional R^2, lowest AIC, as well as the number of observations within the dataset for a particular model. The number of observations were an important factor in choosing predictor models. It could therefore be the case that models with few observations were not chosen as predictor models despite their high Nakagawa Conditional R^2 and low AIC. The predictor models were tested using the testing subset (Figure 11, Appendix 2). The predicted values were plotted against the actual values and the R^2 was determined through linear regression analysis. The best linear mixed-effects predictor model was chosen based on the highest R^2 of the actual versus predicted plot and used to compare to prediction with random forest regression.

Random forest regression

The R package “randomForest” was used to perform random forest regression. The same explanatory variables as the best linear mixed-effects predictor model were used as input variables for random forest regression, for a fair comparison between the two methods of prediction. Predicted values from random forest regression were also plotted against actual values and the R^2 was determined through linear regression. Prediction with random forest regression was also performed on the testing subset.

3. Results

3.1 Descriptive Statistics

The variability of grain N, residue N, NDFAkg, NDFA% and harvest index within the dataset is shown in boxplots in Figure 2. Grain N ranged between 3.61% and 8.16%, with a mean of 5.93%. In comparison, residue N had a lower mean of 1.11%, whilst values ranged from 0.11% to 2.64%. Mean harvest index within the dataset was 0.39, with values ranging between 0.11 and 0.55. NDFAkg ranged between 3 kg N ha^{-1} and 354 kg N ha^{-1} with a mean of 139 kg N ha^{-1}. The mean of NDFA%, expressed as a percentage of total N in aboveground plant biomass, lies at 56%, with a minimum value of 9% and a maximum of value of 98% .

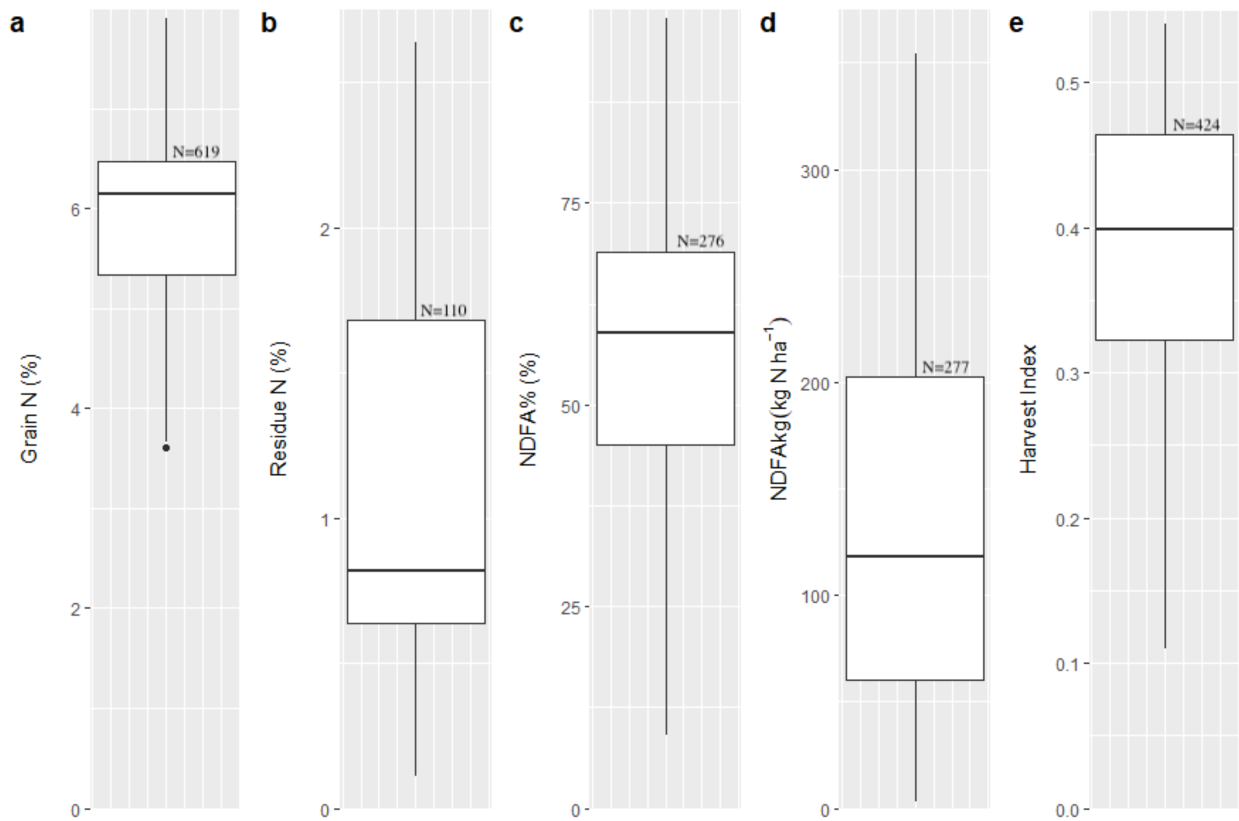

Figure 2: Box plots for the variables **(a)** Grain N, **(b)** Residue N, **(c)** NDFA%, **(d)** NDFAkg and **(e)** Harvest Index. The box plots represent all data within the dataset after outliers were removed.

The distribution of data for all response variables per UN sub-region is shown in Figure 3. Although there is overlap, grain N ranges and means differ between regions (Figure 3a). In Southern Europe and South-eastern Asia, grain N content of soybean is highest, with respective mean values of 6.48% and 7.04%. In Eastern Europe, grain N content of soybean is the lowest with a mean of 4.49%. The density plot for residue N shows that Northern America and Southern Europe have more residue N values at the upper range relative to other regions with mean values of 1.51% and 1.85% respectively (Figure 3b). Southern Asia also has a relatively high mean residue N content of 1.65%, but the data is distributed across a wider range of 0.53% to 2.64%.

In Australia and New Zealand and Latin America and the Caribbean observed soybean NDFAkg values were highest, with respective means of 178 kg N ha^{-1} and 276 kg N ha^{-1} (Figure 3d). Latin America and the

Caribbean also has a relatively high mean NDFA% of 66%, followed by Sub-Saharan Africa and Northern America with respective means of 64% and 61%. Eastern Asia has the highest mean NDFA% at 67% (Figure 3c).

Harvest index distributions between regions show much overlap, with means for most regions ranging between 0.37 and 0.43 (Figure 3e). Australia and New Zealand and Northern Africa are two regions which stand out with a larger share of harvest index values at the lower range. Australia and New Zealand and Northern Africa are the only two regions which fall outside the abovementioned range of means, with respective means of 0.22 and 0.28. Latin America and the Caribbean has the highest mean harvest index of 0.43, followed by Northern America and Western Europe with mean harvest indices of 0.42.

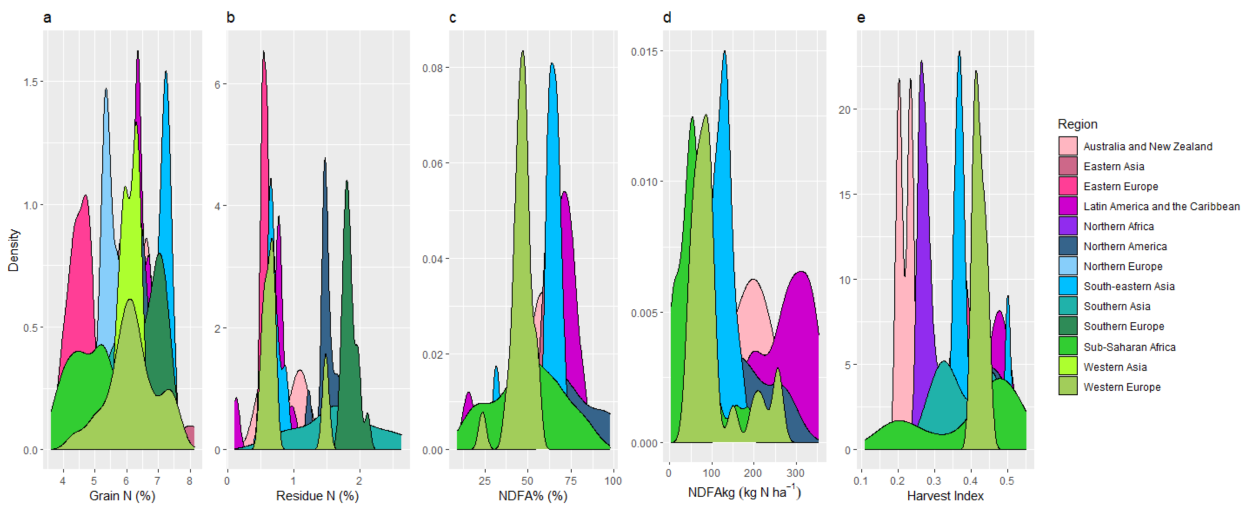

Figure 3: Density plots for the variables **(a)** Grain N, **(b)** Residue N, **(c)** NDFA%, **(d)** NDFAkg and **(e)** Harvest Index. The density plots represent all data within the dataset after outliers were removed.

3.2 Explanatory Variables

Variables explaining variation in harvest index

A positive relationship was observed in which harvest index increases with grain yield. However, high nitrogen fertiliser application rates were consistent with lower harvest indices (Figure 4). Model H1_10, which included GY and FN as fixed effects, had a Nakagawa's R^2 of 0.20 (Table 2).

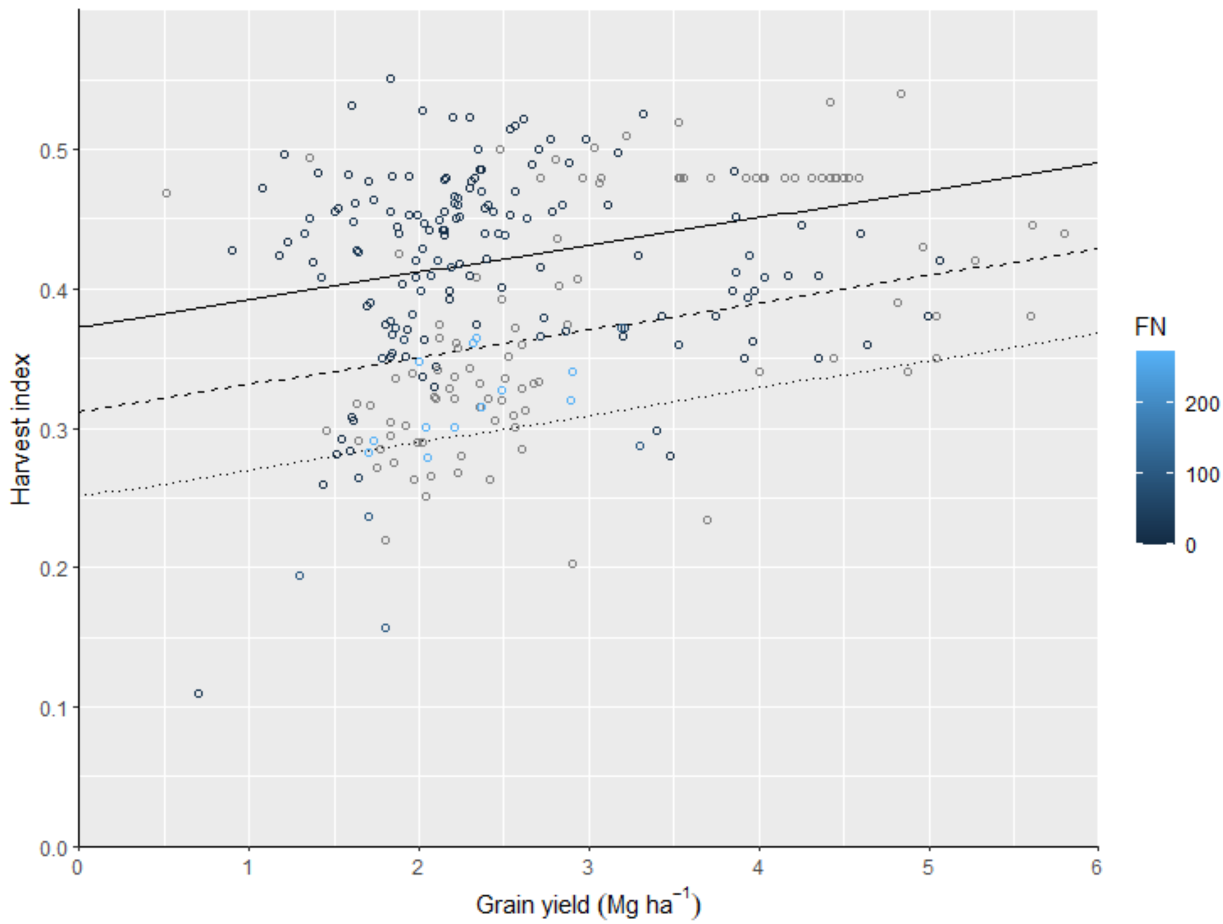

Figure 4: Soybean harvest index in relation to grain yield (GY) and elemental nitrogen fertiliser application rate (FN). Regression lines show linear mixed-effects model with GY and FN as fixed effects, UN sub-region as random effect and years of data x replicates as weight, at 0 kg N ha^{-1} (solid line), 100 kg N ha^{-1} (dashed line) and 200 kg N ha^{-1} (dotted line) FN.

A similar positive relationship between harvest index and grain yield was found for linear mixed-effects model H1_9, however higher SoilP values were consistent with lower harvest indices. Model H1_9, the linear mixed-effects model explaining harvest index with GY and SoilP as fixed effects, has a Nakagawa R^2 value of 0.96.

Model H1_6 and H1_9 had the highest Nakagawa's R^2 value of 0.96 out of all linear mixed-effects models explaining variation in harvest index. Model H1_6 included only SoilP as a fixed effect, whilst H1_9 included both SoilP and GY as fixed effects (Table 2). The addition of GY as a fixed effect to model H1_6 did not increase the explanatory power of the model.

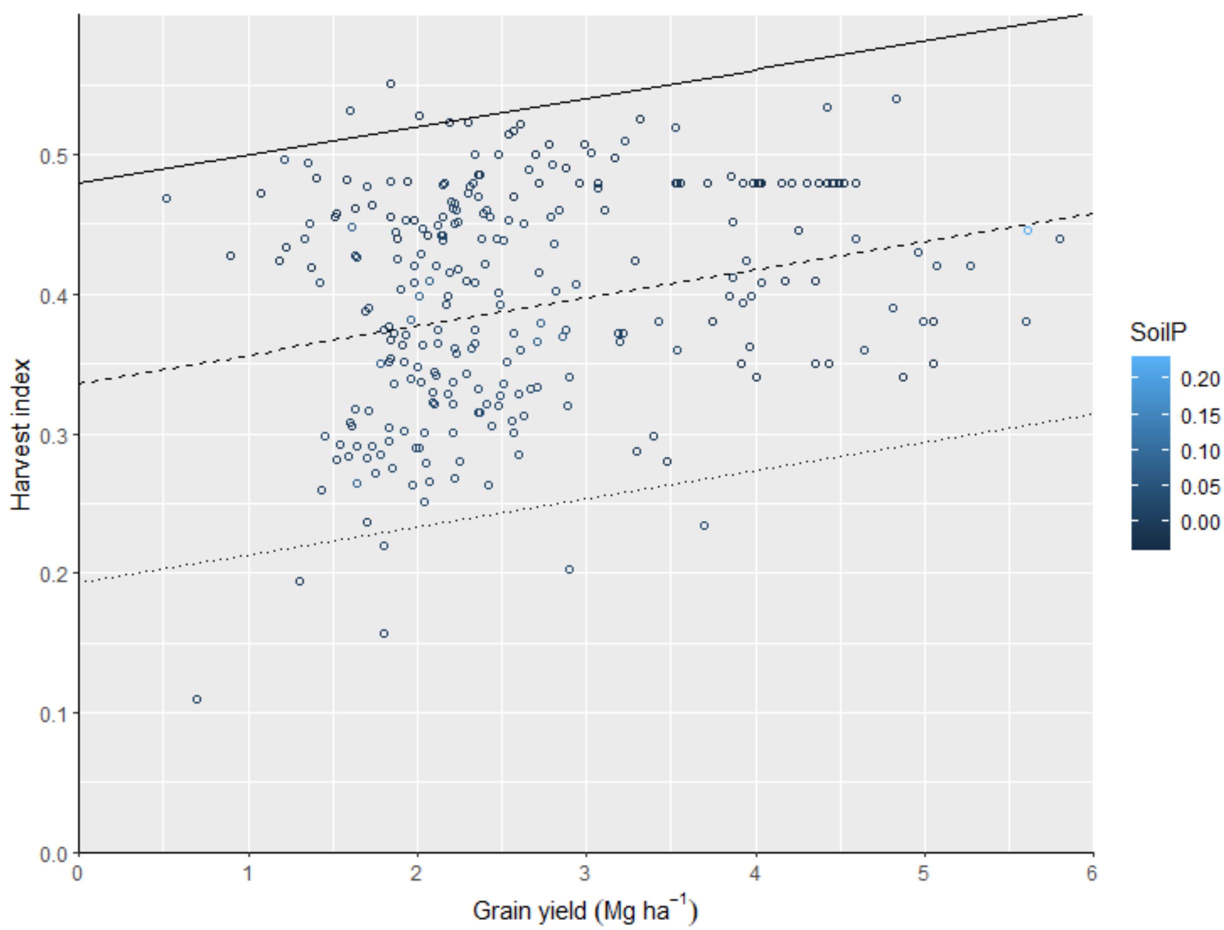

Figure 5: Soybean harvest index in relation to grain yield (GY) and cumulative soil phosphorus budget (SoilP). Regression lines show linear mixed-effects model with GY and SoilP as fixed effects, UN sub-region as random effect and years of data x replicates as weight, at 0.02 kg P m^{-2} (solid line), 0.04 kg P m^{-2} (dashed line) and 0.06 kg P m^{-2} (dotted line) SoilP.

Table 2
Akaike Information Criterion (AIC) and Nakagawa's conditional R^2 values for linear mixed-effects models explaining variation in soybean harvest index (HI). Models in bold were chosen as predictor models.

Model name	Model equation ("Region" was used as a covariate in all equations)	AIC	R^2	Number of observations
H1_1	HI ~ GY	-624	0.09	265
H1_2	HI ~ GY + $(GY)^2$	-631	0.12	265
H1_3	HI ~ FN	-523	0.14	222
H1_4	HI ~ FP	-612	0.12	283
H1_5	**HI ~ FK**	**-465**	**0.19**	**204**
H1_6	**HI ~ SoilP**	**-876**	**0.96**	**340**
H1_7	HI ~ SoilK	-765	0.89	340
H1_8	HI ~ FN + SoilP	-609	0.87	222
H1_9	HI ~ GY + SoilP	-707	0.96	265
H1_10	HI ~ GY + FN	-443	0.20	162
H1_11	**HI ~ GY + FN + FP**	**-404**	**0.34**	**143**
H1_12	HI ~ GY + FN + FP + FK	-343	0.33	125

Variables explaining variation in grain N

A positive trend is observed in which grain N increases with grain yield and FN (Figure 6a). Higher SoilP values, however, are consistent with lower grain N values (Figure 6b). The R^2 value of the linear mixed-effects model including GY and SoilP as fixed effects (Nakagawa's $R^2 = 0.27$) is higher than the model including grain yield and FN (Nakagawa's $R^2 = 0.23$) as fixed effects (Table 3).

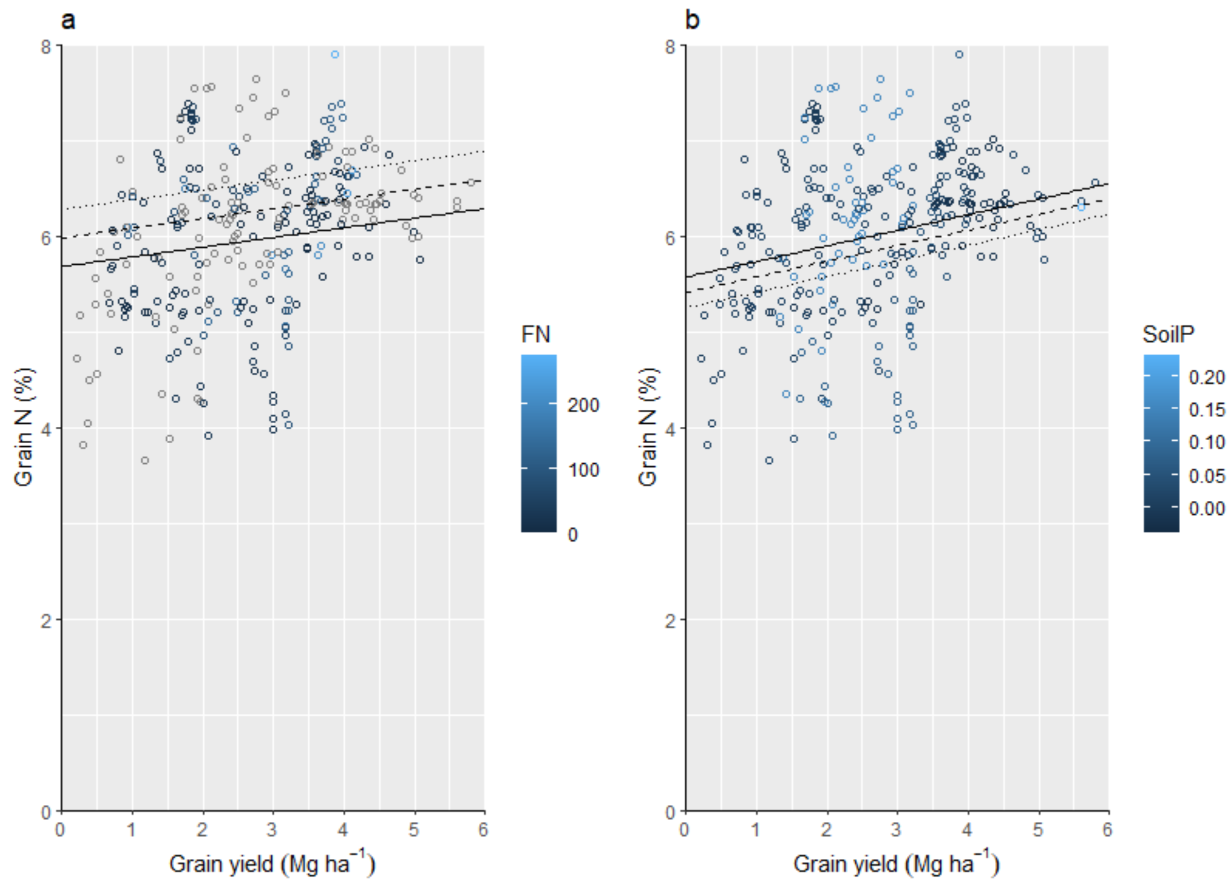

Figure 6: **(a)** Soybean grain N in relation to grain yield (GY) and elemental nitrogen fertiliser application rate (FN). Regression lines show linear mixed-effects model with GY and FN as fixed effects, UN sub-region as random effect and years of data x replicates as weight, at 0 kg N ha^{-1} (solid line), 100 kg N ha^{-1} (dashed line) and 200 kg N ha^{-1} (dotted line) FN. **(b)** Soybean grain N in relation to grain yield (GY) and cumulative soil phosphorus budget (SoilP). Regression lines show linear mixed-effects model with GY and SoilP as fixed effects, UN sub-region as random effect and years of data x replicates as weight, at 0.1 kg P m^{-2} (solid line), 0.3 kg P m^{-2} (dashed line) and 0.5 kg P m^{-2} (dotted line) SoilP.

The linear mixed-effects model with the highest Nakagawa's R^2 value of 0.27 was H2_10, which included FN and SoilP as fixed effects. Model H2_2 had a Nakagawa's R^2 value of 0.26 and a lower AIC compared to H2_10 (Table 3).

Table 3
Akaike Information Criterion (AIC) and Nakagawa's conditional R^2 values for linear mixed-effects models explaining variation in soybean grain N (kg N ha^{-1}). Models in bold were chosen as predictor models.

Model name	Model equation ("Region" was used as a covariate in all equations)	AIC	R^2	Number of observations
H2_1	**Grain N ~ GY**	**702**	**0.18**	**306**
H2_2	**Grain N ~ FN**	**641**	**0.26**	**311**
H2_3	Grain N ~ FP	673	0.20	308
H2_4	Grain N ~ FK	453	0.21	216
H2_5	**Grain N ~ SoilP**	**1168**	**0.15**	**495**
H2_6	Grain N ~ SoilK	1168	0.15	495
H2_7	Grain N ~ GY + FN	426	0.22	189
H2_8	Grain N ~ FN + FP	549	0.23	253
H2_9	Grain N ~ FN + FP + FK	394	0.24	184
H2_10	Grain N ~ FN + SoilP	643	0.27	311
H2_11	Grain N ~ GY + SoilP	704	0.18	306
H2_12	Grain N ~ GY + FN + SoilP	428	0.22	189

Variables explaining variation in residue N

A positive trend was observed between residue N, GY, and FN (Figure 7a). The Nakagawa's R^2 of H3_10, the model with GY and FN as fixed effects, was 0.81 (Table 4). Similarly, a positive trend was also observed for residue N, GY and SoilP. However, lower SoilP values were consistent with higher residue N (Figure 7b). The linear mixed-effects model with GY and SoilP as fixed effects had a Nakagawa's R^2 of 0.73 (Table 4). Model H3_11 and H3_12 had the highest Nakagawa's R^2 value of 0.81. These models both included GY and FN as fixed effects. In addition, H3_12 also included SoilP as a fixed effect.

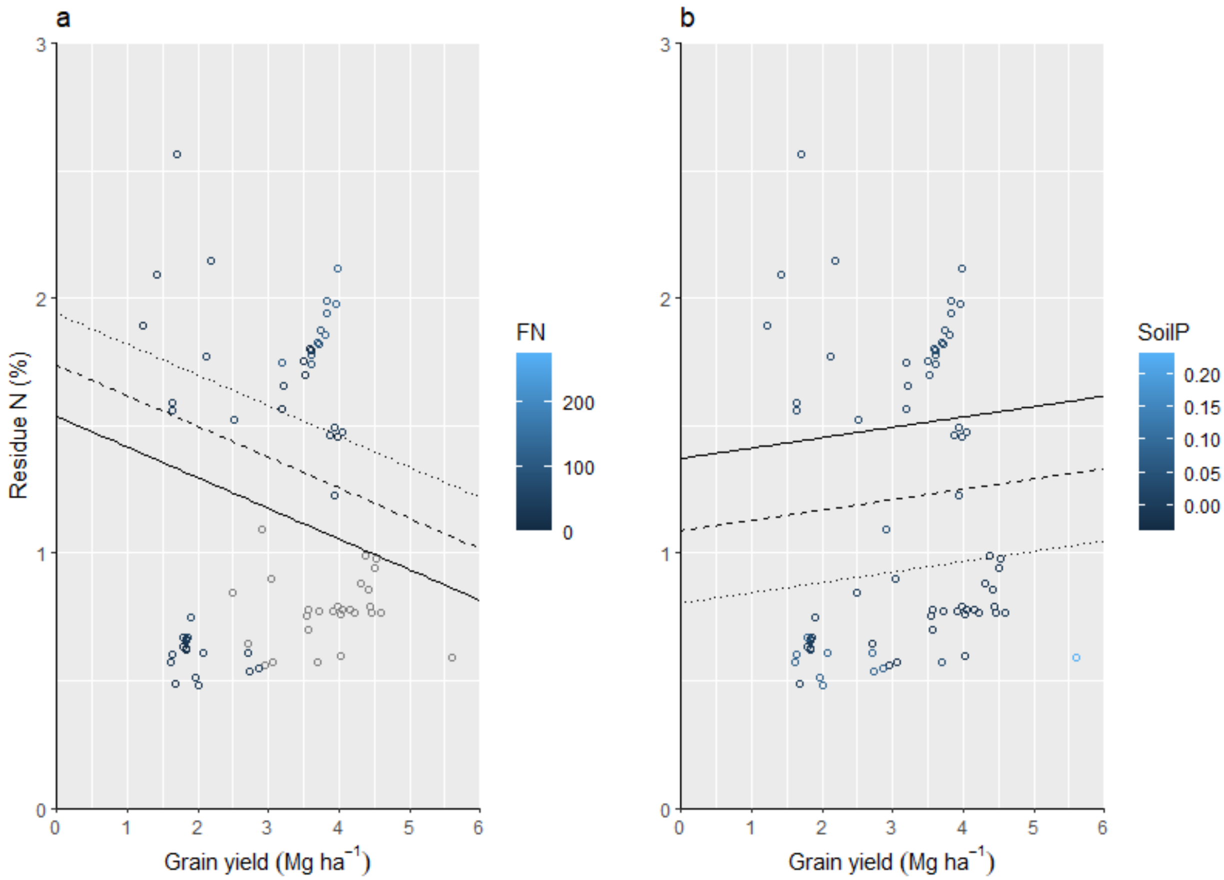

Figure 7: **(a)** Soybean residue N in relation to grain yield (GY) and elemental nitrogen fertiliser application rate (FN). Regression lines show linear mixed-effects model with GY and FN as fixed effects, UN sub-region as random effect and years of data x replicates as weight, at 0 kg N ha^{-1} (solid line), 100 kg N ha^{-1} (dashed line) and 200 kg N ha^{-1} (dotted line) FN. **(b)** Soybean residue N in relation to grain yield (GY) and cumulative soil phosphorus budget (SoilP). Regression lines show linear mixed-effects model with GY and SoilP as fixed effects, UN sub-region as random effect and years of data x replicates as weight, at -0.1 kg P m^{-2} (solid line), 0 kg P m^{-2} (dashed line) and 0.1 kg P m^{-2} (dotted line) SoilP.

Table 4
Akaike Information Criterion (AIC) and Nakagawa's conditional R^2 values for linear mixed-effects models explaining variation in soybean residue N (kg N ha^{-1}). Models in bold were chosen as predictor models.

Model name	Model equation ("Region" was used as a covariate in all equations)	AIC	R^2	Number of observations
H3_1	**Residue N ~ GY**	**-17**	**0.75**	**75**
H3_2	**Residue N ~ FN**	**44**	**0.49**	**55**
H3_3	Residue N ~ FP	75	0.27	62
H3_4	Residue N ~ FK	55	0.38	52
H3_5	Residue N ~ SoilP	87	0.54	91
H3_6	Residue N ~ SoilK	90	0.44	91
H3_7	**Residue N ~ SoilP + SoilK**	**86**	**0.57**	**91**
H3_8	Residue N ~ FN + FP	42	0.40	32
H3_9	Residue N ~ GY + SoilP	-16	0.73	75
H3_10	Residue N ~ GY + SoilP + SoilK	-16	0.71	75
H3_11	Residue N ~ GY + FN	-8	0.81	47
H3_12	Residue N ~ GY + FN + SoilP	-7	0.81	47

Variables explaining variation in NDFA

There is a strong correlation between NDFAkg, GY and FK (Figure 8a). Linear mixed-effects model H4_8, which includes GY and FK as fixed effects has a Nakagawa's R^2 of 0.71. Model H4_11 with GY, FP and SoilK as fixed effect had the highest Nakagawa's R^2 of 0.94 (Table 5).

Linear mixed-effects models explaining NDFA% generally show lower Nakagawa's R^2 values compared to those explaining NDFAkg. Figure 8b shows a weak positive correlation between NDFA%, grain yield and FP. NDFA% was best explained by model H5_10 which included GY, FN, and FP as fixed effects and had a Nakagawa's R^2 of 0.20 (Table 6).

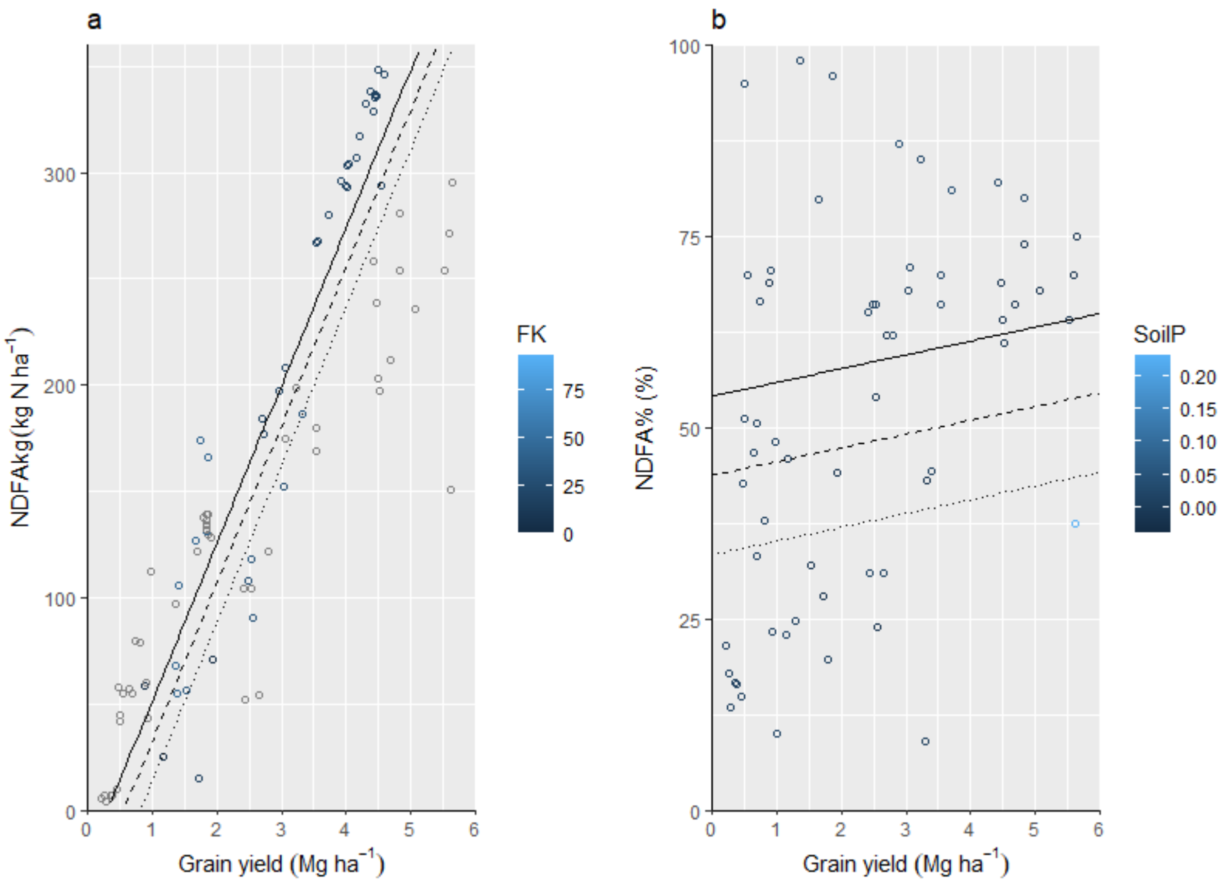

Figure 8: **(a)** Soybean NDFAkg in relation to grain yield (GY) and elemental nitrogen fertiliser application rate (FN). Regression lines show linear mixed-effects model with GY and FK as fixed effects, UN sub-region as random effect and years of data x replicates as weight, at 0 kg K ha^{-1} (solid line), 50 kg K ha^{-1} (dashed line) and 100 kg K ha^{-1} (dotted line) FN. **(b)** Soybean NDFA% in relation to grain yield (GY) and cumulative soil phosphorus budget (SoilP). Regression lines show linear mixed-effects model with GY and SoilP as fixed effects, UN sub-region as random effect and years of data x replicates as weight, at 0 kg P m^{-2} (solid line), 0.2 kg P m^{-2} (dashed line) and 0.4 kg P m^{-2} (dotted line) SoilP.

Table 5

Akaike Information Criterion (AIC) and Nakagawa's conditional R^2 values for linear mixed-effects models explaining variation in soybean NDFAkg (kg N ha^{-1}).

Model name	Model equation ("Region" was used as a covariate in all equations)	AIC	R^2	Number of observations
H4_1	NDFAkg ~ GY	1241	0.51	118
H4_2	NDFAkg ~ FN	951	0.12	91
H4_3	NDFAkg ~ FP	1578	0.33	143
H4_4	NDFAkg ~ FK	865	0.42	79
H4_5	NDFAkg ~ SoilP	3100	0.24	277
H4_6	NDFAkg ~ SoilK	3093	0.36	277
H4_7	NDFAkg ~ GY + FP	723	0.69	71
H4_8	NDFAkg ~ GY + FK	520	0.71	52
H4_9	NDFAkg ~ GY + FP + FK	520	0.72	52
H4_10	NDFAkg ~ GY + SoilK	1242	0.48	118
H4_11	NDFAkg ~ GY + FP + SoilK	709	0.81	71
H4_12	NDFAkg ~ GY + FK + SoilP	521	0.70	52

Table 6

Akaike Information Criterion (AIC) and Nakagawa's conditional R^2 values for linear mixed-effects models explaining variation in soybean NDFA% (%).

Model name	Model equation ("Region" was used as a covariate in all equations)	AIC	R^2	Number of observations
H5_1	NDFA% ~ GY	778	0.09	86
H5_2	NDFA% ~ FN	831	0.12	98
H5_3	NDFA% ~ FP	985	0.11	117
H5_4	NDFA% ~ FK	495	0.10	58
H5_5	NDFA% ~ FN + FP	708	0.13	84
H5_6	NDFA% ~ SoilP	2450	0.02	276
H5_7	NDFA% ~ SoilK	2451	0.02	276
H5_8	NDFA% ~ FN + FP + SoilK	698	0.12	84
H5_9	NDFA% ~ GY + FP	351	0.06	40
H5_10	NDFA% ~ GY + FN + FP	227	0.19	26
H5_11	NDFA% ~ GY + SoilP	780	0.09	86
H5_12	NDFA% ~ GY + SoilK	779	0.10	86

3.3 Prediction Accuracies

Prediction of the response variables harvest index, grain N and residue N with random forest regression was more accurate compared to prediction with the best linear mixed-effects predictor model (Figure 9). The greatest absolute difference in accuracy between prediction with linear mixed-effects models and random forest regression was observed for grain N. The difference between the R^2 values of the linear mixed-effects model (R^2 = 0.75) and random forest (R^2 = 0.83) for grain N was 0.08. In terms of percentages, the difference between prediction accuracies of random forest regression (R^2 = 0.038) and linear mixed-effects models (R^2 = 0.0097) was largest for residue N. A smaller difference between prediction with linear mixed-effects models and random forest regression was found for harvest index. Harvest index prediction was more accurate using random forest regression (R^2 = 0.96) than prediction with linear mixed-effects models (R^2 = 0.92).

In predicting harvest index, grain N and residue N using random forest regression, region is the most important predicting variable. Region has a variable importance percentage of over 65% for all three response variables (Figure 10).

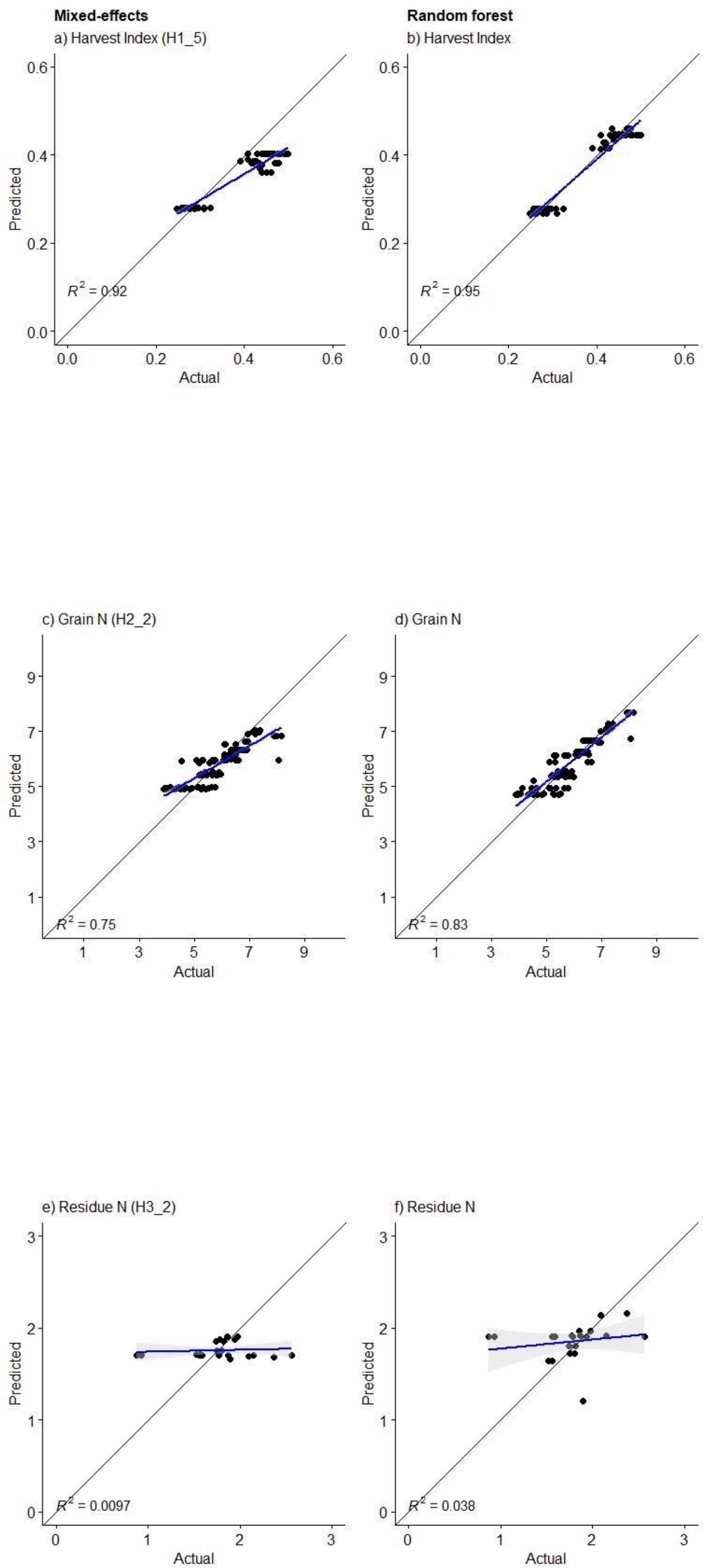

Figure 9: Linear regression of predicted versus actual soybean harvest index (HI), crop product nitrogen concentration (CPN, as % of dry matter yield) and crop residue nitrogen concentration (CRN, as % of dry matter yield), comparing linear mixed-effect models (left column) with random forest (right column).

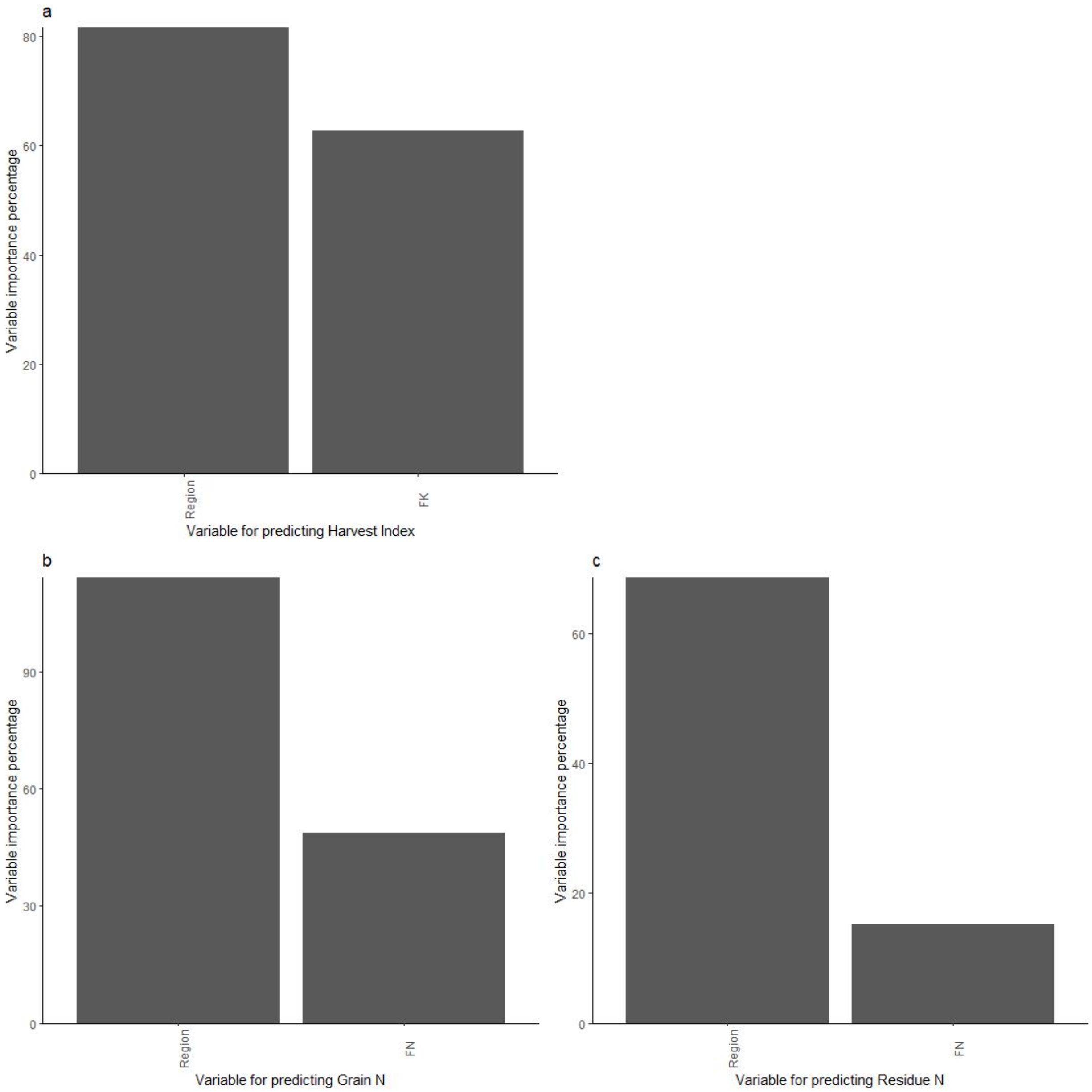

Figure 10: Importance of variables for prediction with random forest regression for **(a)** Harvest Index, **(b)** Grain N and **(c)** Residue N.

4. Discussion

4.1 Data Variability

The variability of grain N and residue N within the dataset exceeded the minimum and maximum limits of nitrogen concentration reported in an article by Nijhof (1987). Nijhof (1987) reported grain N minimum and maximum ranges for soybeans of 4.60% to 7.60% respectively and minimum and maximum residue N ranges from 0.40% to 1.70% respectively. In comparison, grain N within the present dataset ranged between 3.61% and 8.16%, whilst residue N ranged from 0.11% to 2.64%. Although comparable, the extremes go beyond the ranges of Nijhof (1987). This could be because Nijhof (1987) excluded 10% of the most extreme values as outliers and based the minimum and maximum soybean nitrogen concentrations on data from 30 articles. The present study included data from 84 different studies from 31 countries, which could explain the larger variability of harvest index. Collecting more data and re-analysing the data variability of these response variables may allow for a clearer exclusion of outliers, other than on statistical grounds.

Harvest index ranged from 0.11 to 0.55 with a mean of 0.39, within the dataset. However, the interquartile range was much smaller, ranging between 0.32 and 0.47. Unkovich et al. (2010) reported Australian harvest index minimum and maximum ranges for several leguminous crops, including faba bean (0.11-0.58, N = 48), chickpea (0.06-0.55, N = 188) and field pea (0.06-0.58, N = 185). These values suggest that the range found within the present dataset is not uncommon for legumes. The large minimum and maximum range found in the dataset compared to the interquartile range could also be a result of differences in methods used to calculate harvest index between articles from which these values derive. Articles may have calculated harvest index including different plant components and reported measurements at varying moisture contents. Moreover, mature soybean plants regularly drop leaves, and it is often unclear whether articles have included this biomass in their calculations of total aboveground biomass (Herridge et al., 2022). These factors could have contributed to the variability in harvest index found within the dataset.

NDFA% ranged from 9% to 98% within the dataset, with a mean of 56%. NDFAkg ranged between 3 kg ha^{-1} and 354 kg ha^{-1}, with a mean value of 139 kg N ha^{-1}. Comparable ranges were found by Schipanski et al. (2010) which reports NDFAkg values between 40 kg N ha^{-1} and 224 kg N ha^{-1} and NDFA% values ranging from 36% to 82%. The trials in this study were treated with inoculant containing nitrogen fixating bacterial strains, which could explain the higher lower limit reported by Schipanski et al. (2010) compared to the range within this dataset. Furthermore, a review article by Salvagiotti et al. (2008) reports a similar mean value for NDFAkg of 111 kg N ha^{-1} and a minimum and maximum range between 0 and 337 kg N ha^{-1}. The same article reports an NDFA% range of 0 to 98% with a mean of 52% (Salvagiotti et al., 2008). The high variability in NDFA% and NDFAkg found in the present dataset are comparable to those found in literature.

4.2 Explanatory Variables

Variability in harvest index was best explained by the linear mixed-effects model with SoilP as a fixed effect. There was a negative relationship between harvest index and SoilP; higher SoilP values were consistent with lower harvest indices. A similar relationship was found between harvest index and nitrogen fertiliser application, in which higher nitrogen fertiliser application rates were consistent with lower harvest indices. A possible explanation of this observed trend could be that high nitrogen availability leads to increased vegetative growth, resulting in a lower harvest index.

Variability in grain N was best explained by the linear mixed-effects model with FN and SoilP as fixed effects (Nakagawa's $R^2 = 0.27$). Although, SoilP did not add much to the explanatory power of this model. The model with FN as sole fixed effect had a Nakagawa's R^2 of 0.26 and a lower AIC compared to the model including both FN and SoilP. Higher nitrogen fertiliser application rates were consistent with higher grain nitrogen concentrations. High nitrogen availability can lead to nitrogen accumulation and therefore higher nitrogen grain concentrations (Divito et al., 2016; Salvagiotti et al., 2008). On the contrary, higher

SoilP values were consistent with lower grain nitrogen concentrations. This can be explained by high potassium availability leading to nitrogen dilution (Janssen et al., 1990).

Residue N showed the same relationship to nitrogen fertiliser application and SoilP as grain N. Residue nitrogen concentration increase with nitrogen fertiliser application rates and decrease with SoilP. Variability in residue N was best explained by the linear mixed-effects model with GY and FN as fixed effects (Nakagawa's R^2 = 0.81). Although the model with grain yield as sole fixed effect had a comparable Nakagawa's R^2 of 0.75 and a higher number of observations.

NDFAkg was best explained by a linear mixed-effects model which included the variables grain yield, FP and SoilK as fixed effects (Nakagawa's R^2 = 0.81). In all models, FK and FP were shown to be better explanatory variables for NDFAkg than FN. NDFAkg has a strong positive relationship with GY. A negative relationship is found between NDFAkg and fertiliser application rates.

NDFA% variability was best explained by grain yield, FN, and FP as fixed effects in a linear mixed-effects model (Nakagawa's R^2 = 0.19), although this model only has 26 data observations. Unlike NDFAkg, NDFA% variability was better explained by FN than FP or FK as fixed effects in linear mixed-effects models.

4.3 Prediction

Predictor models were chosen subjectively based on the number of observations within the dataset, AIC and Nakagawa's R^2 of the linear mixed-effects models. In some cases, predictor models were chosen which therefore did not have the highest Nakagawa R^2 or lowest AIC. In the case of residue N especially, there were very few observations, making it difficult to assess models on their prediction accuracy. The lack of observations was also the reason no predictor models were tested for NDFAkg and NDFA%.

Prediction with random forest regression was more accurate compared to prediction with linear mixed-effects models across all predicted response variables. Accurate predictions were made for harvest index and grain nitrogen concentration, with R^2 values of 0.83 and 0.95 for linear regression on actual versus predicted plots. Prediction accuracies for residue N were low using linear mixed-effects model (R^2 = 0.0097) as for random forest regression (R^2 = 0.038). The same input variables were used for random forest as the best predictor linear mixed-effects model, for a fair comparison. Using more variables in random forest regression and optimising its prediction may increase its prediction accuracy. This was not done within this study, as the objective was to compare it to linear mixed-effects model predictions. The study by Ludemann et al. (2022) estimating maize harvest index and nitrogen concentrations of grain and residue, also found that random forest regression prediction was more accurate compared to linear mixed-effects models. Ludemann et al. (2022) reported R^2 values for actual versus predicted plots for harvest index, and nitrogen concentration of grain and residue of 0.58, 0.68 and 0.56 respectively.

Random forest outperformed linear mixed-effects models in terms of prediction accuracies. It is expected that machine learning will be used more often in future research. Although it does increase the necessity for standardised, high-quality databases. Machine learning can help get the most out of data collected around the world. A downside of random forest regression is that it is more difficult to interpret how variables relate to one another. Coefficients in linear mixed-effects modelling give insight on the strength of a relationship and its direction. Therefore, linear mixed-effects models allow for a better understanding of the relationships between variables and thus give a better insight into the mechanisms behind the results. Which of the two methods is better, depends on the goal of the research.

4.4 Limitations

The greatest limitation of this study is the lack of data points, especially for response variables residue N, NDFAkg and NDFA%. Few data points means that potential outliers are given more weight, which could influence the outcome of the results. Both in linear mixed-effects modelling and random forest regression, region was an important variable despite uneven data distribution across regions. More data points across regions could further improve explanatory models and prediction accuracies.

This study looked at a limited selection of explanatory variables, due to findings from previous research and the wide availability of data points for variables found within the dataset. There are more variables to be explored which could potentially contribute to explaining variation and prediction of these response variables.

Lastly, the struggle of standardising data from literature is not without bias or human error. Unfortunately, there is still some ambiguity in articles about definition, units and how measurements are taken which makes it a challenge to correctly standardise data. For example, ambiguity in whether grain yield is reported in kilograms of dry matter or fresh weight for soybeans with a harvest index of 0.5 could results in about a 7.6% difference in harvest index assuming a moisture proportion of 0.86. This highlights the importance of unambiguous reporting in scientific literature and transparent data collection in meta-analyses.

4.5 Improvements and Further Research

To improve linear mixed-effects models' explanatory power and prediction accuracies, collecting more data across all regions would deliver more robust results, especially for residue N, NDFAkg and NDFA%. Further research could focus on including explanatory variables to the dataset such as crop variety, climate, growing season length and yield potential variables. An interesting variable for explaining NDFA could include information on inoculation with nitrogen fixing bacterial strains. In the context of soybean cultivation, research into biological nitrogen fixation could further improve nutrient management plans. Creating conditions suitable for nitrogen fixation and looking into inoculation rather than fertilisation may improve yields and reduce nutrient losses further (Sogut, 2006).

5. Conclusions

Creating explanatory models for soybean harvest index, grain N, residue N and NDFA using globally available data helps understand and improve nitrogen use in agriculture. The widely available variables including grain yield, fertiliser application rates, and cumulative soil phosphorus and potassium budgets can be used in linear mixed-effects models to explain these response variables. Furthermore, these variables can be used to accurately predict harvest index and grain N. Random forest regression outperforms linear mixed-effects modelling in terms of prediction accuracy but does not give insight in the relationships between variables. The high Nakagawa's R^2 of explanatory models and accurate prediction for grain N shows that both linear mixed-effects modelling and random forest regression can be important tools to develop nutrient management plans. Expanding the dataset with response variable data and explanatory variables, such as yield potential and variety information, could further improve the understanding and prediction of nitrogen flows in soybean cultivation.

6. References

Adjei-Nsiah, S., Ahiakpa, J. K., Gyan, K., & Kanampiu, F. (2021). Influence of phosphorus fertiliser blends on grain yield, nutrient concentration, and profitability of soyabeans in the southern Guinea Savannah of Ghana. *South African Journal of Plant and Soil*, *38*(5), 381-389. https://doi.org/10.1080/02571862.2021.1959660

Aher, S. B., Lakaria, B. L., Kaleshananda, S., & Singh, A. B. (2022). Yield, nutrient uptake and economics of soybean–wheat cropping system under organic nutrient management in Central India. *Journal of plant nutrition*, *45*(6), 904-919. https://doi.org/10.1080/01904167.2021.1994600

Amadou, A., Song, X., Huang, S., Song, A., Tang, Z., Dong, W., Zhao, S., Zhang, B., Yi, K., & Fan, F. (2021). Effects of long-term organic amendment on the fertility of soil, nodulation, yield, and seed quality of soybean in a soybean-wheat rotation system. *Journal of Soils and Sediments*, *21*(3), 1385-1394. https://doi.org/10.1007/s11368-021-02887-1

Amiri, H., Ghalavand, A., & Mokhtassi-Bidgoli, A. (2021). Growth, Seed Yield and Quality of Soybean as Affected by Integrated Fertilizer Managements and Zeolite Application. *Communications in Soil Science and Plant Analysis*, *52*(15), 1834-1851. https://doi.org/10.1080/00103624.2021.1900222

Andriani, J., Andrade, F., Suero, E., & Dardanelli, J. (1991). Water deficits during reproductive growth of soybeans. I. Their effects on dry matter accumulation, seed yield and its components. *Agronomie*, *11*(9), 737-746. https://doi.org/10.1051/agro:19910904

Asres, A., & Tiruneh, S. (Preprint (2020)). Evaluation of Biomass Yield and Nutritional Composition of Soybean (Glycine Max (L.) Merrill) Varieties Grown in Low Land Areas of Eastern Amhara, Ethiopia. http://scholar.google.com/scholar_url?url=https://www.preprints.org/manuscript/202011.0271/download/final_file&hl=en&sa=X&d=39148358455228125&ei=m_2uX5HgHMCI6rQP29KioAM&scisig=AAGBfm2y4MpnBE_hsmyeucGLNSWH_rTF8A&nossl=1&oi=scholaralrt&html=

Basal, O., & Szabó, A. (2020). Sole and combined effects of drought and phosphorus application on soybean. *Journal of Central European Agriculture*, *21*(4), 796-808.

Bellaloui, N., Ebelhar, M. W., Gillen, A. M., Fisher, D. K., Abbas, H. K., Mengistu, A., Reddy, K. N., & Paris, R. L. (2011). Soybean seed protein, oil, and fatty acids are altered by S and S+ N fertilizers under irrigated or non-irrigated environments. *Agricultural Sciences*, *2*(4), 465-476. https://digitalcommons.cedarville.edu/cgi/viewcontent.cgi?article=1304&context=science_and_mathematics_publications

Bender, R. R., Haegele, J. W., & Below, F. E. (2015). Nutrient uptake, partitioning, and remobilization in modern soybean varieties. *Agronomy Journal*, *107*(2), 563-573. https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2134/agronj14.0435

Bhangoo, M., & Albritton, D. (1972). Effect of Fertilizer Nitrogen, Phosphorus, and Potassium on Yield and Nutrient Content of Lee Soybeans 1. *Agronomy Journal*, *64*(6), 743-746. https://acsess.onlinelibrary.wiley.com/doi/abs/10.2134/agronj1972.00021962006400060010x

Boddey, R. M., Urquiaga, S., Neves, M. C. P., & Peres, J. (1990). Quantification of the contribution of N2 fixation to field-grown grain legumes—a strategy for the practical application of the 15N isotope dilution technique. *Soil Biology and Biochemistry*, *22*(5), 649-655.

Bortolon, L., Bortolon, E. S. O., de Camargo, F. P., Seraglio, N. A., Lima, A. d. O., Rocha, P. H. F., de Souza, J. P., Sousa, W. C., Tomazzi, M., & Lago, B. C. (2018). Yield and nutrient uptake of soybean cultivars under intensive cropping systems. *Embrapa Pesca e Aquicultura-Artigo em periódico indexado (ALICE)*, *10*(12). https://doi.org/10.5539/jas.v10n12p344

Bos, J. F. F. P., ten Berge, H. F. M., Verhagen, J., & van Ittersum, M. K. (2017). Trade-offs in soil fertility management on arable farms. *Agricultural Systems*, *157*, 292-302. https://doi.org/https://doi.org/10.1016/j.agsy.2016.09.013

Cafaro La Menza, N., Monzon, J. P., Lindquist, J. L., Arkebauer, T. J., Knops, J. M. H., Unkovich, M., Specht, J. E., & Grassini, P. (2020). Insufficient nitrogen supply from symbiotic fixation reduces

seasonal crop growth and nitrogen mobilization to seed in highly productive soybean crops [https://doi.org/10.1111/pce.13804]. *Plant, Cell & Environment*, *43*(8), 1958-1972. https://doi.org/https://doi.org/10.1111/pce.13804
Cafaro La Menza, N., Monzon, J. P., Specht, J. E., Lindquist, J. L., Arkebauer, T. J., Graef, G., & Grassini, P. (2019). Nitrogen limitation in high-yield soybean: Seed yield, N accumulation, and N-use efficiency. *Field Crops Research*, *237*, 74-81. https://www.sciencedirect.com/science/article/pii/S0378429019302916?via%3Dihub
Cannon, K., McClure, M., Yin, X., & Sams, C. (2021). Corn and soybean response to sulfur fertilizer in West Tennessee. *Crop, Forage & Turfgrass Management*, e20092.
Chețan, F., Chețan, C., Bogdan, I., Pop, A. I., Moraru, P. I., & Rusu, T. (2021). The Effects of Management (Tillage, Fertilization, Plant Density) on Soybean Yield and Quality in a Three-Year Experiment under Transylvanian Plain Climate Conditions. *Land*, *10*(200), 200.
Coale, F. J., Meisinger, J. J., & Wiebold, W. J. (1985). Effects of plant breeding and selection on yields and nitrogen fixation in soybeans under two soil nitrogen regimes. *Plant and Soil*, *86*(3), 357-367. https://doi.org/10.1007/BF02145456
Deibert, E. J., Bijeriego, M., & Olson, R. A. (1979). Utilization of 15N Fertilizer by Nodulating and Non-Nodulating Soybean Isolines1. *Agronomy Journal*, *71*(5), 717-723. https://doi.org/https://doi.org/10.2134/agronj1979.00021962007100050006x
Di Ciocco, C., Coviella, C., Penón, E., Díaz-Zorita, M., & López, S. (2008). Biological fixation of nitrogen and N balance in soybean crops in the pampas region. *Spanish Journal of Agricultural Research*, *6*(1), 114-119. https://www.cabdirect.org/cabdirect/FullTextPDF/2008/20083107234.pdf
Divito, G. A., Echeverría, H. E., Andrade, F. H., & Sadras, V. O. (2016). Soybean shows an attenuated nitrogen dilution curve irrespective of maturity group and sowing date. *Field Crops Research*, *186*, 1-9.
Domingos, C. d. S., Besen, M. R., Esper Neto, M., Costa, E. J. O., Scapim, C. A., Inoue, T. T., Batista, M. A., & Braccini, A. L. (2021). Can calcium and boron leaf application increase soybean yield and seed quality? *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, *71*(3), 171-181. https://doi.org/10.1080/09064710.2020.1869818
Dragicevic, V., Stoiljkovic, M., Brankov, M., Tolimir, M., Tabaković, M., Dodevska, M. S., & Simić, M. (2022). Status of Essential Elements in Soil and Grain of Organically Produced Maize, Spelt, and Soybean. *Agriculture*, *12*(5), 702. https://doi.org/10.3390/agriculture12050702
Eliçin, A., Öztürk, F., KizilgeÇİ, F., Koca, Y., & Iqbal, M. A. (2021). Coherent fertilization regimes boost productivity and nutritional quality of soybean (Glycine max. L. Merrill). *Fresenius Environmental Bulletin*, *30*, 13323-13330. https://www.researchgate.net/profile/Ferhat-Kizilgeci/publication/356287799_COHERENT_FERTILIZATION_REGIMES_BOOST_PRODUCTIVITY_AND_NUTRITIONAL_QUALITY_OF_SOYBEAN_Glycine_max_L_Merrill/links/61b33638590a0b7ed6352d0a/COHERENT-FERTILIZATION-REGIMES-BOOST-PRODUCTIVITY-AND-NUTRITIONAL-QUALITY-OF-SOYBEAN-Glycine-max-L-Merrill.pdf?origin=publication_detail
Emmerson, M., Morales, M. B., Oñate, J. J., Batáry, P., Berendse, F., Liira, J., Aavik, T., Guerrero, I., Bommarco, R., Eggers, S., Pärt, T., Tscharntke, T., Weisser, W., Clement, L., & Bengtsson, J. (2016). Chapter Two - How Agricultural Intensification Affects Biodiversity and Ecosystem Services. In A. J. Dumbrell, R. L. Kordas, & G. Woodward (Eds.), *Advances in Ecological Research* (Vol. 55, pp. 43-97). Academic Press. https://doi.org/https://doi.org/10.1016/bs.aecr.2016.08.005
Engy, S. A. Z., Mary, Seham, & Salwa. (2020). Impact of foliar and soil fertilization on productivity and quality of some soybean cultivars under calcareous soil conditions. *Zagazig Journal of Agricultural Research*, *47*(4), 867-881. https://zjar.journals.ekb.eg/article_110314_1c016a744255b19f5e919370d056722d.pdf
Erbil, E., Ucak, A. B., Tas, T., Karakus, M., & Yetkin, C. (2020). Effect of foliar application of iron (Fe) and molibydenium (Mo) on yield, protein and nodulation in soybean under agro climatic conditions

of Sanliurfa. *Fresenius Environmental Bulletin*, *29*(4), 3033-3039. https://www.researchgate.net/profile/Timucin_Tas/publication/344237368_Effect_of_Foliar_Application_of_Iron_Fe_and_Molibydenium_Mo_on_Yield_Protein_and_Nodulation_in_Soybean_Under_Agro_Climatic_Conditions_of_Sanliurfa/links/5f5f4d094585154dbbd03262/Effect-of-Foliar-Application-of-Iron-Fe-and-Molibydenium-Mo-on-Yield-Protein-and-Nodulation-in-Soybean-Under-Agro-Climatic-Conditions-of-Sanliurfa.pdf

FAOSTAT. (2022). https://www.fao.org/faostat/en/#data

Gan, Y., Stulen, I., Posthumus, F., van Keulen, H., & Kuiper, P. (2002). Effects of N management on growth, N2 fixation and yield of soybean. *Nutrient Cycling in Agroecosystems*, *62*(2), 163-174. https://doi.org/10.1023/A:1015528132642

Gan, Y., Stulen, I., van Keulen, H., & Kuiper, P. J. C. (2003). Effect of N fertilizer top-dressing at various reproductive stages on growth, N2 fixation and yield of three soybean (Glycine max (L.) Merr.) genotypes. *Field Crops Research*, *80*(2), 147-155. https://doi.org/https://doi.org/10.1016/S0378-4290(02)00171-5

Gelfand, I., & Robertson, G. P. (2015). A reassessment of the contribution of soybean biological nitrogen fixation to reactive N in the environment. *Biogeochemistry*, *123*(1), 175-184. https://doi.org/10.1007/s10533-014-0061-4

George, T., Singleton, P. W., & Bohlool, B. B. (1988). Yield, Soil Nitrogen Uptake, and Nitrogen Fixation by Soybean from Four Maturity Groups Grown at Three Elevations. *Agronomy Journal*, *80*(4), 563-567. https://doi.org/10.2134/agronj1988.00021962008000040004x

Ghani, R. A., Kende, Z., Tarnawa, Á., Omar, S., Kassai, M. K., & Jolánkai, M. (2021). The effect of nitrogen application and various means of weed control on grain yield, protein and lipid content in soybean cultivation. *Acta Alimentaria*. https://doi.org/10.1556/066.2021.00095

Guafa, W., Peoples, M. B., Herridge, D. F., & Rerkasem, B. (1993). Nitrogen fixation, growth and yield of soybean grown under saturated soil culture and conventional irrigation. *Field Crops Research*, *32*(3), 257-268. https://doi.org/https://doi.org/10.1016/0378-4290(93)90035-L

Gyogluu, C., Boahen, S. K., & Dakora, F. D. (2016). Response of promiscuous-nodulating soybean (Glycine max L. Merr.) genotypes to Bradyrhizobium inoculation at three field sites in Mozambique. *Symbiosis*, *69*(2), 81-88. https://doi.org/10.1007/s13199-015-0376-5

Ham, G. E., & Caldwell, A. C. (1978). Fertilizer Placement Effects on Soybean Seed Yield, N2 Fixation, and 33P Uptake1 [https://doi.org/10.2134/agronj1978.00021962007000050020x]. *Agronomy Journal*, *70*(5), 779-783. https://doi.org/10.2134/agronj1978.00021962007000050020x

Herridge, D. F., Giller, K. E., Jensen, E. S., & Peoples, M. B. (2022). Quantifying country-to-global scale nitrogen fixation for grain legumes II. Coefficients, templates and estimates for soybean, groundnut and pulses. *Plant and soil*, *474*(1), 1-15. https://doi.org/10.1007/s11104-021-05166-7

Hiep, N. H., Diep, C. N., & Herridge, D. F. (2002). Nitrogen fixation of soybean and groundnut in the Mekong Delta, Vietnam. *Inoculants and nitrogen fixation of legumes in Vietnam. ACIAR Proceedings*.

Iannetta, P. P. M., Young, M., Bachinger, J., Bergkvist, G., Doltra, J., Lopez-Bellido, R. J., Monti, M., Pappa, V. A., Reckling, M., Topp, C. F. E., Walker, R. L., Rees, R. M., Watson, C. A., James, E. K., Squire, G. R., & Begg, G. S. (2016). A Comparative Nitrogen Balance and Productivity Analysis of Legume and Non-legume Supported Cropping Systems: The Potential Role of Biological Nitrogen Fixation [Original Research]. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.01700

Jansone, I., Sterna, V., Stramkale, V., Stramkalis, A., Justs, A., & Zute, S. (2021). Impact of cultivation technologies on soybean production and quality [soybean, feed, food, fertilisation, protein]. *Environment Technology Resources Proceedings of the International Scientific and Practical Conference*, *1*, 7. https://doi.org/10.17770/etr2021vol1.6605

Janssen, B. H., Guiking, F., van der Eijk, D., Smaling, E. M., Wolf, J., & van Reuler, H. (1990). A system for quantitative evaluation of the fertility of tropical soils (QUEFTS). *Geoderma*, *46*(4), 299-318.

Jarecki, W., & Bobrecka-Jamro, D. (2021). Effect of sowing data on the yield and seed quality of soybeans (Glycine Max (L.) Merr). *Journal of Elementology*, *26*(1), 7-18.

Jat, G., Sharma, S. K., Meena, R., Choudhary, R., Choudhary, R., & Yadav, S. K. (2021). Studies on Effect of Zinc Application on Quality and Yield of Soybean (Glycine max L.) under Typic Haplustepts Soil. *Ind. J. Pure App. Biosci*, *9*(1), 188-193. https://doi.org/10.18782/2582-2845.8511

Jefing, Y., Herridge, D. F., Peoples, M. B., & Rerkasem, B. (1992). Effects of N fertilization on N2 fixation and N balances of soybean grown after lowland rice. *Plant and Soil*, *147*(2), 235-242. https://doi.org/10.1007/BF00029075

Jensen, E. S., & Hauggaard-Nielsen, H. (2003). How can increased use of biological N_2 fixation in agriculture benefit the environment? *Plant and soil*, *252*(1), 177-186. https://doi.org/https://doi.org/10.1023/A:1024189029226

Kakabouki, I., Folina, A., Zisi, C., & Karydogianni, S. (2020). Fertilization expression via nitrogen indices in soybean crop under two system tillage. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, *48*(2), 799-813. https://doi.org/10.15835/nbha48211880

Karhale, A. R. (2021). Effect of nutrient management on nutrient uptake, yield and economics of soybean. *Journal of Soils and Crops*, *31*(1), 116-120.

Kubar, M. S., Shar, A. H., Kubar, K. A., Rind, N. A., Ullah, H., Kalhoro, S. A., Wang, C., Feng, M., Gujar, A., Sun, H., Yang, W., El Enshasy, H., Brestic, M., Zivcak, M., Ondrisik, P., Aljuaid, B. S., El-Shehawi, A. M., & Ansari, M. J. (2021). Optimizing nitrogen supply promotes biomass, physiological characteristics and yield components of soybean (Glycine max L. Merr.). *Saudi Journal of Biological Sciences*. https://doi.org/10.1016/j.sjbs.2021.06.073

Kucey, R. M. N., Chaiwanakupt, P., Snitwongse, P., Toomsan, B., Boonkerd, N., Siripaibool, C., Rennie, R. J., Rungrattanakasin, W., & Wadisirisak, P. (1988). Nitrogen fixation (N-15 dilution) with soybeans under Thai field conditions. III. Effects of Bradyrhizobium japonicum strains and herbicides in Northeast Thailand. *The Journal of General and Applied Microbiology*, *34*(3), 243-253.

Landriscini, M. R., Galantini, J. A., Duval, M. E., & Capurro, J. E. (2019). Nitrogen balance in a plant-soil system under different cover crop-soybean cropping in Argentina. *Applied Soil Ecology*, *133*, 124-131. https://doi.org/https://doi.org/10.1016/j.apsoil.2018.10.005

Langeveld, J. W. A., Verhagen, A., Neeteson, J. J., van Keulen, H., Conijn, J. G., Schils, R. L. M., & Oenema, J. (2007c). Evaluating farm performance using agri-environmental indicators: Recent experiences for nitrogen management in The Netherlands. *Journal of Environmental Management*, *82*(3), 363-376. https://doi.org/https://doi.org/10.1016/j.jenvman.2005.11.021

Latifinia, E., & Eisvand, H. R. (2022). Soybean Physiological Properties and Grain Quality Responses to Nutrients, and Predicting Nutrient Deficiency Using Chlorophyll Fluorescence. *Journal of Soil Science and Plant Nutrition*. https://doi.org/10.1007/s42729-022-00785-0

Liu, J., You, L., Amini, M., Obersteiner, M., Herrero, M., Zehnder, A. J., & Yang, H. (2010). A high-resolution assessment on global nitrogen flows in cropland. *Proceedings of the National Academy of Sciences*, *107*(17), 8035-8040. https://doi.org/https://doi.org/10.1073/pnas.0913658107

Lohar, R., Hase, C., & Kapurkar, J. (2020). Applications of Organic and Synthetic Fertilizers for Sustainable Production of Soybean Crop (Glycine max L.) and Improvement in Soil Health. *Research Journal of Agricultural Sciences*, *11*(6), 1332-1342. https://www.researchgate.net/profile/Gh_Akbari/publication/346719530_Evaluation_of_Maize_Grain_Yield_and_Quality_Affected_by_GEM_interactions_Case_study_Pakdasht_area_in_Iran/links/5fcf9a9092851c00f85f0244/Evaluation-of-Maize-Grain-Yield-and-Quality-Affected-by-GEM-interactions-Case-study-Pakdasht-area-in-Iran.pdf

Ludemann, C. I., Hijbeek, R., Loon, M. v., Murrell, S. T., Dobermann, A., & Ittersum, M. K. v. (2023). *Published field experimental data of crop yields, nutrient concentrations, and harvest indices from around the world*. https://doi.org/https://doi.org/10.5061/dryad.djh9w0w4m

Ludemann, C. I., Hijbeek, R., van Loon, M. P., Murrell, T. S., Dobermann, A., & van Ittersum, M. K. (2022). Estimating maize harvest index and nitrogen concentrations in grain and residue using

globally available data. *Field Crops Research*, *284*, 108578. https://doi.org/https://doi.org/10.1016/j.fcr.2022.108578
Ma, Z., Luo, M., Wang, Q., Li, C., Guo, L., Wu, S., Kazery, J., Yan, Y., Zhai, L., & Liu, H. (2022). Long-Term Agronomic Practices in Reducing Nitrogen and Phosphorus Loss from Intensive Croplands. *ACS Earth and Space Chemistry*.
Machado, F. R., Possenti, J. C., Vismara, E. S., Deuner, C., Favoretto, V. R., & Machado, F. R. (2021). Yield components and nutritional quality of soybean grain as a function of foliar fertilizers application time. *Revista Brasileira de Ciências Agrárias - Brazilian Journal of Agricultural Sciences*, *16*(2), 1-7. https://doi.org/10.5039/agraria.v16i2a8851
Machiani, M. A., Javanmard, A., Morshedloo, M. R., Janmohammadi, M., & Maggi, F. (2021). Funneliformis mosseae Application Improves the Oil Quantity and Quality and Eco-physiological Characteristics of Soybean (Glycine max L.) Under Water Stress Conditions. *Journal of Soil Science and Plant Nutrition*, *21*(4), 3076-3090. https://doi.org/10.1007/s42729-021-00590-1
Mamun, M. A. A., Sarker, U., Mannan, M. A., Rahman, M. M., Karim, M. A., Ercisli, S., Marc, R. A., & Golokhvast, K. S. (2022). Application of Potassium after Waterlogging Improves Quality and Productivity of Soybean Seeds. *Life*, *12*(11), 1816.
Mandić, V., Đorđević, S., Đorđević, N., Bijelić, Z., Krnjaja, V., Petričević, M., & Brankov, M. (2020). Genotype and Sowing Time Effects on Soybean Yield and Quality. *Agriculture*, *10*(11), 502. https://www.mdpi.com/2077-0472/10/11/502
Movalia, J., & Savalia, S. G. (2021). Effect of sources and levels of sulphur on yield and quality of kharif soybean [Glycine max (L.) Merrill]. *the Pharma Innovation*, *10*(6), 142-150.
Munyinda, K., Karamanos, R. E., Legg, J. O., & Sanogho, S. (1988). Nitrogen fixation by soybeans (Glycine max L.) in Zambia. *Plant and Soil*, *109*(1), 57-63. https://doi.org/10.1007/BF02197580
Nassar, K., EL-Shaboury, H. A., & EL-Sonbaty, A. E. (2021). Influence of some potassium humate application methods and mineral fertilization of nitrogen and phosphorus on soybean yield and quality. *Menoufia Journal of Soil Science*, *6*(4), 133-145.
Nijhof, K. (1987). *The concentration of macro-elements in economic products and residues of (sub) tropical field crops*. Centre for World Food Studies.
Oberson, A., Nanzer, S., Bosshard, C., Dubois, D., Mäder, P., & Frossard, E. (2007). Symbiotic N2 fixation by soybean in organic and conventional cropping systems estimated by 15N dilution and 15N natural abundance. *Plant and Soil*, *290*(1), 69-83. https://doi.org/10.1007/s11104-006-9122-3
Ojiem, J. O., Vanlauwe, B., de Ridder, N., & Giller, K. E. (2007). Niche-based assessment of contributions of legumes to the nitrogen economy of Western Kenya smallholder farms. *Plant and Soil*, *292*, 119-135.
Okogun, J. A., Sanginga, N., Abaidoo, R., Dashiell, K. E., & Diels, J. (2005). On-farm Evaluation of Biological Nitrogen Fixation Potential and Grain Yield of Lablab and Two Soybean Varieties in the Northern Guinea Savanna of Nigeria. *Nutrient Cycling in Agroecosystems*, *73*(2), 267-275. https://doi.org/10.1007/s10705-005-3821-7
Peoples, M. B., Gault, R. R., Lean, B., Sykes, J. D., & Brockwell, J. (1995). Nitrogen fixation by soybean in commercial irrigated crops of central and southern New South Wales. *Soil Biology and Biochemistry*, *27*(4), 553-561. https://doi.org/https://doi.org/10.1016/0038-0717(95)98631-W
R Core Team. (2022). *R: A Language and Environment for Statistical Computing* In R Foundation for Statistical Computing. https://www.R-project.org/
Radzka, E., Katarzyna, R., & Wysokinski, A. (2021). Nitrogen Uptake from Different Sources by Soybean Grown at Different sowing densities. *Agronomy Journal*, *11*(720). https://doi.org/10.3390/agronomy11040720
Raj, K. K., Pandey, R., Singh, B., Meena, M., Talukdar, A., & Chakraborty, D. (2021). Response of soybean genotypes to iron limiting stress in calcareous vertisol under ambient and elevated CO2 and temperature conditions. *Journal of Environmental Biology*, *42*(2), 295-301.
Rashmi, I., Kuldeep, K., Ashok, K., Kala, S., Anita, K., Pal, R., & Sao, Y. (2022). Photosynthetic Pigments, Symbiotic Traits, Yield Attributes and Nutrient Uptake of Soybean Crop as Influenced by Soil

Amendments on Degraded Vertisol of Chambal Tablelands. *Legume Research-An International Journal*, *1*, 6. http://arccarticles.s3.amazonaws.com/GalleyProof/Attachment-atgalley-proof-LR-4923-6089603e9087340b9bdcbbb1.pdf

Rennie, R. J., & Dubetz, S. (1984). Effect of Fungicides and Herbicides on Nodulation and N2 Fixation in Soybean Fields Lacking Indigenous Rhizobium japonicum1 [https://doi.org/10.2134/agronj1984.00021962007600030019x]. *Agronomy Journal*, *76*(3), 451-454. https://doi.org/https://doi.org/10.2134/agronj1984.00021962007600030019x

Roberts, T. (2010). Nutrient best management practices: Western perspectives on global nutrient stewardship. 19th World Congress of Soil Science, Soil Solutions for a Changing World, p172-75, Brisbane, Australia.

Rochester, I., Peoples, M., Hulugalle, N., Gault, R., & Constable, G. (2001). Using legumes to enhance nitrogen fertility and improve soil condition in cotton cropping systems. *Field Crops Research*, *70*(1), 27-41.

Rurangwa, E., Vanlauwe, B., & Giller, K. E. (2018). Benefits of inoculation, P fertilizer and manure on yields of common bean and soybean also increase yield of subsequent maize. *Agriculture, Ecosystems & Environment*, *261*, 219-229. https://doi.org/https://doi.org/10.1016/j.agee.2017.08.015

Rushovich, D. (2020). *Sulfur Management to Enhance Yield and Protein Quality of Grain Legumes* University of Maryland, College Park].

Rymuza, K., Radzka, E., & Wysokiński, A. (2020). Nitrogen uptake from different sources by non-GMO soybean varieties. *Agronomy*, *10*(9), 1219.

Sainju, U. M., Lenssen, A. W., Allen, B. L., Stevens, W. B., & Jabro, J. D. (2016). Nitrogen balance in response to dryland crop rotations and cultural practices. *Agriculture, Ecosystems & Environment*, *233*, 25-32. https://doi.org/https://doi.org/10.1016/j.agee.2016.08.023

Salvagiotti, F., Cassman, K. G., Specht, J. E., Walters, D. T., Weiss, A., & Dobermann, A. (2008). Nitrogen uptake, fixation and response to fertilizer N in soybeans: A review. *Field Crops Research*, *108*(1), 1-13.

Sanginga, N., Dashiell, K., Okogun, J. A., & Thottappilly, G. (1997). Nitrogen fixation and N contribution by promiscuous nodulating soybeans in the southern Guinea savanna of Nigeria. *Plant and Soil*, *195*(2), 257-266. https://doi.org/10.1023/A:1004207530131

Santachiara, G., Borrás, L., Salvagiotti, F., Gerde, J. A., & Rotundo, J. L. (2017). Relative importance of biological nitrogen fixation and mineral uptake in high yielding soybean cultivars. *Plant and Soil*, *418*(1), 191-203. https://doi.org/10.1007/s11104-017-3279-9

Santachiara, G., Salvagiotti, F., Gerde, J. A., & Rotundo, J. L. (2018). Does biological nitrogen fixation modify soybean nitrogen dilution curves? *Field Crops Research*, *223*, 171-178. https://doi.org/https://doi.org/10.1016/j.fcr.2018.04.001

Sathyanarayana, E., Padmaja, G., Sharma, S. H. K., & Vidya, G. (2021). Effect of different sources and levels of nitrogen, phosphorus and zinc on soil fertility, nutrient content, uptake and yield of soybean. *The Pharma Journa*, *10*(7), 1152-1156.

Schapaugh Jr., W. T., & Wilcox, J. R. (1980). Relationship Between Harvest Indices and other Plant Characteristics in Soybean1. *Crop Science*, *20*(4). https://doi.org/https://doi.org/10.2135/cropsci1980.0011183X002000040028x

Schipanski, M., Drinkwater, L., & Russelle, M. (2010). Understanding the variability in soybean nitrogen fixation across agroecosystems. *Plant and soil*, *329*, 379-397.

Schulte-Uebbing, L. F., Beusen, A. H. W., Bouwman, A. F., & de Vries, W. (2022). From planetary to regional boundaries for agricultural nitrogen pollution. *Nature*, *610*(7932), 507-512. https://doi.org/https://doi.org/10.1038/s41586-022-05158-2

Silva, J. V., van Ittersum, M. K., ten Berge, H. F. M., Spätjens, L., Tenreiro, T. R., Anten, N. P. R., & Reidsma, P. (2021). Agronomic analysis of nitrogen performance indicators in intensive arable cropping systems: An appraisal of big data from commercial farms. *Field Crops Research*, *269*, 108176. https://doi.org/https://doi.org/10.1016/j.fcr.2021.108176

Singh, R., Prasad, R., Delaney, D. P., & Watts, D. B. (2021). Does Soybean Yield and Seed Nutrient Content Change Due to Broiler Litter Application? *Agronomy*, *11*(8), 1523. https://www.mdpi.com/2073-4395/11/8/1523

Sogut, T. (2006). Rhizobium inoculation improves yield and nitrogen accumulation in soybean (Glycine max) cultivars better than fertiliser. *New Zealand journal of crop and horticultural science*, *34*(2), 115-120.

Stajković, O., Kuzmanović, D., Rasulić, N., Knežević, M., Ugrenović, V., Buntić, A., & Delić, D. (2021). Effect of inoculation with Bradyrhizobium and phosphate solubilizing bacteria on soybean seed yield and composition. *Zemljiste i biljka*, *70*(2), 56-66. https://doi.org/10.5937/ZemBilj2102056S

Stewart, W., Dibb, D., Johnston, A., & Smyth, T. (2005). The contribution of commercial fertilizer nutrients to food production. *Agronomy journal*, *97*(1), 1-6. https://doi.org/ https://doi.org/10.2134/agronj2005.0001

Suciu, L., Tarau, A., Urda, C., Chetan, F., Muresanu, F., Sopterean, L., Boiu-Sicuia, O., & Barbu, L. (2021). The influence of bacterial inoculats on pathogens, yield and quality in soybean crop. *AgroLife Sci. J*, *10*, 221-226.

Tanwar, S., & Shaktawat, M. (2003). Influence of phosphorus sources, levels and solubilizers on yield, quality and nutrient uptake of soybean (Glycine max)—wheat (Triticum aestivum) cropping system in southern Rajasthan. *Indian Journal of Agricultural Sciences*, *73*(1), 3-7.

Tavares, C. J., Ribeiro Junior, W. Q., Ramos, M. L. G., Pereira, L. F., Casari, R. A. d. C. N., Pereira, A. F., de Sousa, C. A. F., da Silva, A. R., Neto, S. P. d. S., & Mertz-Henning, L. M. (2022). Water Stress Alters Morphophysiological, Grain Quality and Vegetation Indices of Soybean Cultivars. *Plants*, *11*(4), 559. https://www.mdpi.com/2223-7747/11/4/559

Tolokonnikov, V., Novikov, A., Koshkarova, T., Vorontsova, E., & Fomin, S. (2021). Agromeliorative methods of cultivation of a new variety of soybeans Volgogradka 2 under irrigation conditions. IOP Conference Series: Earth and Environmental Science,

Toomsan, B., McDonagh, J. F., Limpinuntana, V., & Giller, K. E. (1995). Nitrogen fixation by groundnut and soyabean and residual nitrogen benefits to rice in farmers' fields in Northeast Thailand. *Plant and Soil*, *175*(1), 45-56. https://doi.org/10.1007/BF02413009

United Nations. (2015). World population prospects: The 2015 revision. *United Nations Econ Soc Aff*, *33*(2), 1-66.

Unkovich, M., Baldock, J., & Forbes, M. (2010). Variability in harvest index of grain crops and potential significance for carbon accounting: examples from Australian agriculture. *Advances in agronomy*, *105*, 173-219.

Urquiaga, S., Jantalia, C., Zotarelli, L., Araujo, E., Alves, B., Boddey, R., Cabezas, W., Dos Santos, H., & Torres, E. (2006). Nitrogen dynamics in soybean-based crop rotations under conventional and zero tillage in Brazil. *IAEA Proceedings Series "Management Practices for Improving Sustainable Crop Production in Troicla Acid Soils". IAEA-STI-PUB*, *1285*, 13-46.

Van Egmond, K., Bresser, T., & Bouwman, L. (2002). The European Nitrogen Case. *AMBIO: A Journal of the Human Environment*, *31*(2), 72-78, 77. https://doi.org/10.1579/0044-7447-31.2.72

Van Ittersum, M. K., Van Bussel, L. G., Wolf, J., Grassini, P., Van Wart, J., Guilpart, N., Claessens, L., De Groot, H., Wiebe, K., & Mason-D'Croz, D. (2016). Can sub-Saharan Africa feed itself? *Proceedings of the National Academy of Sciences*, *113*(52), 14964-14969. https://doi.org/https://doi.org/10.1073/pnas.1610359113

Van Vugt, D., Franke, A. C., & Giller, K. E. (2018). Understanding variability in the benefits of N2-fixation in soybean-maize rotations on smallholder farmers' fields in Malawi. *Agriculture, Ecosystems & Environment*, *261*, 241-250. https://doi.org/https://doi.org/10.1016/j.agee.2017.05.008

Vanlauwe, B., Hungria, M., Kanampiu, F., & Giller, K. E. (2019). The role of legumes in the sustainable intensification of African smallholder agriculture: Lessons learnt and challenges for the future. *Agriculture, Ecosystems & Environment*, *284*, 106583. https://doi.org/https://doi.org/10.1016/j.agee.2019.106583

Vasilas, B., & Ham, G. (1985). Intercropping nodulating and non-nodulating soybeans: effects on seed characteristics and dinitrogen fixation estimates. *Soil Biology and Biochemistry*, *17*(4), 581-582. https://www.academia.edu/58364184/Intercropping_nodulating_and_non_nodulating_soybeans_Effects_on_seed_characteristics_and_dinitrogen_fixation_estimates

Walker, W., Raines, G., & Peck, T. (1985). Effect of soybean cultivar, phosphorus and potassium upon yield and chemical composition. *Journal of plant nutrition*, *8*(1), 73-87. https://www.tandfonline.com/doi/pdf/10.1080/01904168509363325?needAccess=true

West, P. C., Gerber, J. S., Engstrom, P. M., Mueller, N. D., Brauman, K. A., Carlson, K. M., Cassidy, E. S., Johnston, M., MacDonald, G. K., Ray, D. K., & Siebert, S. (2014). Leverage points for improving global food security and the environment. *Science*, *345*(6194), 325-328. https://doi.org/doi:10.1126/science.1246067

Zhang, X., Davidson, E. A., Mauzerall, D. L., Searchinger, T. D., Dumas, P., & Shen, Y. (2015). Managing nitrogen for sustainable development. *Nature*, *528*(7580), 51-59. https://doi.org/10.1038/nature15743

Zhang, X., Zou, T., Lassaletta, L., Mueller, N. D., Tubiello, F. N., Lisk, M. D., Lu, C., Conant, R. T., Dorich, C. D., Gerber, J., Tian, H., Bruulsema, T., Maaz, T. M., Nishina, K., Bodirsky, B. L., Popp, A., Bouwman, L., Beusen, A., Chang, J., . . . Davidson, E. A. (2021). Quantification of global and national nitrogen budgets for crop production. *Nature Food*, *2*(7), 529-540. https://doi.org/10.1038/s43016-021-00318-5

Zimmer, S., Messmer, M., Haase, T., Piepho, H.-P., Mindermann, A., Schulz, H., Habekuß, A., Ordon, F., Wilbois, K.-P., & Heß, J. (2016). Effects of soybean variety and Bradyrhizobium strains on yield, protein content and biological nitrogen fixation under cool growing conditions in Germany. *European Journal of Agronomy*, *72*, 38-46. https://doi.org/https://doi.org/10.1016/j.eja.2015.09.008

Zingore, S., Murwira, H. K., Delve, R. J., & Giller, K. E. (2008). Variable grain legume yields, responses to phosphorus and rotational effects on maize across soil fertility gradients on African smallholder farms. *Nutrient Cycling in Agroecosystems*, *80*, 1-18.

Zotarelli, L., Zatorre, N. P., Boddey, R. M., Urquiaga, S., Jantalia, C. P., Franchini, J. C., & Alves, B. J. (2012). Influence of no-tillage and frequency of a green manure legume in crop rotations for balancing N outputs and preserving soil organic C stocks. *Field Crops Research*, *132*, 185-195.

Appendices

Appendix 1

Table 7
Mean values, sample size and standard deviation per UN sub-region for Harvest Index, Grain N, Residue N, NDFAkg and NDFA%.

Region	Mean value and standard deviation for each variable									
	Harvest Index (-)	SD (-)	Grain N (%)	SD (%)	Residue N (%)	SD (%)	NDFAkg (kg N ha^{-1})	SD (kg N ha^{-1})	NDFA% (%)	SD (%)
Australia and New Zealand	0.22(N=2)	0.02	6.23(N=2)	0.56	0.83(N=2)	0.37	178(N=33)	62	55(N=35)	16
Eastern Asia	-	-	5.72(N=36)	1.17	-	-	94(N=27)	34	67(N=27)	6
Eastern Europe	0.37(N=12)	0.05	4.49(N=12)	0.32	0.57(N=12)	0.05	-	-	-	-
Latin America and the Caribbean	0.43(N=54)	0.08	6.34(N=114)	0.36	0.75(N=36)	0.36	276(N=49)	57	66(N=38)	18
Northern Africa	0.28(N=27)	0.02	5.44(N=38)	1.10	-	-	-	-	-	-
Northern America	0.42(N=90)	0.07	6.14(N=129)	0.50	1.51(N=9)	0.15	138(N=39)	80	61(N=42)	23
Northern Europe	-	-	5.44(N=24)	0.29	-	-	-	-	-	-
South-eastern Asia	0.39(N=11)	0.06	7.04(N=19)	0.38	0.69(N=11)	0.11	118(N=28)	31	60(N=11)	11
Southern Asia	0.38(N=159)	0.07	5.82(N=47)	0.75	1.65(N=19)	0.57	-	-	-	-
Southern Europe	-	-	6.48(N=25)	1.03	1.85(N=16)	0.11	-	-	-	-
Sub-Saharan Africa	0.40(N=64)	0.13	5.02(N=78)	0.85	-	-	69(N=78)	58	64(N=100)	20
Western Asia	-	-	6.12(N=18)	0.26	-	-	-	-	-	-
Western Europe	0.42(N=5)	0.02	6.18(N=77)	0.79	0.79(N=5)	0.39	106(N=23)	65	46(N=23)	7
World	**0.39(N=424)**	**0.09**	**5.93(N=619)**	**0.87**	**1.11(N=110)**	**0.59**	**139(N=277)**	**93**	**56(N=276)**	**19**

Appendix 2

Linear Regression of predicted versus actual harvest index, grain N and residue N showing all three chosen linear mixed-effects models for prediction

Harvest Index **Grain N** **Residue N**

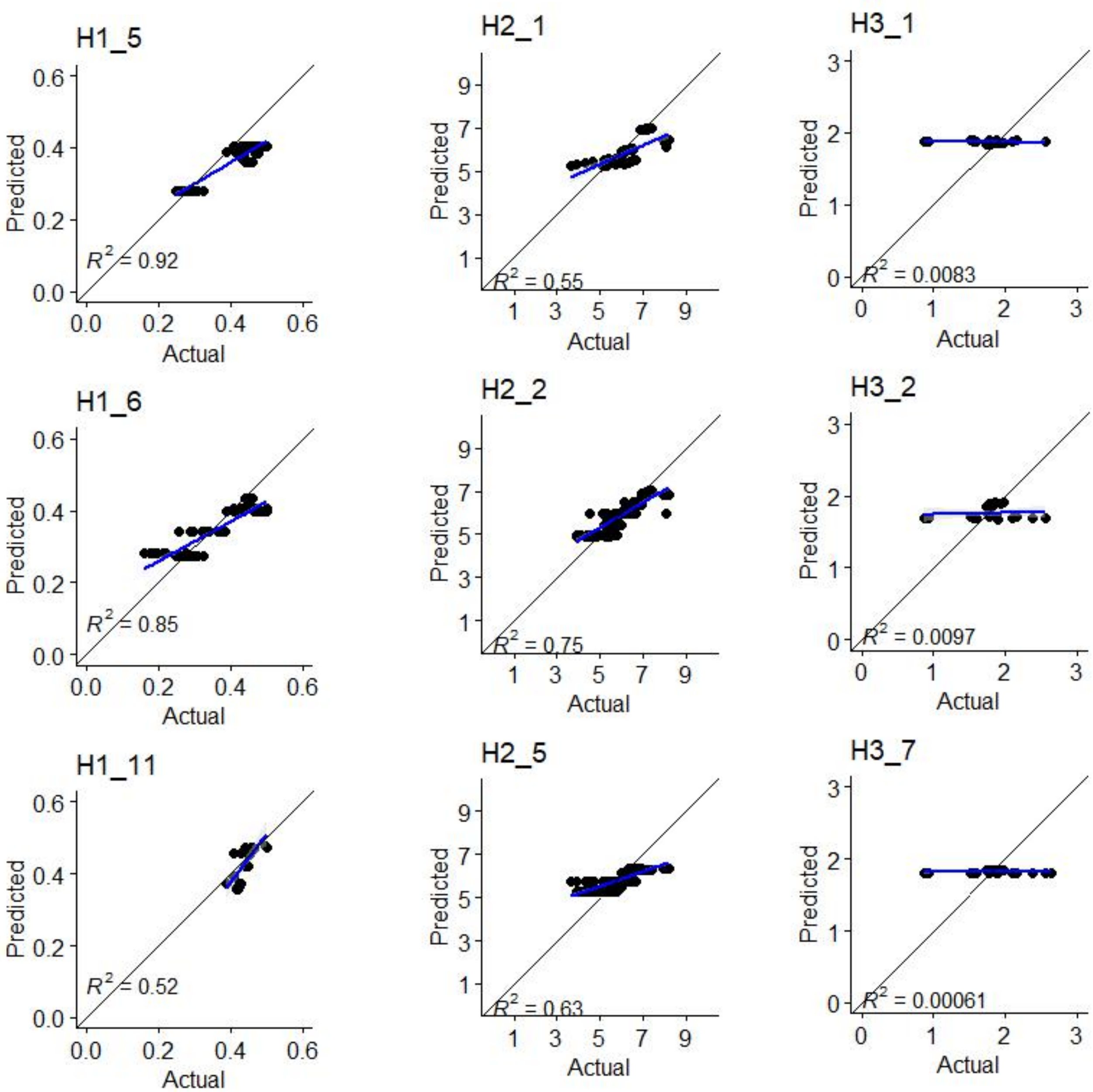

Figure 11: Scatter plots of actual values versus predicted values for all linear mixed-effects predictor models for Harvest Index (left), Grain N (middle) and Residue N (right). Model names are mentioned in the top left of each scatter plot. Model equation, Nakagawa's R^2 and Akaike's Information Criterion per model can be found in Table 3 for Harvest Index, Table 4 for Grain N and Table 5 for Residue N.

Appendix 3

Correlation matrix including response variables and key explanatory variables

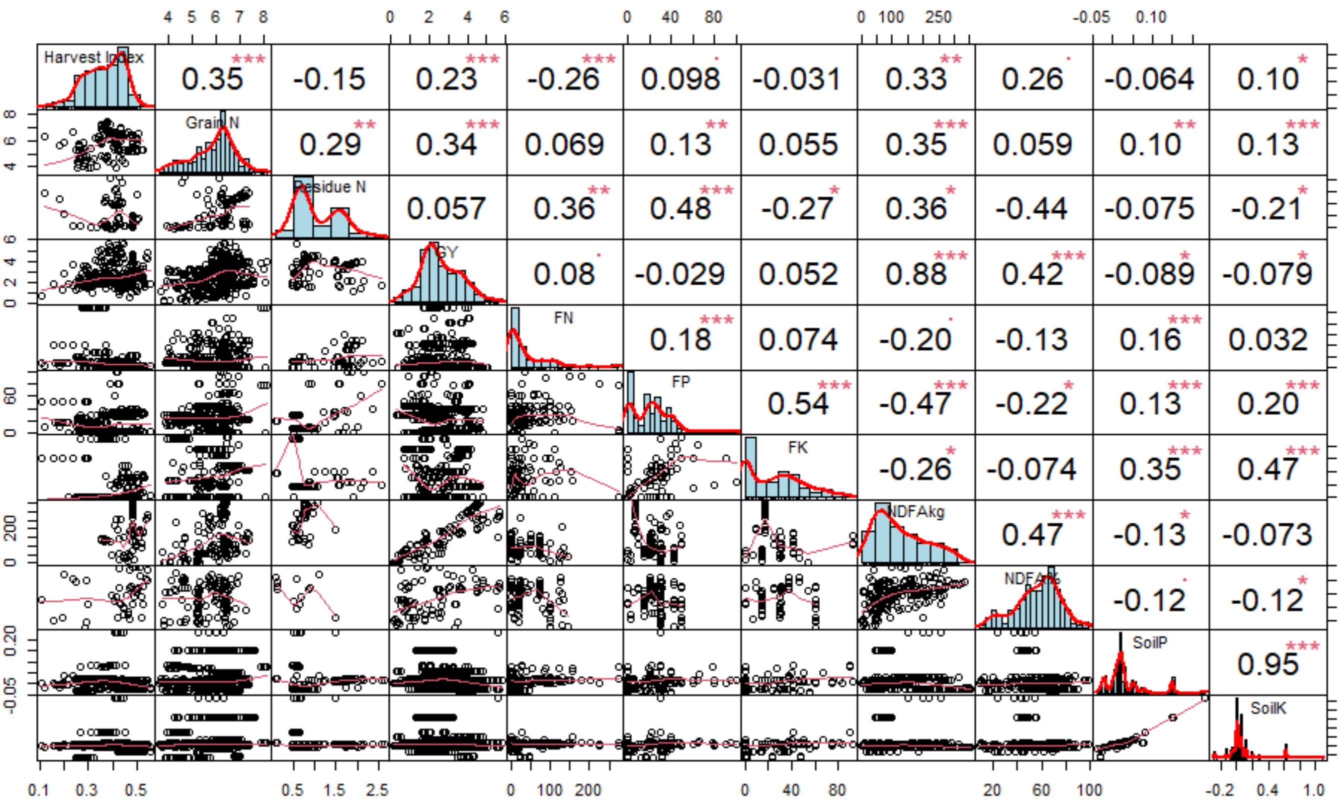

Figure 12: Correlation matrix including the variables Harvest Index (unitless), Grain N (kg N ha^{-1}), Residue N (kg N ha^{-1}), GY (Mg ha^{-1}), FN (kg N ha^{-1}), FP (kg P ha^{-1}), FK (kg K ha^{-1}), SoilP (kg P m^{-2}) and SoilK (kg K m^{-2}). Variable definitions can be found on Page 6. The diagonal shows the distribution of data for each variable in histograms. Below the diagonal are bivariate scatter plots with fitted density lines in red. Above the diagonal are correlation coefficients, with level of significance is displayed by symbols indicating p-values as: 0 – 0.001 = ***, 0.001 – 0.01 = ** , 0.01–0.05 = *, 0.05–0.1 = ".", and 0.1–1 = " ".

Appendix 4

Table 9
Publications from which data is included in the dataset used in this study

#	Publication
1	Adjei-Nsiah et al. (2021)
2	Aher et al. (2022)
3	Amadou et al. (2021)
4	Amiri et al. (2021)
5	Andriani et al. (1991)
6	Asres and Tiruneh (Preprint (2020))
7	Basal and Szabó (2020)
8	Bellaloui et al. (2011)
9	Bender et al. (2015)
10	Bhangoo and Albritton (1972)
11	Boddey et al. (1990)
12	Bortolon et al. (2018)
13	Cafaro La Menza et al. (2019)
14	Cafaro La Menza et al. (2020)
15	Cannon et al. (2021)
16	Chețan et al. (2021)
17	Coale et al. (1985)
18	Deibert et al. (1979)
19	Di Ciocco et al. (2008)
20	Domingos et al. (2021)
21	Dragicevic et al. (2022)
22	Eliçin et al. (2021)
23	Engy et al. (2020)
24	Erbil et al. (2020)
25	Gan et al. (2002)
26	Gan et al. (2003)
27	Gelfand and Robertson (2015)
28	George et al. (1988)
29	Ghani et al. (2021)
30	Guafa et al. (1993)
31	Gyogluu et al. (2016)
32	Ham and Caldwell (1978)
33	Hiep et al. (2002)
34	Jansone et al. (2021)
35	Jarecki and Bobrecka-Jamro (2021)
36	Jat et al. (2021)
37	Jefing et al. (1992)
38	Kakabouki et al. (2020)
39	Karhale (2021)
40	Kubar et al. (2021)

41	Kucey et al. (1988)
42	Landriscini et al. (2019)
43	Latifinia and Eisvand (2022)
44	Lohar et al. (2020)
45	Machado et al. (2021)
46	Machiani et al. (2021)
47	Mamun et al. (2022)
48	Mandić et al. (2020)
49	Movalia and Savalia (2021)
50	Munyinda et al. (1988)
51	Nassar et al. (2021)
52	Oberson et al. (2007)
53	Ojiem et al. (2007)
54	Okogun et al. (2005)
55	Peoples et al. (1995)
56	Radzka et al. (2021)
57	Raj et al. (2021)
58	Rashmi et al. (2022)
59	Rennie and Dubetz (1984)
60	Rochester et al. (2001)
61	Rurangwa et al. (2018)
62	Rushovich (2020)
63	Rymuza et al. (2020)
64	Sanginga et al. (1997)
65	Santachiara et al. (2017)
66	Santachiara et al. (2018)
67	Sathyanarayana et al. (2021)
68	Schapaugh Jr. and Wilcox (1980)
69	Singh et al. (2021)
70	Stajković et al. (2021)
71	Suciu et al. (2021)
72	Tanwar and Shaktawat (2003)
73	Tavares et al. (2022)
74	Tolokonnikov et al. (2021)
75	Toomsan et al. (1995)
76	Urquiaga et al. (2006)
77	Van Vugt et al. (2018)
78	Vasilas and Ham (1985)
79	Walker et al. (1985)
80	Zimmer et al. (2016)
81	Zingore et al. (2008)
82	Zotarelli et al. (2012)