Agricultural Systems 209 (2023) 103667

Contents lists available at ScienceDirect

Agricultural Systems

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

Identifying drivers for variability in maize (*Zea mays* L.) yield in Ghana: A meta-regression approach

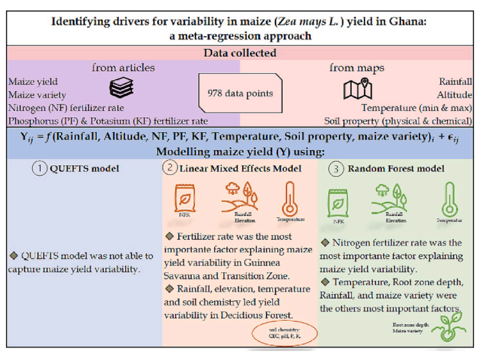
Anselme K.K. Kouame [a], Prem S. Bindraban [a,*], Isaac N. Kissiedu [a], Williams K. Atakora [a], Khalil El Mejahed [b]

[a] *International Fertilizer Development Center (IFDC), Muscle Shoals, AL 35662, USA*
[b] *Mohammed VI Polytechnic University (UM6P), AgroBioSciences, Lot 660, Hay Moulay Rachid, 43150 Ben Guerir, Morocco*

HIGHLIGHTS

- The factors influencing maize yield variability in the agro-ecological zones of Ghana are still unclear.
- Overall, nitrogen fertilizer rate was the most important factor explaining maize yield variability.
- In the Guinea Savanna and Transition Zone, fertilizer rates and maize variety determined the variability in yield.
- In the Deciduous Forest, the environmental factors and soil chemistry were predominant in explaining the yield variability.
- Random forest modelling showed that root zone depth was also a key factor in explaining maize yield variability in Ghana.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Jagadish Timsina

Keywords:
Maize yield variability
Fertilizer rate
QUEFTS
Linear Mixed Effects Model
Random Forest
Maize yield prediction

ABSTRACT

CONTEXT: Maize is the main cereal crop in Ghana, but its production is adversely affected by various biotic and abiotic factors.
OBJECTIVE: This study aimed to highlight the factors related to maize yield variability. To this end, yields from 978 data points within 3 agro-ecological zones (AEZs) were used in crop-based and statistical modelling.
METHODS: The QUantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) model, the Linear Mixed Effects Model (LMM), and the Random Forest (RF) model were used to evaluate multiple effect sizes.
RESULTS AND CONCLUSIONS: Analyzing an entire set of yield data points with QUEFTS, and LMM explained 19%, and 26% of yield variability, respectively. Considering all data points in the RF model, nitrogen fertilizer (NF) rate, temperature, root zone depth, rainfall, and variety accounted for 27%, 15%, 13%, 10%, and 9% of yield variation, respectively. In Guinea Savanna (GS), Transition Zone (TZ), and Deciduous Forest (DF), QUEFTS explained 30%, 20%, and 4% of yield variability, respectively. LMM, however, explained 47%, 51%, and 79% of yield variability in those AEZs. LMM showed that the phosphorus fertilizer (PF) rate was very important and exceeded the importance of the NF rate in GS. LMM showed also that yield variability was significantly related to maize variety at the AEZ scale. In DF, soil chemistry (marginal $R^2 = R^2_m = 0.48$) and environmental variables ($R^2_m = 0.43$) contributed more to explaining yield variability, whereas in GS and TZ, fertilizer rates ($R^2_m = 0.35$ in GS

* Corresponding author.
E-mail address: pbindraban@ifdc.org (P.S. Bindraban).

https://doi.org/10.1016/j.agsy.2023.103667
Received 21 April 2022; Received in revised form 18 April 2023; Accepted 19 April 2023

and 0.26 in TZ) and variety ($R^2_m = 0.04$ in GS and 0.20 in TZ) played a much larger role. In GS, TZ, and DF, the RF model explained 74%, 79%, and 84% of the variance in yield, respectively. These findings suggest low impact of fertilization on yield on the inherently fertile soils in the DF, while fertilization drives yield increase in the less fertile TZ and GS AEZs. We may conclude that QUEFTS was unable to capture yield variability and, according to RF and LMM analysis, the NF rate was the most important factor in explaining yield variability in the data. It can also be concluded that the factors responsible for yield variability are AEZ dependent.

SIGNIFICANCE: We discuss the implications of these findings to uncover factors driving maize yield variability. It also provides information to guide and prioritize actions to be taken based on the importance of these factors in contributing to yield variability.

1. Introduction

Global demand for food will continue to increase for at least 50 years (Tilman et al., 2011; Cicin-Sain, 2018), and climate change is not helping matters. Agricultural production in sub-Saharan Africa (SSA) must at least triple to meet this growing food demand (Godfray et al., 2010; Rahman et al., 2021). Additionally, the agri-food system is essential to achieving at least 12 of the 17 Sustainable Development Goals (SDGs) of the United Nations by 2030, and it plays a significant role in the economy of the SSA nations (FAO/OECD, 2018). For many years, it has been the economy's fastest-growing industry in Ghana (Diao et al., 2019). Thus, agricultural growth is the main driver of poverty reduction and the largest source of employment for rural communities, mainly smallholder farmers with 2 ha of land or less (USAID, 2022).

However, farmers face changing and increasingly unpredictable weather conditions, drastically reducing soil fertility, and typically use local or inbred crop varieties. Bationo et al. (2018) reported that soil nutrient depletion rates of about 35 kg N, 4 kg P, and 20 kg K per hectare are worrisome and prevalent in all agro-ecological zones (AEZs) in Ghana, with nitrogen (N) and phosphorus (P) being the most deficient nutrients (Zingore et al., 2015). As a result, yields obtained by smallholder farmers are far below the potentially attainable yields, hampering agricultural production and jeopardizing economic development and food security (Adzawla et al., 2021a, 2021b).

One solution is to increase fertilizer application by farmers. However, it is increasingly understood that crop yield in many areas of Africa, including Ghana, is depressed by a variety of soil degradation problems and many other factors, such as crop variety, soil organic matter, and soil depth (Sadras and Calvino, 2001; Kpotor et al., 2014; Tetteh et al., 2016; Guilpart et al., 2017; Leenaars et al., 2018). Furthermore, high variability in climatic conditions (rainfall and temperature) causes uncertainties in agricultural productivity, with profound impacts on the ecology, economy, and social welfare of rural farmers (Onduru and Du Preez, 2007; Kyei-Mensah et al., 2019).

Despite current low crop productivity, Ghana could intensify production and significantly close the current yield gaps of major cereals (Bationo et al., 2018; van Loon et al., 2019), since it has been estimated that only about 20%, on average of the potential maize yield is being achieved across the country (GYGA, 2021). For example, addressing nutrient deficiencies by applying fertilizer alone would help to reduce the maize yield gap to 50% of the attainable yield (Mueller et al., 2012). However, Adzawla et al. (2021a, 2021b) and SRID/MoFA (2021) have reported that the maize yield is around 2 t ha^{-1} even with the application of nitrogen (N), phosphorus (P), and potassium (K) compound fertilizers. Subsequently, Bua et al. (2020) found that maize yield is highly variable, with yields ranging from a mere 500 kg ha^{-1} to >8 t ha^{-1}. This large yield variability depresses farmers' incentive and ability to purchase fertilizers in subsequent seasons (Njoroge, 2019).

It is important, therefore, to identify the key drivers for the observed variabilities in maize yield, which can be done using model-based approaches. Various studies around the world have shown that the application of system models (Wallach et al., 2018) can be useful in determining and prioritizing the relative importance of factors that contribute to yield variability (Jeong et al., 2016; Lamos-Díaz et al., 2020; Nevavuori et al., 2020; Paudel et al., 2021; Timsina et al., 2021).

Models such as the QUantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) have been advocated by several studies for estimating crop yield (Tabi et al., 2007; Tittonell et al., 2008; Wijayanto and Prastyanto, 2012; Xu et al., 2013; Ren et al., 2015), and specifically recommended for use in Ghana as well (Wijayanto and Prastyanto, 2012; Antwi et al., 2017). QUEFTS was developed for maize and considers soil chemical properties (pH, organic carbon, extractable and total phosphorus, exchangeable potassium, and organic nitrogen) and fertilizer application as the input variables (Janssen et al., 1990; Sattaria et al., 2014). The model assumes that all other production factors are optimal and does not consider the maize variety, soil physical properties, or climatic variables. But in real life, maize grain yield could be expressed by Eq. (1) according to Giller et al. (2013), as the result of "Variety", biophysical "Environment" (temperature and precipitation), "Soil" (acidity, texture, root depth, limiting nutrients), and management, including mineral fertilizers interactions:

$$\text{Maize yield} = \text{Environment}*\text{Soil}*\text{Variety}*\text{Fertilizer} \tag{1}$$

Therefore, it would be useful to complement the QUEFTS model with statistical modelling as part of further research to determine how these variables not included in the QUEFTS relate to maize yield in Ghana (Kihara et al., 2016; van Loon et al., 2019; Atiah et al., 2021). Thus, QUEFTS was utilized in conjunction with the statistical models Linear Mixed Effects Model (LMM) and Random Forest (RF).

In this paper, we present the findings of our analysis of almost a thousand maize research data points with N, P, and K fertilizers in the Deciduous Forest (DF), Transition Zone (TZ), and Guinea Savanna (GS) to elucidate the biotic and/or abiotic factors relating to the observed maize yield variability.

2. Materials and methods

2.1. Study area

Maize trials were conducted in three of Ghana's AEZs located between 5° and 15° East and 4° and 16° North: DF ($n = 186$), TZ ($n = 227$), and GS ($n = 565$) (Fig. 1). The rainy season defined the planting periods of the trials conducted by researchers. Indeed, a large part of the GS has a mono-modal rainy season (Kranjac-Berisavljevic et al., 1999) and the average monthly temperatures varied between 27 °C and 36 °C (Ghansah et al., 2018; Darko et al., 2019). Conversely, TZ and DF have 2 rainy seasons; the major season lasts from March to July and a minor season occurs from September to November, with June registering the highest rainfall (Nkrumah et al., 2014). The average temperatures in these AEZs are between 24 °C and 34 °C (Ghansah et al., 2018; Darko et al., 2019).

2.2. Yield data

Maize yields from 978 data points from research experiments conducted between 2001 and 2017 were collected from scientific articles and local institutional reports (Supplementary Table S 1). Dates of sowing and/or the growing seasons were not mentioned in most of the reports collected. The planting season was therefore assigned to each data points according to the rainy period in each AEZ (Fig. 1), considering that the trials were rainfed (Adu et al., 2014).

2.3. Fertilizer data

Fertilizer treatments were heterogeneous and included organic fertilizers only, organic fertilizers in combination with inorganic fertilizers, and inorganic fertilizers only. Organic fertilizers, including cow dung; poultry, goat, and sheep droppings; compost; town waste; biochar and palm bunch ash, were all converted into N, P_2O_5, and K_2O rates (Fening et al., 2009; Adjei-Nsiah, 2012; Kanton et al., 2016; Badu et al., 2019). The lowest and highest rates of nitrogen fertilizer (NF) were 7 kg ha^{-1} and 281 kg ha^{-1}, with common rates of 30 kg ha^{-1}, 60 kg ha^{-1}, and 90 kg ha^{-1}. The lowest and highest rates of P_2O_5 (PF) and K_2O (KF) were 3 kg ha^{-1} and 90 kg ha^{-1}, with common rates of 20 kg ha^{-1}, 40 kg ha^{-1}, and 60 kg ha^{-1}. To assess the fertilizer contribution to yield, data points were categorized into 2 treatments: "With_fertilizer" and "Without_fertilizer".

2.4. Soil and environmental data

Table 1 presents the continuous variables used in the models. Soil chemical and physical properties data were obtained from the African SoilGrids (ISRIC), at 250 m of resolution for the 0–30 cm topsoil. Rainfall and temperature datasets are from the WorldClim (Fick and Hijmans, 2017) database, at a spatial resolution of 1 km^2 and land elevation data with a resolution of 30 m from Shuttle Radar Topography Mission (SRTM) Digital Terrain Elevation.

The geographic coordinates of each data point were overlaid on the maps to extract soil and environmental data. Mean, minimum, and maximum temperatures and total rainfall for each data point were obtained from aggregated monthly measurements according to the recommended rainy season for planting maize in Ghana, on the basis that maize was harvested no later than 120 days after the date of sowing (Adu et al., 2014).

Table 1
Quantitative variables (soil property and environmental factors) used as input data in the models.

Continuous variables	Unit	Abbreviation in models	Authors
Land elevation	m	ELV	(Farr and Kobrick, 2000)
Monthly precipitation	mm	MP[1]	
Monthly mean temperature	°C	MMeT[2]	
Monthly min temperature	°C	MMiT[2]	(Fick and Hijmans, 2017)
Monthly max temperature	°C	MMaT[2]	
Soil organic carbon	g kg^{-1}	SOC	
Total nitrogen	g kg^{-1}	TOTN	
Total phosphorus	mg kg^{-1}	TOTP	
Extractable phosphorus	mg kg^{-1}	P	
Extractable potassium	mmol kg^{-1}	K	
Soil pH	–	pH	(Hengl et al., 2015)
Cation exchange capacity	mmol $_{(+)}$ kg^{-1}	CEC	
Sand content	%	SAND	
Clay content	%	CLAY	
Silt content	%	SILT	
Root zone depth	cm	RootDEP	

[1] Precipitation is the sum for the estimated maize growing season.
[2] Minimum, maximum, and mean temperatures are the mean for the estimated maize growing season.

2.5. QUEFTS model

QUEFTS was implemented in the R software, based on Sattaria et al. (2014). QUEFTS' calibration and validation parameters from Wijayanto and Prastyanto (2012) and Antwi et al. (2017) were used to replace the

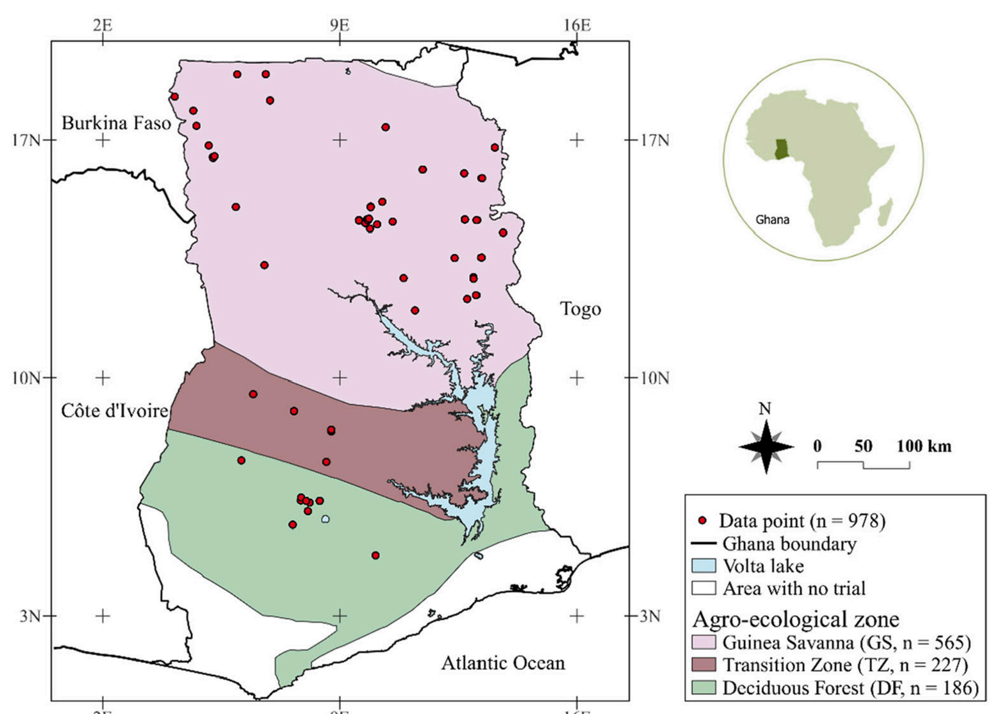


Fig. 1. Geographic distribution of the experimental locations on a map of Ghana that shows the country's agro-ecological zones. Many experimental locations overlap.
Agroecological map source: (Antwi et al., 2014).

Janssen et al. (1990)' default parameterization. The maximum yield was set to 10 t ha^{-1}. Soil-available P in P-Mehlich3 and the exchangeable K in K-NH$_4$Ac extracted from the ISRIC maps were converted into P-Olsen and K-Mehlich3 extractable based on Sawyer and Mallarino (1999)' pedo-transfer functions. The performance of the QUEFTS model was assessed using the coefficient of determination (R^2) (Krause et al., 2005), through a linear regression between QUEFTS-estimated and observed yield, and the significance of correlation (r) was determined based on a $p < 0.05$.

2.6. Linear mixed effects model

Prior to statistical modelling, data points were subjected to the Shapiro-Wilk normality test and were not found to be normally distributed (Supplementary Table S 2). Therefore, maize yield was transformed and normalized to meet the assumptions of homoscedasticity and homogeneity of analysis of variance (ANOVA) and LMM errors variance (Supplementary Fig. S 1). To evaluate the effectiveness of the normalization technique applied, the "bestNormalize()" function of the bestNormalize R package (Peterson, 2021) was used. Thus, the out-of-sample method via 10-fold cross-validation with 3 repeats was performed to estimate the Pearson P-statistic (normality statistic). The transformation technique was selected according to the calculated value of the normality statistic.

The numerical variables in Table 1 were all standardized using the "scale()" function in R. In effect, they were standardized for all observation by first subtracting the mean and next dividing by the standard deviation.

Variance inflation factors (VIFs) were calculated to measure the inflation of the predictor coefficients due to collinearities between the independent numerical variables in Table 1. The "vif()" function of the car R package (Fox and Weisberg, 2019) was used to calculate the VIF values. Silt and sand had VIFs >10, so multicollinearities were very likely. They were considered as potential predictors to be eliminated. However, the VIF for sand became <3 when silt was removed from the explanatory variables. Soil organic carbon (SOC) and total nitrogen (TOTN) also had VIFs >5. As with sand, the VIF for SOC became <5 when TOTN was removed. Therefore, in the rest of the ANOVA and LMM analysis, silt and total nitrogen were removed as predictors. On the other hand, the maximum temperature (MMaT) and minimum temperature (MMiT) also had a VIF > 5. Instead of deleting one of them, they were combined, thus creating a new explanatory variable (i.e., mean temperature - MMeT).

The R function "TukeyHSD()" was used to perform the Tukey-Kramer test, and the multcompView R package (Graves et al., 2019) was used to compact the letter display to indicate significant differences between AEZ, Treatment, and Treatment*AEZ subgroups following eq. 2, where Y_{ijk} is the kth observed value of yield formed by the ith level of treatment effect, the jth level of AEZ effect, the ijth level of interaction between treatment and AEZ, and the error (ϵ_{ijk}).

$$Y_{ijk} = \text{Treatment}_i + \text{AEZ}_j + (\text{Treatment*AEZ})_{ij} + \epsilon_{ijk} \tag{2}$$

The means of the maize yield were significantly different at $p < 0.05$. Tukey's Honest Significant Difference (TukeyHSD) multiple comparison analysis method tests were used because there were unequal data point sizes among AEZ and Treatments subgroups (Marusteri and Bacarea, 2010; McHugh, 2011).

One-way ANOVA Eqs. (3)–(5) were used to measure the granularity of the random effect (i.e., Intraclass Correlation Coefficient — ICC) by Treatment, AEZ, and Year on maize yields, respectively. The ICC was calculated as the ratio of the variance between the random effect and the model's total variance (Nakagawa and Schielzeth, 2010). The one-way ANOVA equations with random effects were constructed as follows:

$$Y_{ij\_Treat} = \text{Treatment}_i + \epsilon_{ij} \tag{3}$$

$$Y_{ij\_AEZ} = \text{AEZ}_i + \epsilon_{ij} \tag{4}$$

$$Y_{ij\_Year} = \text{Year}_i + \epsilon_{ij} \tag{5}$$

where Y_{ij} represents the yields, Treatment_i is the effect of the ith treatment, AEZ_i is the effect of the ith AEZ, Year_i is the effect of the ith year, and the error (ϵ_{ij}).

Since the objective was to determine if the variability in maize yield was related and could be explained by the variables listed in Table 1 and maize variety type, a LMM was chosen. According to Dargie et al. (2022), LMM accounts for sample size imbalance and confounding effects of uncontrolled variables, as in our case. So, maize yield was modelled in three majors components: a linear function with a fixed effect trend "$f()_i$" with ith level of the explanatory variable (environmental factors — ENV, soil physical properties — SPP, soil chemical properties — SCP, maize variety type — VAR, fertilizer rate — FER), where the intercept was allowed to vary by AEZ_j (random effect of the jth AEZ) and Year_k (random effect of the ith year) in LMM 6_{ENV}, 7_{SPP}, 8_{SCP}, 9_{VAR}, 10_{FER}, 11_{step}, and only by Year_j (random effect of the jth year) in LMM 12_{ENV_AEZ}, 13_{SPP_AEZ}, 14_{SCP_AEZ}, 15_{VAR_AEZ}, and 16_{FER_AEZ}, with an error term (ϵ_{ijk} or ϵ_{ij}) denoting the small-scale fluctuations around "$f()_i$". The LMMs 11_{step} and 17_{step_AEZ} were optimized by backward selection of variables using the lmerTest R package (Kuznetsova et al., 2017) "step()" function. At the level of each AEZ, LMM 17_{step_AEZ} was used to reveal the AEZ's variables that were linked to maize yield variability. As a result, $Y_{ij\_AEZ\_step}$ represented the maize yield modelled in a specific and known AEZ (i.e., GS, TZ, or DF). However, stepwise LMM 17_{step_AEZ}, in DF, showed that the year's ICC was 0%. Therefore, LMM 17_{step_AEZ} was converted into a Multiple Linear Regression (MLR $_{step\_DF}$) formed with the trend function "$f()_i$" and the error term (ϵ_{ij}). To avoid confusing models, GS, TZ, or DF was assigned as a subscript to those LMMs or MLRs that referred to yield modelling in a particular AEZ, and all LMM where the AEZs were not assigned as a subscript referred to the yield modelling across the entire set of data points. The LMMs were constructed as follows:

$$(\text{LMM 6ENV})\ \ Y_{ijk\_ENV} = f(\text{ENV})_i + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \tag{6}$$

$$(\text{LMM 7SPP})\ Y_{ijk\_SPP} = f(\text{SPP})_i + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \tag{7}$$

$$(\text{LMM 8SCP})\ Y_{ijk\_SCP} = f(\text{SCP})_i + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \tag{8}$$

$$(\text{LMM 9VAR})\ Y_{ijk\_VAR} = f(\text{VAR})_i + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \tag{9}$$

$$(\text{LMM 10FER})\ Y_{ijk\_FER} = f(\text{FER})_i + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \tag{10}$$

$$(\text{LMM 11step})\ Y_{ijk\_step} = f(\text{ENV}, \text{SPP}, \text{SCP}, \text{VAR}, \text{FER})_i + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \tag{11}$$

$$(\text{LMM 12ENV\_AEZ})\ Y_{ij\_ENV\_AEZ} = f(\text{ENV})_i + \text{Year}_j + \epsilon_{ij} \tag{12}$$

$$(\text{LMM 13SPP\_AEZ})\ Y_{ij\_SPP\_AEZ} = f(\text{SPP})_i + \text{Year}_j + \epsilon_{ij} \tag{13}$$

$$(\text{LMM 14SCP\_AEZ})\ Y_{ij\_SCP\_AEZ} = f(\text{SCP})_i + \text{Year}_j + \epsilon_{ij} \tag{14}$$

$$(\text{LMM 15VAR\_AEZ})\ Y_{ij\_VAR\_AEZ} = f(\text{VAR})_i + \text{Year}_j + \epsilon_{ij} \tag{15}$$

$$(\text{LMM 16FER\_AEZ})\ Y_{ij\_FER\_AEZ} = f(\text{FER})_i + \text{Year}_j + \epsilon_{ij} \tag{16}$$

$$(\text{LMM 17step\_AEZ})\ Y_{ij\_AEZ\_step} = f(\text{ENV}, \text{SPP}, \text{SCP}, \text{VAR}, \text{FER})_i + \text{Year}_j + \epsilon_{ij} \tag{17}$$

Parameter estimation for the variance components in the LMMs was done with the REstricted Maximum Likelihood (REML) approach (Searle et al., 1992) using the "lmer()" function in the lme4 R package (Bates et al., 2015). To assess the significance of LMM and MLR fixed effects and also to accommodate imbalances in data point sizes among some predictors, an ANOVA table with F-tests and *p*-values, using Kenward-Roger's method for denominator degrees-of-freedom and F-statistic, was used (Spilke et al., 2005). R^2 statistics for mixed-effects model from Nakagawa et al. (2017)' were used to assess the goodness of fit of the LMMs and MLR $_{step\_DF}$ fixed effect trend "$f()_i$". The predictors groups, i.e., ENV, SPP, SCP, VAR, and FER, involved in the fixed-effect statistics modellings, except for LMMs 11_{step} and 17_{step_AEZ} and MLR_{step_DF}, were compared using the marginal R^2 (R^2_m) to explain the maize yield variability related to the trend function (*f*) and also to rank the predictor groups according to their explanatory power.

2.7. Random Forest model

The RF regression model was trained to (*i*) identify, prioritize, and rank the variables in Table 1 that were most important in explaining the variability of maize yields across AEZs, and (*ii*) predict maize yields within each AEZ based on the most important factors highlighted in (*i*). The measure of variable importance worked by calculating the increase in RF's prediction error after permuting the variables in eq. 18. RF is a non-parametric modelling machine learning technique that has been gaining popularity in agricultural data analysis; it is resistant to outliers and can better handle both straightforward linear and complex nonlinear associations compared to LMM/MLR (Han and Kim, 2021).

$$Y_{i\_rf} = f(\mathrm{ENV}, \mathrm{SPP}, \mathrm{SCP}, \mathrm{VAR}, \mathrm{FER})_i + \epsilon_i \tag{18}$$

To build the RF model, a ratio of 75:25 was used to split the complete dataset into training and test datasets, meaning that 75% of the data point-AEZ combinations were used as training data and 25% of the data point-AEZ combinations were used as a separate test dataset. The data was divided so that each AEZ from the training dataset and its matching AEZ from the test set had a maize yield distribution that was similar. To do so, the caret R package (Probst et al., 2019) "createDataPartition()" function was used.

Variable importance ranking and yield prediction were done using the Ranger R package (Wright and Ziegler, 2017). To prevent the model from being overfitted, the number of variables was reduced for RF modelling (by reducing noise), thus a 10-fold cross-validation with 3 repeats was used to enhance the effectiveness of the variable elimination strategy. This was implemented using the "rfe()" and "rfeControl()" functions of the caret R package. The "rfe()" function applies a backward selection process to find the optimal combination of variables that were most relevant in predicting yield. The RF model fitted on the reduced variables was fine-tuned using a grid search approach to select the best hyperparameters for the model, thus the "mtry" varied between 1 and 8, the "nodesize" ranged from 1 to 100 at an interval of 5, and the "ntree" ranged from 1 to 500 at an interval of 5.

RF model performance was evaluated using the coefficient of determination (R^2) (Krause et al., 2005), accuracy using the root mean square error (RMSE) (Krause et al., 2005), and efficiency using the Nash-Sutcliffe coefficient (NSE) (Nash and Sutcliffe, 1970). The linear regression between the observed maize yield and that predicted by the RF was visualized using a 1:1 plot. All data manipulation, model simulations, analysis, and visualization were performed in RStudio©software (RStudio Team, 2022).

3. Results

3.1. Cropping system characteristics

In soils across AEZs, sand predominated over silt and clay, which were present in roughly equal amounts (Supplementary Table S 3). DF soils contained by far the greatest levels of soil organic carbon (SOC), total phosphorus (TOTP), and cation exchange capacity (CEC) than TZ and GS soils. Across the AEZs, potassium (K) and phosphorus (P) were generally present in comparable amounts but at levels relatively below optimal values for P (Kugbe et al., 2019; Daniel et al., 2021) and at a good level for K (Antwi et al., 2016). In addition, soil pH was around 6 and fluctuated only very slightly from one AEZ to another (Bationo et al., 2018). Generally, the data points with the deepest root zones (RootDEP) on average were observed in soils of the TZ, and the data points with the shallowest soils were in the GS (Supplementary Table S 3), as also reported by Bationo et al. (2018).

The coolest average temperatures — MMeT (26 °C) were observed in DF and TZ; however, GS registered the highest average amount of precipitation during the maize growing season (Supplementary Table S 3). In general, the highest land elevations (ELV) were in DF at an average of 270 m, followed by TZ at an average ELV of 247 m.

Across AEZs, the average rate of nitrogen fertilizer (NF), phosphorus fertilizer (PF), and potassium fertilizer (KF) was 67 kg ha^{-1}, 30 kg ha^{-1}, and 29 kg ha^{-1}, respectively. The highest average NF rate was in DF at 80 kg ha^{-1} (SupplementaryTable S 3). On the other hand, the average PF (34 kg ha^{-1}) and KF (33 kg ha^{-1}) rates were higher in GS.

Open-pollinated varieties (OPVs) accounted for 69% of the maize varieties across all data points, with Obatanpa being the most widely OPV variety (Supplementary Table S 4). GH 110, Etubi, Mamaba, and Pannar53 were the only 4 hybrid varieties, representing 20.5% of all data points. Other variety types, categorized as "Others," included "QPM," "Entry," and "Local variety" and accounted for 10.5% of all data points.

3.2. Maize yield characteristics

Considering all data points, yields varied from 11 kg ha^{-1} to 8.2 t ha^{-1}, with an average of 2.2 t ha^{-1}, similar to those reported by SRID/MoFA (2021). The main effect of treatments was significant and large ($p < 0.05$) (Supplementary Table S 5). Tukey's HSD test found that the mean yield was significantly different between fertilized and control data points ($p < 0.05$) (Supplementary Table S 5, Fig. 2 A).

The main effect of AEZ was statistically significant and medium ($p < 0.05$), and the interaction between treatment and AEZ was also statistically significant but weak ($p < 0.05$) (Supplementary Table S 5). Mean yield was significantly ($p < 0.05$) different between GS and TZ data points and between GS and DF data points, but not different between DF and TZ data points. In DF, mean yield was not significantly ($p < 0.05$) different between fertilized and control data points; however, mean yield was significantly ($p < 0.05$) different between fertilized and control data points in TZ and GS (Supplementary Table S 5, Fig. 2 B).

In the one-way ANOVA with random effects, yield varied with treatment, AEZ, and year. The ICC indicated that the random effects (i.e., treatments, AEZs, and years) accounted for 46%, 20%, and 33% of the total variation in yield, respectively (Supplementary Table S 6, S 7 and S 8).

3.3. QUEFTS model yield estimated

When considering all the data points, the average estimated yield from the QUEFTS model was 3.3 t ha^{-1}. A linear regression between observed and QUEFTS-estimated yields, based on the entire data points involved in the simulation (size effects of pH, SOC, P, TOTP, TOTN, K, NF, NP, and KF were confounded), revealed that soil chemical properties and fertilizer rate explained only 19% ($r = 0.44$, $p < 0.05$) of total yield variability (Fig. 3 A). Fig. 3 shows the linear regression between observed and QUEFTS-estimated yields in GS, TZ, and DF (size effects of pH, SOC, P, TOTP, TOTN, K, NF, NP, and KF were confounded in each AEZ). Indeed, there were weak correlations between observed and QUEFTS-estimated yields in all 3 AEZs (R^2_{DF} [4%] < R^2_{TZ} [19%] < R^2_{GS} [30%], Fig. 3B, C, D). In DF, observed and QUEFTS-estimated yields were negatively correlated ($r = -0.21$, $p < 0.05$), but were positively correlated in GS ($r = 0.55$, $p < 0.05$) and TZ ($r = 0.43$, $p < 0.05$).

Furthermore, when yields from the fertilized (size effects of pH, SOC, P, TOTP, TOTN, K, NF, NP, and KF were confounded) and non-fertilized (size effects of pH, SOC, P, TOTP, TOTN, K were confounded) trials were

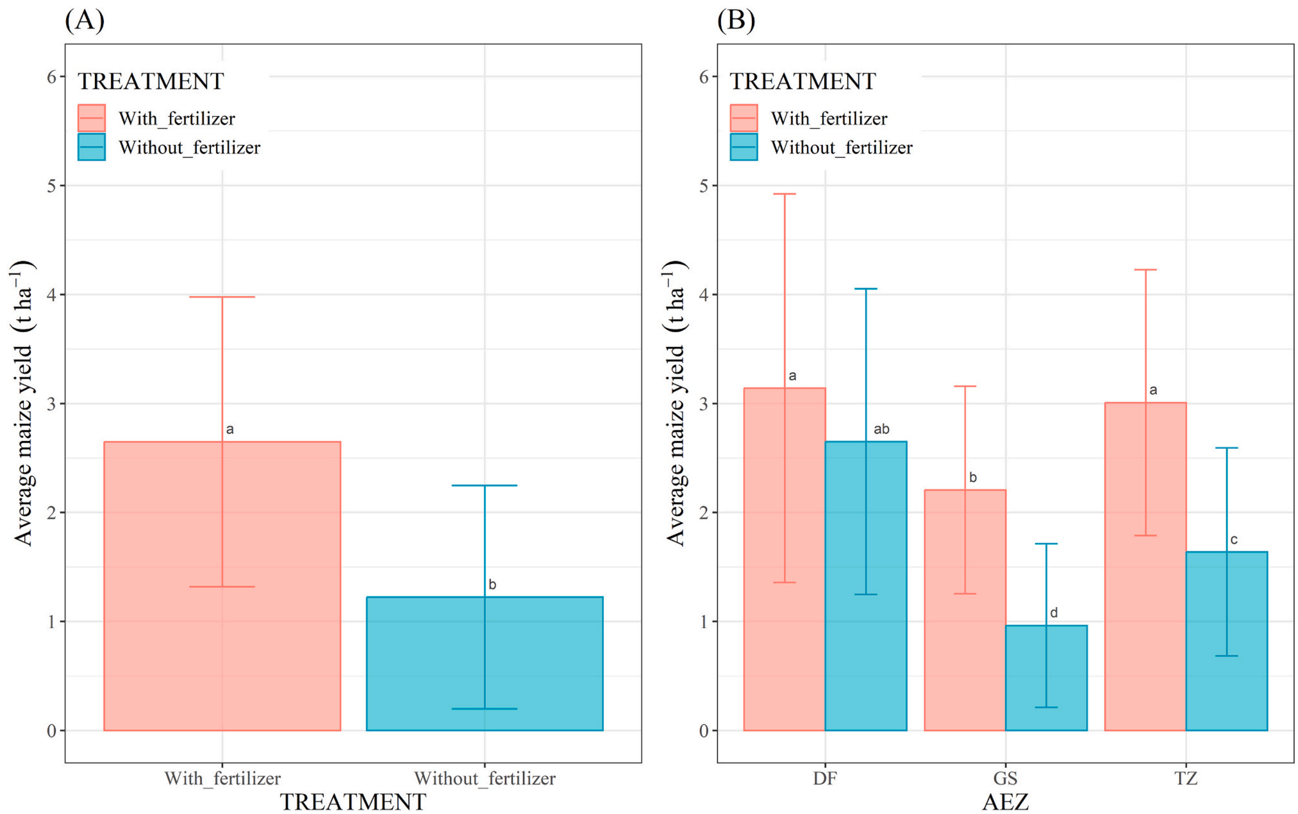


Fig. 2. Average maize yield (t ha^{-1}) of treatments with fertilizer and without fertilizer (A) and the interaction between treatments and agro-ecological zone (AEZ) (B). DF (Deciduous Forest), GS (Guinea Savanna), and TZ (Transition Zone). Error bars are standard deviation bars. Different letters (fisher letters) indicate significant difference ($p < 0.05$).

simulated separately, linear regression between observed and QUEFTS-estimated yields showed that QUEFTS explained only 2% ($r = 0.15$, $p < 0.05$) of the variability in yield in the fertilized data points and 8% ($r = 0.28$, $p < 0.05$) in the non-fertilized data points (Supplementary Fig. S 2 A, B).

At the annual level, in 2001, 2002, 2007, 2008, 2010, 2011, and 2012, linear regressions between observed and QUEFTS-estimated yields showed a good and statistically significant positive correlation (Supplementary Fig. S 2). In 2008, the effects size of soil chemical properties and fertilizer rate, which were confounded, explained 75% of the variability in maize yield. However, the QUEFTS model over-estimated maize yield in most simulation scenarios.

3.4. Modelling maize yield variability across all data points

Variability in maize yield was greatly associated with soil chemical properties (SCP) across the 3 AEZs. In the LMM, yield varied significantly ($p < 0.05$) with SOC, P, and K (Table 2Table 2, Supplementary Table S 9). A correlation matrix and a plot of yield on the CEC did not show a clear trend (Supplementary Fig. S 3 a, d). In addition, pH was not a significant predictor of yield in LMM 11_{step} but was significantly ($p < 0.05$) related to yield in LMM 8_{SCP}, and a plot of yield on pH also revealed a clear and negative trend (Supplementary Fig. S 3 a, c). Compared to P, K, pH, and CEC, the variation in yield due to SOC was greater (F-value = 35.3) (Table 2).

The LMM revealed that, among soil physical properties (SPP), CLAY and RootDEP were significantly ($p < 0.05$) related to the variation in maize yield (Table 2Table 2, Supplementary Table S 9). In addition, there was a positive and significant ($p < 0.05$) trend between maize yield with SAND, CLAY, and RootDEP according to the correlation matrix (Supplementary Fig. S 3 a).

In the LMM analysis, MP was the only significant ($p < 0.05$) environmental predictor (ENV) of yield (Supplementary Table S 9), but the correlation matrix showed that MMeT and ELV were likewise significantly ($p < 0.05$) related to the variation in maize yield (Supplementary Fig. S 3 a). A plot of maize yield on MP showed a significant ($p < 0.05$) and clear negative trend (Supplementary Fig. S 3 b).

Maize variety type (VAR) was significantly ($p < 0.05$) related to yield (Table 2Table 2, Supplementary Table S 9). Across all data points, a one-way ANOVA revealed that there was a significant ($p < 0.05$) difference in mean yield between at least two variety types (Supplementary Table S 10). Tukey's HSD test found that there was no significant difference in mean yield between OPV and hybrid ($p = 0.54$), but there was a significant ($p < 0.05$) difference in mean yield between "Others" and OPV and between "Others" and hybrid varieties.

Across all data points, the fertilizer rates (FER)were significantly ($p < 0.05$) associated with maize yield variability (Table 2, Supplementary Table S 9). The variation in yields with NF rate was the largest of all the variables. A comparison LMMs based on R^2_m values revealed that FER had the highest explanatory power ($R^2_m = 0.17$) of maize yield variability, followed by the SCP ($R^2_m = 0.16$), the VAR ($R^2_m = 0.05$), ENV ($R^2_m = 0.03$) and finally the SPP ($R^2_m = 0.02$) (Supplementary Table S 9). The ICC values in Table 2 and Supplementary Table S 9 indicate that AEZ and year also played important roles in the variance of total maize yield across all the data points, with AEZ having the largest ICC.

3.5. Maize yield determinants for each agro-ecological zone

In GS, LMM demonstrated that maize yield significantly ($p < 0.05$) varied with ELV, CLAY, SOC, VARIETY, NF, PF, SAND, and CEC (Table 3Table 3, Supplementary Table S 11). However, the contributions of ELV and CEC to the variations in maize yield in GS did not show a clear and significant trend (Supplementary Fig. S 4 a, b, e). The PF rate accounted for 42% of the variation explained from the type 2 sum of

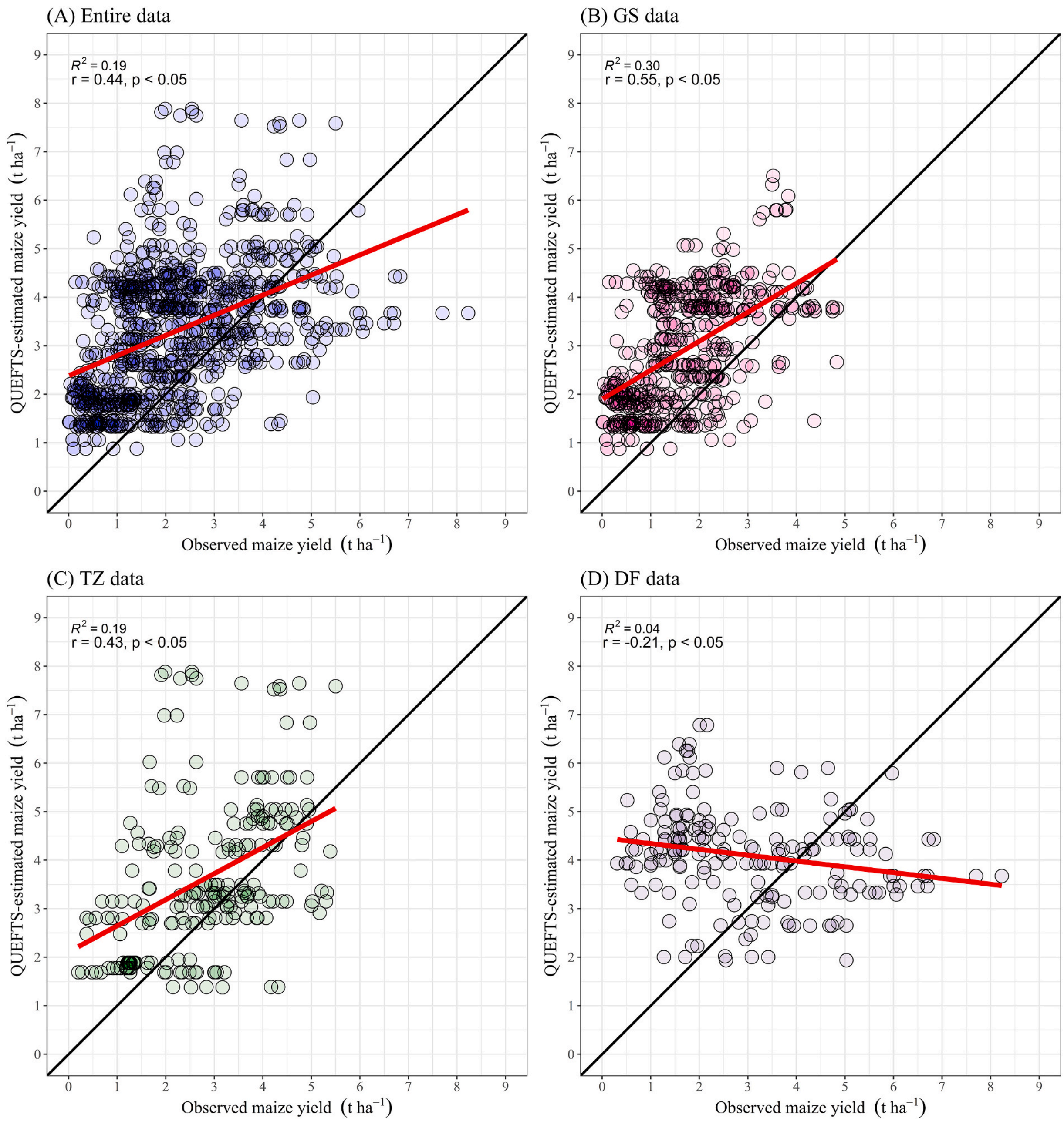


Fig. 3. Relationship between observed and QUEFTS-estimated maize yield (t ha^{-1}) across all data points (A) and per AEZ (B, C, D). The R^2 indicates the portion of yield variability explained by soil chemical properties and fertilizers (confounded effect). The bold red line represents the linear regression line, and the fine black line from left to right is the 1:1 line. As the points overlapped, the ggplot2 R package (Wickham, 2016) function "Jitter" was applied for easier visualization. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Significance of effects of explanatory variables in linear mixed effects modelling of the set of data point yields.

Model	Variable	Sum Sq	DF	Den DF	F-value	Pr(>F)	R^2_m	ICC	
								AEZ	Year
	CLAY	5.4	1	959.9	12.3	<0.05			
	SOC	15.6	1	717.8	35.3	<0.05			
	P	3.7	1	889.5	8.4	<0.05			
	K	2.3	1	817.9	5.1	<0.05			
	pH	1.7	1	466.9	3.8	0.05			
	CEC	7.1	1	660.1	16.0	<0.05			
	VARIETY	53.8	2	960.4	60.9	<0.05			
	NF	84.9	1	957.9	192.3	<0.05			
LMM 11_{step}[†]	PF	27.4	1	963.5	61.9	<0.05	0.26	0.53	0.17

† Entire set of data points of maize yield modelled with Eq. (11) using stepwise linear mixed effects modelling.

squares (Table 3); therefore, GS was found to have a larger significant yield fluctuation due to PF rate than the other AEZs. Random effects (i.e., year) accounted for 2–32% of the total variation in maize yield in GS. The highest explanation power of yield variability ($R^2_m = 0.35$) in GS was shown by FER, followed by VAR ($R^2_m = 0.05$), SPP and SCP ($R^2_m = 0.03$), and ENV ($R^2_m = 0.02$).

In TZ, LMM analysis showed that maize yield variability was significantly ($p < 0.05$) related to MMeT, ELV, CLAY, VARIETY, SOC, NF, and PF (Table 4, Supplementary Table S 12). The combined contribution of ELV, CLAY, VARIETY, NF, and PF explained 55% of the variation in maize yield in TZ, whereas the year's ICC was 37% (Table 4Table 4). Fertilizer rates had the best explanatory power of maize yield variability ($R^2_m = 0.26$) in TZ, followed by the ENV ($R^2_m = 0.21$), then VAR ($R^2_m = 0.20$), SCP ($R^2_m = 0.19$), and finally SPP ($R^2_m = 0.19$).

In the MLR analysis, ELV, MP, SOC, CLAY, SAND, RootDEP, TOTP, P, K, CEC, VARIETY, NF and PF were significant ($p < 0.05$) predictors of maize yield in DF, explaining 79% of maize yield variability (Table 5). The analysis also revealed that maize yield significantly ($p < 0.05$) varied with K but was not significantly related to the KF rate (supplementary Table S 13). The ICC indicated that the random effects (i.e., year) accounted for 4–64% of the total variation in maize yield in DF. NF and PF rates and maize variety did not exhibit the same level of explanatory power of yield variability as in GS and TZ. Indeed, the strongest explanatory power for the variation in yield in DF was found in the SCP ($R^2_m = 0.48$), followed by ENV ($R^2_m = 0.43$), VAR ($R^2_m = 0.32$), and SPP ($R^2_m = 0.24$).

LMM analysis showed that the contribution of maize variety type in the variations in yield in the 3 AEZs was also significant ($p < 0.05$) (Tables 3–5, Supplementary Table S 14, S 15, S 16). A one-way ANOVA revealed that there was a significant ($p < 0.05$) difference in mean yield between at least two variety types in GS, TZ, and DF (Supplementary Table S 14, S 15, S 16). Tukey's HSD test found that there was no significant difference in mean yield between OPV and hybrid in GS ($p = 0.25$) and TZ ($p = 0.69$), but there was significant difference in mean yield in DF ($p < 0.05$). There was a significant ($p < 0.05$) difference in mean yield between OPV and "Others" variety types in GS, TZ, and DF, and also between hybrid and "Others" variety types in the 3 AEZs.

3.6. Modelling maize yield using Random Forest

Considering 75% of all data points, the results of the RF model showed that the NF rate explained the largest portion (26.7%) of the variability in maize yield. The RF model revealed that 4 variables (NF, MMeT, RootDEP, and MP) accounted for nearly 61% of the variability in maize yield (Supplementary Fig. S 7 A). In addition, a strong importance of RootDEP in the entire maize yield variability was spotlighted, whereas LMM yield modelling did not.

The R^2 ranged between 0.71 (training data) and 0.75 (testing data), the RMSE ranged between 750 kg ha^{-1} (training data) and 700 kg ha^{-1} (testing data), and the NSE ranged between 0.58 (training data) and 0.64 (testing data), indicating that the RF model performed well in yield prediction (Fig. 4A, Supplementary Fig. S 7 B). Based on the importance of the variables in explaining yield using the training data points, maize yield prediction was made in each AEZ (Figure 4Figure 4 B, C, D). In comparison to TZ and DF, the RF model's performance in GS was somewhat subpar. Indeed, in GS, RF-predicted maize yield showed the lowest values of NSE = 0.55 and $R^2 = 0.74$.

4. Discussion

4.1. Soil chemical properties and fertilizer rates impacting yield variability

The QUEFTS model was calibrated and validated for the northern regions (i.e., the GS) and explained more variability there than in the other 2 AEZs. Yet, the low spatial and temporal explanatory power of QUEFTS suggests that other factors than soil chemical properties and fertilizer rates only, contribute to the variability in maize yield as also reported by Debtanu et al. (2006) and Onduru and Du Preez (2007). Previous studies in Kenya, Benin, and Rwanda have also shown low precision and accuracy of QUEFTS maize yield predictions (Mulder, 2000; Onduru and Du Preez, 2007; Breure et al., 2022). While soil properties do not reveal strong temporal dynamics, rainfall, and temperature in 2001 (MP = 502–698 mm, MMaT = 27–29°) and 2008 (MP = 709 mm, MMaT = 30°) were favorable better mimicking the presumed growth conditions in QUEFTS when its explanatory power was indeed highest ($R^2 > 50\%$). In addition, the importance of soil physical properties, such as texture and soil depth, are also not considered in QUEFTS.

The LMM/MLR and RF models unveiled the importance of fertilizer rate (FER) and soil chemical properties (SCP) in maize yield variability as well, as corroborated by Braimoh and Vlek (2006). Of the nine factors that were significantly ($p < 0.05$) associated with maize yield variability in the LMM analysis (Table 2), 5 were related to the SCP. The NF rate was found to be the most significant ($p < 0.05$) contributor to the variation in maize yield across the 3 AEZs in Ghana. Extractable P and K demonstrated a positive trend in the main effect among these soil chemical variables (Fig. S 3 a), also identified by Braimoh and Vlek (2006), Yeboah et al. (2016) and Akolgo et al. (2020).

Ghanaian soils are generally not very acidic (Supplementary Table S 3) and fairly optimal, however the slightly negative relation with maize yield likely results from the narrow pH range and associations with other variables such as temperature (Supplementary Fig. S 3 a, c, 5 a). With tropical soils like in the DF having typical CEC values of around 10 cmol$_{(+)}$ kg^{-1} (Osei, 1995), the high values (83 cmol$_{(+)}$ kg^{-1}) of the 30 data points in the DF suggested an negative though not significant ($p = 0.09$) relation with yield when the entire data point is considered Supplementary Fig. S 3 d). Excluding these data points turned the relation significantly ($p < 0.05$) positive (Supplementary Fig. S 3 e).

The LMM analyses revealed that compared to NF and PF rates, the KF rate was not a strong predictor of maize yield in any of the AEZs. LMM and RF modelling of yields with all data points also confirmed this result, with KF explaining 2.5–3.5% of maize yield variability. The significant ($p < 0.05$) association of K and maize yield in Table 2 was probably due to the data points of the DF (Supplementary Table S 13, Supplementary Fig. S 6 a). According to Yawson et al. (2011), Ghana's forest soils will

Table 3
Significance of the effects of the explanatory variables in the linear mixed effects modelling of the set of yield data points from Guinea Savanna.

Model	Variable	Sum Sq	DF	Den DF	F value	Pr(>F)	R^2_m	ICC Year
LMM 17_{step_GS}[†]	ELV	4.6	1	462.3	9.1	<0.05	0.39	0.02
	SAND	3.7	1	350.4	7.9	<0.05		
	CEC	6.7	1	556.0	13.2	<0.05		
	VARIETY	15.3	2	549.4	14.8	<0.05		
	NF	8.7	1	554.5	16.8	<0.05		
	PF	28.6	1	551.1	55.5	<0.05		

[†] Maize yield modelled with Eq. (17) in Guinea Savanna using stepwise linear mixed effects modelling.

Table 4
Significance of effects in stepwise linear mixed effects modelling using data points from the Transition Zone only.

Model	Variable	Sum Sq	DF	Den DF	F value	Pr(>F)	R^2_m	ICC
								Year
	ELV	8.2	1	6.1	23.1	<0.05		
	CLAY	19.5	1	25.9	54.8	<0.05		
LMM 17_{step_TZ}[†]	VARIETY	37.6	2	143.3	52.7	<0.05	0.55	0.37
	NF	44.9	1	215.4	125.9	<0.05		
	PF	4.2	1	215.9	11.7	<0.05		

[†] Maize yield modelled with Eq. (17) in the Transition Zone using stepwise linear mixed effects modelling.

Table 5
Significance of the effects of the explanatory variables in the linear mixed effects modelling of the set of yield data points from the Deciduous Forest.

Model	Variable	Sum Sq	DF	F value	Pr(>F)	R^2
	ELV	1.6	1	7.4	<0.05	
	MP	2.1	1	9.6	<0.05	
	CLAY	4.2	1	19.3	<0.05	
	SAND	4.4	1	20.3	<0.05	
	RootDEP	1.7	1	8.8	<0.05	
	SOC	2.2	1	10.3	<0.05	
MLR_{step_DF}[†]	TOTP	2.9	1	13.8	<0.05	0.79
	P	2.3	1	10.9	<0.05	
	K	5.9	1	27.4	<0.05	
	CEC	4.2	1	19.6	<0.05	
	VARIETY	37.0	2	86.0	<0.05	
	NF	11.7	1	54.5	<0.05	
	PF	0.7	1	3.4	<0.05	

[†] Maize yield modelled in Deciduous Forest using stepwise multiple linear regression model.

require frequent and split KF applications since they have a low capacity to maintain a long-term supply of K; however, the savanna soils will require less frequent but higher K fertilization to satisfy the exchangeable pool and immediate plant nutrition requirement. Different KF rates were not significantly related to maize yield, probably due to Liebig's Law of the Minimum (Essel et al., 2020). Nevertheless, application of KF will remain important to prevent soil nutrient mining.

We observed significant but downward linear trends in the main effect for SOC in 2 AEZs (Supplementary Fig. S 5 a, S 6 a). The bulk of high-yielding data points in TZ and DF were found in soils with a comparatively lower SOC. This negative trend suggests yield variability to be driven by other factors as well. In research by Kihara et al. (2016), Shehu et al. (2019) and Sileshi et al. (2022), maize fertilization trials on soils with a relatively high SOC level obtained poor yields, while highest yields were found on soils with a low SOC content. Across the entire set of data points, with a wider SOC range, Supplementary Fig. S 3 f revealed a significant ($p < 0.05$) positive trend of maize yield on SOC. Logah et al. (2011) and many other scholars found the same result (Lal, 2006; Solomon et al., 2016; Owoade et al., 2020; Owoade et al., 2021). Therefore, SOC is used as an indicator of soil fertility (Tittonell et al., 2008), with application of NF in such soil promoting maize yield.

According to Tables 3 and 5, in GS and DF respectively, PF had the highest (F-value = 55.5) and lowest (F-value = 3.4) contributions to maize yield variability. This might explain why lower rates of PF (19 kg ha^{-1}) are applied in the DF as opposed to 34 kg ha^{-1} in the GS. The levels of soil P and TOTP in DF were greater than those in GS (Supplementary Table S 3), which may have prevented the PF rate in DF soils from strongly contributing to maize yield.

4.2. Role of soil physical characteristics on yield variability

Only RootDEP demonstrated a significant ($p < 0.05$) positive trend with yield variation when the entire set of maize yield data points was using in LMM. It was indeed identified by the RF model as the third most important variable, right after the NF rate and MMeT (Supplementary Fig. S 7 A). A more specific analysis showed that this influence originated mostly from the DF data points (Supplementary Table S 13, Fig. S 6 a) with deeper soils. Guilpart et al. (2017) revealed this importance of RootDEP on the sensitivity of rainfed maize yields as function of the water-holding capacity of the root zone.

Sand and clay were found to be significantly ($p < 0.05$) associated with maize yield variability in both LMM and MLR modelling. While a negative relation of sand content is expected with yield due to various associated soil properties like low water-holding capacity, the positive trend (Supplementary Fig. S 3 a, S 4 a, c, and S 6 a, g) could be related to the influence of organic fertilizers (poultry and cow manure) in this study (Dapaah et al., 2008; Quansah, 2010; Adjei-Nsiah, 2012; Kanton et al., 2016; Badu et al., 2019). According to Obi and Ebo (1995), Zingore et al. (2007), Uzoma et al. (2011) and Frimpong et al. (2021), the application of organic fertilizers such as cow and poultry manure on sandy soil significantly improves the physicochemical properties of sandy soils favorably affecting maize growth. Moreover, the GS has a higher potential for maize production (Boullouz et al., 2022) that can be better exploited with appropriate fertilization and water management practices; organic manure being one of these favorable practices.

The contribution of clay content to the variations in maize yield was significant (Table 2). Clay plays an important role in the supply, retention, and fixation of many macronutrients and micronutrients in the soil that improve maize crop nutrition (O'Halloran et al., 1985; Batjes, 2011; Florence et al., 2017), and ameliorates soil physical characteristics such as enhancing water holding capacity. In GS, however, a sizable downward trend in the maize yield with clay was observed (Supplementary Fig. S 4 d). While the correlation coefficient was significant ($p < 0.05$), the distribution of data points in the Supplementary Fig. S 4 d does clearly underline this negative trend, likely for similar reasons as indicated for sand. Sileshi et al. (2022) reports similar results, and Njoroge et al. (2018) reported no significant difference in clay content between sites with high and low yield under NPK application in western Kenya.

4.3. Contribution of environmental factors to yield variation

The analysis of LMM (Table 2, Supplementary Table S 9, S 11, S 12, S 13, Fig. S 3 a, b, S 5 b, S 6 a) and RF yield modelling (Supplementary Fig. S 7 A) showed that environmental variables (MMeT, MP, and ELV) were related to maize yield variability, as also reported by Onduru and Du Preez (2007), Mugwe et al. (2009), Fosu-Mensah et al. (2019), Kyei-Mensah et al. (2019), and Cudjoe et al. (2021). However, the MP of the entire data point reveals a negative trend with yield (Supplementary Fig. S 3 a, b). In the DF where very high yields of 7–8 t ha^{-1} were reported, the amount of rainfall was sufficient, though lower compared to other areas with lower yields. Their larger water holder capacity of the deeper soils (RootDEP of 150 mm) made sufficient water available for maize growth, as also reported by Durodola and Mourad (2020) and Bagula et al. (2022). The negative trend between MP and RootDEP supports this logic and why high MP in some locations with low RootDEP did not result in high maize yields (Supplementary Fig. S 6 a, h).

Baffour-Ata et al. (2021) reported that temperature was significantly and positively related to maize yield in some regions of Ghana. This result was also observed in our analysis in TZ (Supplementary Fig. S 5 a,

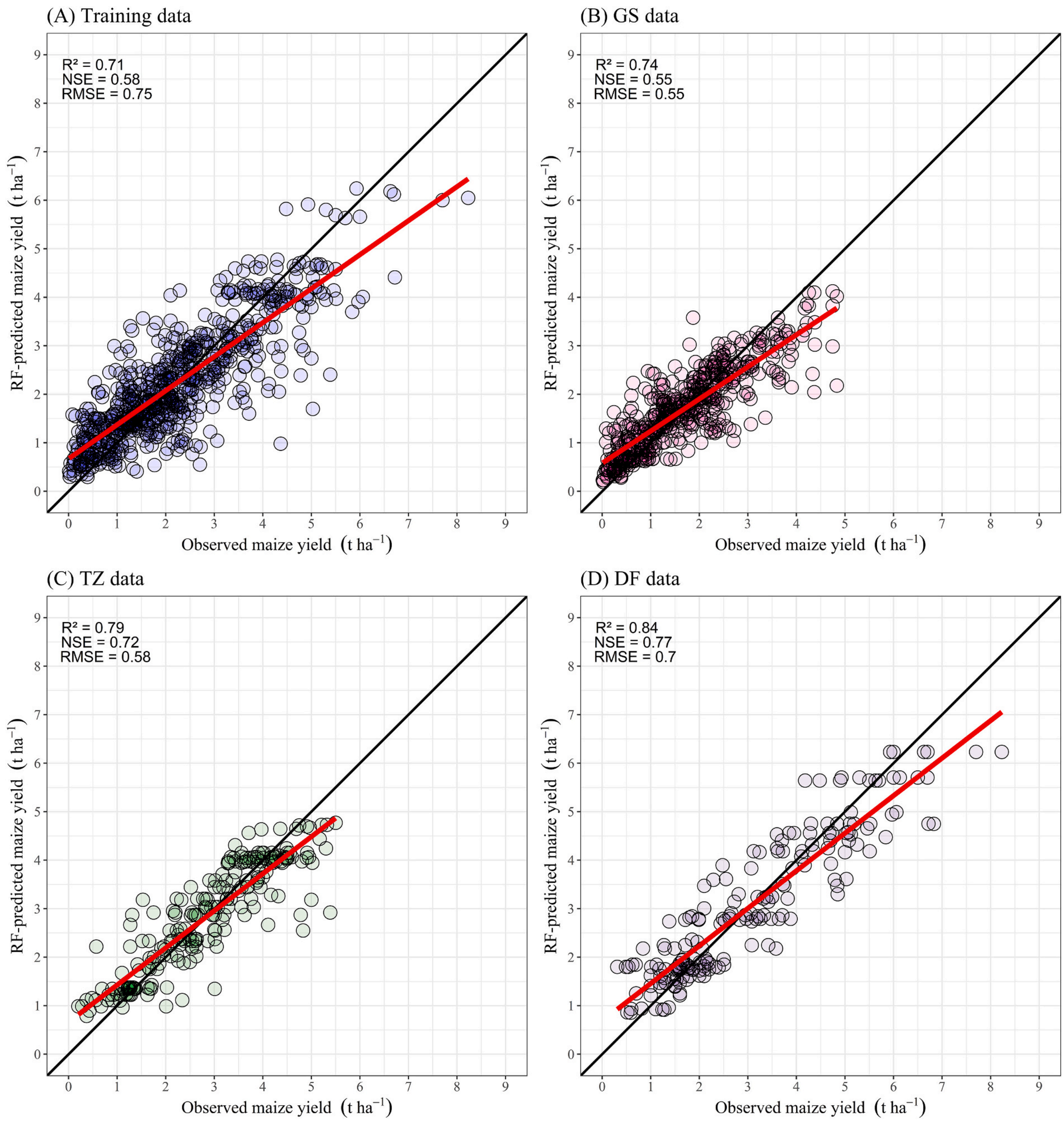


Fig. 4. Relationship between observed maize yield and that predicted by the Random Forest (RF) model using 3 repeats of a 10-fold cross-validation approach. (A) The linear regression on 75% of the entire data point (training data point), (B) only the data point from GS, (C) only the data point from TZ, and (D) only the data point from DF. The solid red lines show the linear regressions fitted to the data point, and the fine black line from left to right is the 1:1 line, with the coefficient of determination (R^2), root mean square error (RMSE), and Nash-Sutcliffe model coefficient efficiency (NSE). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

b). However, according to Bationo et al. (2018), high air temperatures, high light levels, and heat-trapping sandy soils combine to make the local environment too hot for good plant growth. Therefore, the positive relation found in TZ may have resulted from improved high-potential, drought-tolerant variety and fertilizer rate (Atiah et al., 2021). Drought-tolerant maize varieties in some trials, such as GH 110, Mamaba, and Akposoe, could have performed relatively well at higher temperatures.

ELV, in addition to MP and MMeT, was also a significant factor in explaining maize yield variability. In fact, the correlation matrices (Supplementary Fig. S 3 a, S 4 a and S 5a) revealed that ELV and some soil variables, including RooTDEP, CLAY, and SOC content, were positively correlated. Thus, for example, the significant ($p < 0.05$) trend between ELV and RootDEP demonstrates that a high yield is more likely because the higher the ELV, the larger the water pool available to the plant. A study by Jiang and Thelen (2004) also confirmed that ELV is a very useful factor in understanding variation in maize production and that many soil properties are significantly dependent upon ELV (Cooper, 1979; Ovalles and Collins, 1986; Kravchenko and Robertson, 2007).

4.4. *Yield variability related to maize variety*

Overall, the results showed that maize variety type was significantly ($p < 0.05$) related to maize yield variability, as found by Kpotor et al. (2014) also. Bawa and Tang (2021) evaluated the influence of different NF rates on maize yield and found differential varietal responses, as did

Adu et al. (2014) in Ghana, and Abera et al. (2017) in Ethiopia. Afreh et al. (2022) reported that, depending on the interaction between maize variety and NF rate, farmers would need to apply their NF during planting for Omankwa (OPV) and 14–28 days after planting for hybrid varieties to achieve 4.7 t ha^{-1} and 6.5 t ha^{-1}, respectively. The coupled contribution of environmental factors and fertilizer rates could explain the aberrant responses of varieties between the AEZ's. A study conducted in DF and TZ by Kpotor et al. (2014) also revealed substantial interactions between variety, NF rate, and trial site, which were very important in determining maize yield.

4.5. Model performance

Crop yield is intricately influenced by a variety of genetic, environmental, and management factors as well as their interactions. For those involved in agriculture, the ability to precisely predict crop yield in a variety of geographic settings with changing environmental circumstances is becoming more and more crucial (Wang, 2021). However, the identifying of variables influencing maize yield (in Ghana) is highly complex because yields are influenced by interdependent and frequently drastically different climates, soil, and management variables (see also van Loon et al. (2019). Model cross-validation showed that RF performed well in maize yield prediction across AEZs (Fig. 4). The small value of RMSE observed in GS in yield prediction, compared to TZ and DF, could be explained by the low mean yield observed (μ) and the standard deviation (σ) less dispersed in GS ($\mu = 1$ 0.68 t ha-1, $\sigma = 1$) than in TZ ($\mu = 2.82$ t ha-1, $\sigma = 1.3$) and DF ($\mu = 3.05$ t ha-1, $\sigma = 2$).

QUEFTS is a model based on empirical processes and internal interaction (Janssen et al., 1990), whereas statistical and machine learning models (i.e., LMM, MLR, and RF) directly explain and predict yield without many internal processes. This could also justify why LMM/MLR and RF performed better than QUEFTS at both identifying the driving variables of maize yield variability and estimating yield (Bonilla-Cedrez et al., 2021). Additionally, it was demonstrated that RF outperformed in terms of maize yield explanation (R^2) since it is not subject to the same amount of normality assumptions that LMM and MLR are. It was thus possible to clearly illustrate the importance of RootDEP in explaining yield variability. The use of linear models, such as LMM and MLR, alongside RF facilitates the interpretation of how variables are related to maize yield variability in cases where RF has only highlighted and ranked the importance of factors related to maize yield variability and not how these factors vary with maize yield across AEZs. Our multi-model approach used in this study reveals this added value to unravel driving factors of maize yield.

5. Conclusion

Increasing maize yields to meet food demand while increasing farmers' profitability remains a major challenge for Ghanaian farming systems. The overall objective of this study was to characterize rainfed maize yield variation, understand the sources of variability, and predict maize yield using robust statistical and modelling tools. The high variability in yield to fertilizer application, both within and across AEZs and years, reflects a high degree of heterogeneity in soil characteristics, environmental factors, maize varieties, and growing conditions at various spatial and temporal scales. This study provided new information on maize yield variability in Ghana. When the entire set of data points was involved in the yield modelling, LMM/MLR and RF models showed that the NF rates were the most important factors explaining the maize yield variability in Ghana. RF revealed also that the second most important factor was MMeT, followed by RootDEP and MP. Since yield variability was significantly related to AEZ, in DF, soil chemical properties and environmental factors explained most of variability. In TZ and GS, NF and PF drove yield explanation. This suggests that the inherently high soil fertility in DF overrules the importance of fertilizers, while fertilizers drive yield increase in the less-fertile TZ and GS. In all 3 AEZs, the use of different types of maize varieties also played an important role in the overall yield variability observed. Importantly, our multi-model approach that combines advanced statistical methods, crop-soil modelling, and machine learning, reveals its ability to identify drivers for yield despite the huge complexity of the production system.

CRediT authorship contribution statement

Anselme K.K. Kouame: Data curation, Methodology, Formal analysis, Writing – review & editing. **Prem S. Bindraban:** Supervision, Formal analysis, Writing – review & editing. **Isaac N. Kissiedu:** Writing – review & editing. **Williams K. Atakora:** Writing – review & editing. **Khalil El Mejahed:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Data availability

The authors do not have permission to share data.

Acknowledgement

This study received funding from the Université Mohammed VI Polytechnique and the OCP group, through the Fertilizer Research and Responsible Implementation Program, led by the International Fertilizer Development Center (IFDC). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.agsy.2023.103667.

References

Abera, T., Debele, T., Wegary, D., 2017. Effects of varieties and nitrogen fertilizer on yield and yield components of maize on farmers field in mid altitude areas of Western Ethiopia. Int. J. Agronomy 2017, 1–13. https://doi.org/10.1155/2017/4253917.

Adjei-Nsiah, S., 2012. Response of maize (Zea mays L.) to different rates of Palm bunch ash application in the semi-deciduous Forest agro-ecological zone of Ghana. Appl. Environ. Soil Sci. 870948. https://doi.org/10.1155/2012/870948.

Adu, G.B., Abdulai, M.S., Alidu, H., Nustugah, S.K., Buah, S.S., Kombiok, J.M., Obeng-Antwi, K., Abudulai, M., Etwire, P.M., 2014. Recommended Production Practices for Maize in Ghana. https://doi.org/10.13140/2.1.4376.3527.

Adzawla, W., Atakora, W.K., Gouzaye, A., Bindraban, P.S., 2021a. Crop Yield and Fertilizer Use Among Farmers in Guinea Savannah and Transitional Zones of Ghana. https://ifdc.org/wp-content/uploads/2021/07/IFDC-FERARI-Research-Report-No-5-Final.pdf.

Adzawla, W., Atakora, W.K., Kissiedu, I.N., Martey, E., Etwire, P.M., Gouzaye, A., Bindraban, P.S., 2021b. Characterization of farmers and the effect of fertilization on maize yields in the Guinea Savannah, Sudan Savannah, and transitional agroecological zones of Ghana. EFB Bioecon. J. 1 https://doi.org/10.1016/j.bioeco.2021.100019.

Afreh, D.N., Afari, M.A.B., Adjei, R.R., Sarfo Boateng, A., Santo, K.G., Abdulai, M., Popovic, V., 2022. Response of two maize (Zea mays L.) varieties to times of NPK (15-15-15) fertilizer application. Int. J. Agronomy 2022, 1–7. https://doi.org/10.1155/2022/7186913.

Akolgo, G.A., Kemausuor, F., Awafo, E.A., Amankwah, E., Atta-Darkwa, T., Essandoh, E. O., Bart-Plange, A., Maia, C.M.B., d. F., 2020. Biochar as a soil amendment tool: effects on soil properties and yield of maize and cabbage in Brong-Ahafo region, Ghana. Open J. Soil Sci. 10 (03), 91–108. https://doi.org/10.4236/ojss.2020.103005.

Antwi, E.K., Boakye-Danquah, J., Asabere, S.B., Yiran, G.A.B., Loh, S.K., Awere, K.G., Abagale, F.K., Asubonteng, K.O., Attua, E.M., Owusu, A.B., 2014. Land use and landscape structural changes in the ecoregions of Ghana. J. Disaster Res. 9 (4), 452–467.

Antwi, M., Duker, A.A., Fosu, M., Abaidoo, R.C., Pirasteh, S., 2016. Geospatial approach to study the spatial distribution of major soil nutrients in the northern region of Ghana. Cogent Geosci. 2 (1), 1201906. https://doi.org/10.1080/23312041.2016.1201906.

Antwi, A., Duker, A., Fosu, M., Abaidoo, R.C., 2017. Simulation of major soil nutrients requirement for maize production using the QUEFTS model in the northern region of Ghana. Direct Res. J. Agricult. Food Sci. 5 (3), 133–140. Article DRJA43071532. http://directresearchpublisher.org/aboutjournal/drjafs. Article DRJA43071532.

Atiah, W.A., Amekudzi, L.K., Akum, R.A., Quansah, E., Antwi-Agyei, P., Danuor, S.K., 2021. Climate variability and impacts on maize (Zea mays) yield in Ghana, West Africa. Q. J. R. Meteorol. Soc. https://doi.org/10.1002/qj.4199.

Badu, E., Kaba, J.S., Abunyewa, A.A., Dawoe, E.K., Agbenyega, O., Barnes, R.V., 2019. Biochar and inorganic nitrogen fertilizer effects on maize (Zea mays L.) nitrogen use and yield in moist semi-deciduous forest zone of Ghana. J. Plant Nutr. 42 (19), 2407–2422. https://doi.org/10.1080/01904167.2019.1659347.

Baffour-Ata, F., Antwi-Agyei, P., Nkiaka, E., Dougill, A.J., Anning, A.K., Kwakye, S.O., 2021. Effect of climate variability on yields of selected staple food crops in northern Ghana. J. Agricult. Food Res. 6 https://doi.org/10.1016/j.jafr.2021.100205.

Bagula, E.M., Majaliwa, J.-G.M., Basamba, T.A., Mondo, J.-G.M., Vanlauwe, B., Gabiri, G., Tumuhairwe, J.-B., Mushagalusa, G.N., Musinguzi, P., Akello, S., Egeru, A., Tenywa, M.M., 2022. Water Use Efficiency of Maize (Zea mays L.) Crop under Selected Soil and Water Conservation Practices along the Slope Gradient in Ruzizi Watershed, Eastern D.R. Congo. Land 11 (10). https://doi.org/10.3390/land11101833.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bationo, A., Fening, J.O., Kwaw, A., 2018. Assessment of soil fertility status and integrated soil fertility Management in Ghana. In: Bationo, A., et al. (Eds.), Improving the Profitability, Sustainability and Efficiency of Nutrients Through Site Specific Fertilizer Recommendations in West Africa Agro-Ecosystems, vol. 1. Springer International Publishing, pp. 93–138. https://doi.org/10.1007/978-3-319-58789-9_7. AG 2018.

Batjes, N.H., 2011. Global Distribution of Soil Phosphorus Retention Potential (ISRIC Report 2011/06), 42 pp, Issue. https://library.wur.nl/WebQuery/wurpubs/fulltext/185289.

Bawa, A., Tang, Y., 2021. Yield and growth response of maize (Zea mays L.) to varietal and nitrogen application in the Guinea savanna agro-ecology of Ghana. Adv. Agricult. 2021, 1–8. https://doi.org/10.1155/2021/1765251.

Bonilla-Cedrez, C., Chamberlin, J., Hijmans, R.J., 2021. Fertilizer and grain prices constrain food production in sub-Saharan Africa. Nature Food 2 (10), 766–772. https://doi.org/10.1038/s43016-021-00370-1.

Boullouz, M., Bindraban, P.S., Kissiedu, I.N., Kouame, A.K.K., Devkota, K.P., Atakora, W. K., 2022. An integrative approach based on crop modeling and geospatial and statistical analysis to quantify and explain the maize (Zea mays) yield gap in Ghana. Front. Soil Sci. 2, 1037222. https://doi.org/10.3389/fsoil.2022.1037222.

Braimoh, A.K., Vlek, P.L.G., 2006. Soil quality and other factors influencing maize yield in northern Ghana. Soil Use Manag. 22 (2), 165–171. https://doi.org/10.1111/j.1475-2743.2006.00032.x.

Breure, M.S., Kempen, B., Hoffland, E., 2022. Spatial predictions of maize yields using QUEFTS – a comparison of methods. Geoderma 425. https://doi.org/10.1016/j.geoderma.2022.116018.

Bua, S., El-Mejahed, K., MacCarthy, D., Adogoba, D.S., Kissiedu, I.N., Atakora, W.K., Fosu, M., 2020. Yield responses of maize to fertilizers in Ghana (IFDC FERARI n°2), Issue. IFDC. https://ifdc.org/wp-content/uploads/2020/09/FERARI-Policy-Brief-2-Yield-Responses-of-Maize-to-Fertilizers-in-Ghana.pdf.

Cicin-Sain, I., 2018. Population Growth and Food: A Systems Perspective. Retrieved 01 Jully 2021 from. https://sustainablefoodtrust.org/articles/population-growth-and-food-a-systems-perspective/.

Cooper, P.J.M., 1979. The association between altitude, environmental variables, maize growth and yields in Kenya. J. Agric. Sci. 93 (3), 635–649. https://doi.org/10.1017/S0021859600039058.

Cudjoe, G.P., Antwi-Agyei, P., Gyampoh, B.A., 2021. The effect of climate variability on maize production in the Ejura-Sekyedumase municipality, Ghana. Climate 9 (10). https://doi.org/10.3390/cli9100145.

Daniel, K.A.N., Alice, A.A., Thomas, A.-G., 2021. Response of maize (Zea mays L.) to foliar and soil applied fertilizers in the semi-deciduous forest zone of Ghana. Afr. J. Agric. Res. 17 (8), 1114–1122. https://doi.org/10.5897/AJAR2021.15561.

Dapaah, H.K., Ennin, S.A., Asafu-Agyei, J.N., 2008. Combining inorganic fertilizer with poultry manure for sustainable production of quality protein maize in Ghana. *Ghana*. J. Agric. Sci. 41 (1) https://doi.org/10.4314/gjas.v41i1.46144.

Dargie, S., Girma, T., Chibsa, T., Kassa, S., Boke, S., Abera, A., Haileselassie, B., Addisie, S., Amsalu, S., Haileselassie, M., Soboka, S., Abera, W., Weldesemayat, S.G., 2022. Balanced fertilization increases wheat yield response on different soils and agroecological zones in Ethiopia. Exp. Agric. 58 https://doi.org/10.1017/s0014479722000151.

Darko, D., Adjei, K.A., Odai, S.N., Obuobie, E., Asmah, R., Trolle, D., 2019. Recent climate trends for the Volta Basin in West Africa. Weather 14 (51). https://doi.org/10.1002/wea.3303.

Debtanu, M., Das, D.K., Pathak, H., 2006. Simulation of fertilizer requirement for irrigated wheat in eastern India using the QUEFTS model. Arch. Agron. Soil Sci. 52 (4), 403–418. https://doi.org/10.1100/tsw.2006.43.

Diao, X., Hazell, P., Kolavalli, S., Resnick, D., 2019. Ghana's economic and agricultural transformation: past performance and future prospects, p. 277. https://doi.org/10.1093/oso/9780198845348.001.0001.

Durodola, O.S., Mourad, K.A., 2020. Modelling Maize Yield and Water Requirements under Different Climate Change Scenarios. Climate 8 (11). https://doi.org/10.3390/cli8110127.

Essel, B., Abaidoo, R.C., Opoku, A., Ewusi-Mensah, N., 2020. Economically optimal rate for nutrient application to maize in the semi-deciduous Forest zone of Ghana. J. Soil Sci. Plant Nutr. 20 (4), 1703–1713. https://doi.org/10.1007/s42729-020-00240-y.

FAO/OECD, 2018. Food Security and Nutrition: Challenges for Agriculture and the Hidden Potential of Soil. FAO. http://www.fao.org/3/CA0917EN/ca0917en.pdf.

Farr, T.G., Kobrick, M., 2000. Shuttle radar topography Mission produces a wealth of data. Eos Trans. AGU 81, 583.

Fening, J.O., Yeboah, E., Gyapong, T.A., Gaizie, E., 2009. On farm evaluation of the contribution of three green manures to maize yield in the semi-deciduous forest zone of Ghana. Afr. J. Environ. Sci. Technol. 3 (9). https://www.ajol.info/index.php/ajest/article/view/46071.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37 (12), 4302–4315. https://doi.org/10.1002/joc.5086.

Florence, A., Ransom, M., Mengel, D., 2017. Potassium fixation by oxidized and reduced forms of phyllosilicates. Soil Sci. Soc. Am. J. 81 (5), 1247–1255. https://doi.org/10.2136/sssaj2016.12.0420.

Fosu-Mensah, B.Y., Manchadi, A., Vlek, P.L.G., 2019. Impacts of climate change and climate variability on maize yield under rainfed conditions in the sub-humid zone of Ghana: a scenario analysis using APSIM. *West African*. J. Appl. Ecol. 27 (1), 108 126-108 126. https://www.ajol.info/index.php/wajae/article/view/189216, 108 126-108 126.

Fox, J., Weisberg, S., 2019. An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Frimpong, K.A., Phares, C.A., Boateng, I., Abban-Baidoo, E., Apuri, L., 2021. One-time application of biochar influenced crop yield across three cropping cycles on tropical sandy loam soil in Ghana. Heliyon 7 (2), e06267. https://doi.org/10.1016/j.heliyon.2021.e06267.

Ghansah, B., Forkuo, E.K., Osei, E.F., Appoh, R.K., Asare, M.Y., Kluste, N.A.B., 2018. Mapping the spatial distribution of small reservoirs in the white Volta sub-basin of Ghana. Remote Sens. Appl. Soc. Environ. 9, 107–115. https://doi.org/10.1016/j.rsase.2017.12.003.

Giller, K.E., Franke, A.C., Abaidoo, R., Baijukya, F., Bala, A., Boahen, S., Dashiell, K., Kantengwa, S., Sanginga, J.-M., Sanginga, N., Simmons, A., Turner, A., 2013. N2Africa: Putting nitrogen fixation to work for smallholder farmers in Africa. In: Agro-Ecological Intensification of Agricultural Systems in the African Highlands. Routledge, pp. 176–194. https://doi.org/10.4324/9780203114742.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food security: the challenge of feeding 9 billion people. Sci. (80-.) 327 (5967), 812–818. https://doi.org/10.1126/science.1185383.

Graves, S., Piepho, H.-P., Selzer, L., 2019. Visualizations of Paired Comparisons: Package 'multcompView'. https://cran.r-project.org/web/packages/multcompView/index.html.

Guilpart, N., Grassini, P., van Wart, J., Yang, H., van Ittersum, M.K., van Bussel, L.G.J., Wolf, J., Claessens, L., Leenaars, J.G.B., Cassman, K.G., 2017. Rooting for food security in sub-Saharan Africa. Environ. Res. Lett. 12 (11), 1–8. https://doi.org/10.1088/1748-9326/aa9003.

GYGA, 2021. Agriculture in Ghana. University of Nebraska-Lincoln and Wageningen University & Research. Retrieved 24th March 2021 from. https://www.yieldgap.org/ghana.

Han, S., Kim, H., 2021. Optimal feature set size in random Forest regression. Appl. Sci. 11 (8) https://doi.org/10.3390/app11083428.

Hengl, T., Heuvelink, G.B., Kempen, B., Leenaars, J.G., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLoS One 10 (6), 1–26. https://doi.org/10.1371/journal.pone.0125814.

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. https://doi.org/10.1016/0016-7061(90)90021-Z.

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 (6), e0156571. https://doi.org/10.1371/journal.pone.0156571.

Jiang, P., Thelen, K.D., 2004. Effect of soil and topographic properties on crop yield in a North-Central corn–soybean cropping system, 96 (1), 252–258. https://doi.org/10.2134/agronj2004.0252.

Kanton, R.A.L., Prasad, P.V.V., Mohammed, A.M., Bidzakin, J.K., Ansoba, E.Y., Asungre, P.A., Lamini, S., Mahama, G., Kusi, F., Sugri, I., 2016. Organic and inorganic fertilizer effects on the growth and yield of maize in a dry agro-ecology in northern Ghana. J. Crop Improv. 30 (1), 1–16. https://doi.org/10.1080/15427528.2015.1085939.

Kihara, J., Nziguheba, G., Zingore, S., Coulibaly, A., Esilaba, A., Kabambe, V., Njoroge, S., Palm, C., Huising, J., 2016. Understanding variability in crop response to fertilizer and amendments in sub-Saharan Africa. Agric. Ecosyst. Environ. 229, 1–12. https://doi.org/10.1016/j.agee.2016.05.012.

Kpotor, P., Akromah, R., Ewool, M.B., Kena, A.W., Owusu-Adjei, E., Tuffour, H.O., 2014. Assessment of the relative yielding abilities and stability of maize (Zea mays L.) genotypes under different levels of nitrogen fertilization across two agro-ecological zones in Ghana. Int. J. Scientific Res. Agricult. Sci. 1 (7), 128–141. https://doi.org/10.12983/ijsras-2014-p0128-0141.

Kranjac-Berisavljevic, G., Bayorbor, T.B., Abdulai, A.S., Obeng, F., Blench, R.M., Turton, C.N., Boyd, C., Drake, E., 1999. Rethinking Natural Resource Degradation in Semi-Arid Sub-Saharan Africa: The Case of Semi-Arid Ghana. University for Development Studies, Tamale ODI, London.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97. https://doi.org/10.5194/adgeo-5-89-2005.

Kravchenko, A.N., Robertson, G.P., 2007. Can topographical and yield data substantially improve Total soil carbon mapping by regression kriging ? Agron. J. (99), 12–17. https://doi.org/10.2134/agronj2005.0251.
Kugbe, J.X., Kombat, R., Atakora, W., Tejada Moral, M., 2019. Secondary and micronutrient inclusion in fertilizer formulation impact on maize growth and yield across northern Ghana. Cogent Food Agricult. 5 (1) https://doi.org/10.1080/23311932.2019.1700030.
Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82 (13), 1–23. https://doi.org/10.18637/jss.v082.i13.
Kyei-Mensah, C., Kyerematen, R., Adu-Acheampong, S., 2019. Impact of rainfall variability on crop production within the Worobong ecological area of Fanteakwa District, Ghana. Adv. Agricult.e 2019, 1–7. https://doi.org/10.1155/2019/7930127.
Lal, R., 2006. Enhancing crop yields in the developing countries through restoration of the soil organic carbon pool in agricultural lands. Land Degrad. Dev. 17 (2), 197–209. https://doi.org/10.1002/ldr.696.
Lamos-Díaz, H., Puentes-Garzón, D.E., Zarate-Caicedo, D.A., 2020. Comparison between machine learning models for yield forecast in cocoa crops in Santander, Colombia. Revista Facultad de Ingeniería 29 (54), e10853. https://doi.org/10.19053/01211129.v29.n54.2020.10853.
Leenaars, J.G.B., Claessens, L., Heuvelink, G.B.M., Hengl, T., Ruiperez Gonzalez, M., van Bussel, L.G.J., Guilpart, N., Yang, H., Cassman, K.G., 2018. Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa. Geoderma 324, 18–36. https://doi.org/10.1016/j.geoderma.2018.02.046.
Logah, V., Ewusi-Mensah, N., Tetteh, F.K.M., 2011. Soil organic carbon and crop yield under different soil amendments and cropping Systems in the Semi-deciduous Forest Zone of Ghana. J. Plant Sci. 6 (4), 165–173. https://doi.org/10.3923/jps.2011.165.173.
Marusteri, M., Bacarea, V., 2010. Comparing groups for statistical differences: how to choose the right statistical test? Biochemia medica 20 (1), 15–32. https://hrcak.srce.hr/file/73801.
McHugh, M.L., 2011. Multiple comparison analysis testing in ANOVA. Biochemia medica 21 (3), 203–209. https://www.biochemia-medica.com/en/journal/21/3/10.11613/BM.2011.029.
Mueller, N.D., Gerber, J.S., Johnston, M., Ray, D.K., Ramankutty, N., Foley, J.A., 2012. Closing yield gaps through nutrient and water management. Nature 490 (7419), 254–257. https://doi.org/10.1038/nature11420.
Mugwe, J., Mugendi, D., Kungu, J., Muna, M.M., 2009. Maize yields response to application of organic and inorganic input under on-station and on-farm experiments in Central Kenya. Exp. Agric. 45 (1), 47–59. https://doi.org/10.1017/S0014479708007084.
Mulder, I., 2000. Soil Fertility: QUEFTS and Farmer's Perceptions (Vol. Working Paper No 30). IIED. https://www.researchgate.net/publication/228856865_Soil_fertility_QUEFTS_and_farmers%27_perceptions.
Nakagawa, S., Schielzeth, H., 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biol. Rev. 85 (4), 935–956. https://doi.org/10.1111/j.1469-185x.2010.00141.x.
Nakagawa, S., Johnson, P.C.D., Schielzeth, H., 2017. The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. J. R. Soc. Interface 14 (134), 20170213. https://doi.org/10.1098/rsif.2017.0213.
Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. J. Hydrol., 10(3), 282–290. https://en.wikipedia.org/wiki/Nash–Sutcliffe_model_efficiency_coefficient.
Nevavuori, P., Narra, N., Linna, P., Lipping, T., 2020. Crop yield prediction using multitemporal UAV data and Spatio-temporal deep learning models. Remote Sens. 12 (23), 4000. https://www.mdpi.com/2072-4292/12/23/4000.
Njoroge, K.S., 2019. Explaining variability in maize yield responses to nutrient applications in smallholder farms of western Kenya [(Publication Number 978-94-6395-164-7) Wageningen University & Research]. The Netherlands. https://doi.org/10.18174/503185.
Njoroge, R., Otinga, A.N., Okalebo, J.R., Pepela, M., Merckx, R., 2018. Maize (Zea mays L.) response to secondary and micronutrients for profitable N, P and K fertilizer use in poorly responsive soils. Agronomy 8 (4). https://doi.org/10.3390/agronomy8040049.
Nkrumah, F., Klutse, N.A.B., Adukpo, D.C., Owusu, K., Quagraine, K.A., 2014. Rainfall variability over Ghana: model versus rain gauge observation. Int. J. Geosci. 5 (7) https://doi.org/10.4236/ijg.2014.57060.
Obi, M.E., Ebo, P.O., 1995. The effects of organic and inorganic amendments on soil physical properties and maize production in a severely degraded sandy soil in southern Nigeria. Bioresour. Technol. 51 (2–3), 117–123. https://doi.org/10.1016/0960-8524(94)00103-8.
O'Halloran, I.P., Kachanoski, R.G., Stewart, J.W.B., 1985. Spatial variability of soil phosphorus as influenced by soil texture and management. Can. J. Soil Sci. 65 (3), 475–487. https://doi.org/10.4141/cjss85-051.
Onduru, D.D., Du Preez, C.C., 2007. Spatial and temporal aspects of agricultural sustainability in the semi-arid tropics: a case study in Mbeere district, Eastern Kenya. Trop. Sci. 47 (3), 134–148. https://doi.org/10.1002/ts.207.
Osei, B.A., 1995. Effects of different lime application rates and time on some chemical properties of an acid soil in Ghana. Soil Use Manag. 11, 25–29. https://doi.org/10.1111/j.1475-2743.1995.tb00491.x.
Ovalles, F.A., Collins, M.E., 1986. Soil-landscape relationships and soil variability in north Central Florida. Soil Sci. Soc. Am. J. 50 (2), 401–408. https://doi.org/10.2136/sssaj1986.03615995005000020029x.
Owoade, F.M., Adiku, S.G.K., Atkinson, C.J., MacCarthy, D.S., Kumahor, S.K., Kolawole, G.O., 2020. Location and land use effects on soil carbon accretion and productivity in the coastal savanna agro-ecological zone of Ghana. West Afr J. App. Ecol. 28 (2), 1–13. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwijgM_d2cD5AhVNLewKHfCfCYgQFnoECAIQAQ&url=https%3A%2F%2Fwww.ajol.info%2Findex.php%2Fwajae%2Farticle%2Fview%2F202673%2F191158&usg=AOvVaw1WdXQq6X2x3K-dU3roFeYD.
Owoade, F.M., Adiku, S.G.K., Atkinson, C.J., MacCarthy, D.S., 2021. Differential impact of land use types on soil productivity components in two agro-ecological zones of southern Ghana. African Handb. Climate Change Adapt. 1721–1733. https://doi.org/10.1007/978-3-030-45106-6_144.
Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., Athanasiadis, I.N., 2021. Machine learning for large-scale crop yield forecasting. Agric. Syst. 187 https://doi.org/10.1016/j.agsy.2020.103016.
Peterson, R.A., 2021. Finding optimal normalizing transformations via bestNormalize. R J. 13 (1), 310–329. https://doi.org/10.32614/RJ-2021-041.
Probst, P., Wright, M.N., Boulesteix, A.L., 2019. Hyperparameters and tuning strategies for random forest. WIREs Data Min. Knowl. Disc. 9 (3) https://doi.org/10.1002/widm.1301.
Quansah, G.W., 2010. Effect of Organic and Inorganic Fertilizers and their Combinations on the Growth and Yield of Maize in the Semi-Deciduous Forest Zone of Ghana [Msc, Kwame Nkrouma University of Science and Technology]. Kumsi, Ghana. http://csirspace.csirgh.com/handle/123456789/2379.
Rahman, A.N., Larbi, A., Kotu, B., Asante, M.O., Akakpo, D.B., Mellon-Bedi, S., Hoeschle-Zeledon, I., 2021. Maize–legume strip cropping effect on productivity, income, and income risk of farmers in northern Ghana. Agron. J. 1-12 https://doi.org/10.1002/agj2.20536.
Ren, T., Zou, J., Wang, Y., Li, X.K., Cong, R.H., Lu, J.W., 2015. Estimating nutrient requirements for winter oilseed rape based on QUEFTS analysis. J. Agric. Sci. 154 (3), 425–437. https://doi.org/10.1017/S0021859615000301.
RStudio Team, 2022. RStudio: Integrated Development for R. RStudio (2021.09.2). PBC, Boston, MA. *URL. http://www.rstudio.com/*.
Sadras, V.O., Calvino, P.A., 2001. Quantification of grain yield response to soil depth in soybean, maize, sunflower, and wheat. Agron. J. 93 (3), 577–583. https://doi.org/10.2134/agronj2001.933577x.
Sattaria, S.Z., van Ittersuma, M.K., Bouwmanb, A.F., Smitd, A.L., Janssena, B.H., 2014. Crop yield response to soil fertility and N, P, K inputs in different environments: testing and improving the QUEFTS model. Field Crop Res. 157, 35–46. https://doi.org/10.1016/j.fcr.2013.12.005.
Sawyer, J.E., Mallarino, A.P., 1999. Differentiating and understanding the Mehlich 3, bray, and Olsen soil phosphorus tests. In: Agronomy Conference Proceedings and Presentations, 12. https://dr.lib.iastate.edu/handle/20.500.12876/4349.
Searle, S.R., Casella, G., McCulloch, C.E., 1992. Variance Components. John Wiley & Sons. https://doi.org/10.1002/9780470316856.
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 https://doi.org/10.1016/j.fcr.2019.107585.
Sileshi, G.W., Kihara, J., Tamene, L., Vanlauwe, B., Phiri, E., Jama, B., 2022. Unravelling causes of poor crop response to applied N and P fertilizers on African soils. Exp. Agric. 1-17 https://doi.org/10.1017/S0014479721000247.
Solomon, D., Lehmann, J., Fraser, J.A., Leach, M., Amanor, K., Frausin, V., Kristiansen, S. M., Millimouno, D., Fairhead, J., 2016. Indigenous African soil enrichment as a climate-smart sustainable agriculture alternative. Front. Ecol. Environ. 14 (2), 71–76. https://doi.org/10.1002/fee.1226.
Spilke, J., Piepho, H.P., Hu, X., 2005. Analysis of unbalanced data by mixed linear models using the MIXED procedure of the SAS system. J. Agron. Crop Sci. 191 (1), 47–54. https://doi.org/10.1111/j.1439-037X.2004.00120.x.
SRID/MoFA, 2021. Agriculture in Ghana: Facts and Figures (2020). https://srid.mofa.gov.gh/sites/default/files/Agriculture%20In%20Ghana%20Facts%20%26%20Figures_%202020%20FINAL.pdf.
Tabi, F.O., Diels, J., Ogunkunle, A.O., Iwuafor, E.N.O., Vanlauwe, B., Sanginga, N., 2007. Potential nutrient supply, nutrient utilization efficiencies, fertilizer recovery rates and maize yield in northern Nigeria. Nutr. Cycl. Agroecosyst. 80 (2), 161–172. https://doi.org/10.1007/s10705-007-9129-z.
Tetteh, F., Larbi, A., Nketia, K.A., Senayah, J.K., Hoeschle-Zeledon, I., Abdulrahman, N., 2016. Suitability of soils for cereal cropping in northern Ghana. Ibadan. IITA, Nigeria, pp. 1–18. https://cgspace.cgiar.org/handle/10568/96198.
Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proc. Natl. Acad. Sci. U. S. A. 108 (50), 20260–20264. https://doi.org/10.1073/pnas.1116437108.
Timsina, J., Dutta, S., Devkota, K.P., Chakraborty, S., Neupane, R.K., Bishta, S., Amgain, L.P., Singh, V.K., Islam, S., Majumdar, K., 2021. Improved nutrient management in cereals using nutrient expert and machine learning tools: productivity, profitability and nutrient use efficiency. Agric. Syst. 192 https://doi.org/10.1016/j.agsy.2021.103181.
Tittonell, P., Corbeels, M., van Wijk, M.T., Vanlauwe, B., Giller, K.E., 2008. Combining organic and mineral fertilizers for integrated soil fertility Management in Smallholder Farming Systems of Kenya: explorations using the crop-soil model FIELD. Agron. J. 100 (5), 1511–1526. https://doi.org/10.2134/agronj2007.0355.
USAID, 2022. Agriculture and Food Security. USAID. Retrieved 18/02/2022 from. https://www.usaid.gov/ghana/agriculture-and-food-security.
Uzoma, K.C., Inoue, M., Andry, H., Fujimaki, H., Zahoor, A., Nishihara, E., 2011. Effect of cow manure biochar on maize productivity under sandy soil condition. Soil Use Manag. 27 (2), 205–212. https://doi.org/10.1111/j.1475-2743.2011.00340.x.

van Loon, M.P., Adjei-Nsiah, S., Descheemaeker, K., Akotsen-Mensah, C., van Dijk, M., Morley, T., van Ittersum, M.K., Reidsma, P., 2019. Can yield variability be explained? Integrated assessment of maize yield gaps across smallholders in Ghana. Field Crop Res. 236, 132–144. https://doi.org/10.1016/j.fcr.2019.03.022.

Wallach, D., Makowski, D., Jones, J.W., Brun, F., 2018. Working with Dynamic Crop Models. Academic Press-Elsevier. https://doi.org/10.1016/C2016-0-01552-8.

Wang, L., 2021. Data driven explanation of temporal and spatial variability of maize yield in the United States. Front. Plant Sci. 12, 701192 https://doi.org/10.3389/fpls.2021.701192.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. Retrieved from. https://ggplot2.tidyverse.org.

Wijayanto, Y., Prastyanto, E., 2012. A study of using QUEFTS model for establishing site specific fertilizer recommendation in maize on the basis of farmer fields. AGRIVITA. J. Agric. Sci. 33 (3), 273–278. https://media.neliti.com/media/publications/53171-EN-a-study-of-using-quefts-model-for-establ.pdf.

Wright, M.N., Ziegler, A., 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77 (1), 1–17. https://doi.org/10.18637/jss.v077.i01.

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. https://doi.org/10.1016/j.fcr.2013.06.006.

Yawson, D.O., Kwakye, P.K., Armah, F.A., Frimpong, K.A., 2011. The dynamics of potassium (K) in representative soil series of Ghana. ARPN J. Agricult. Biol. Sci. 6 (1), 48–55. https://www.researchgate.net/publication/260793689_The_Dynamics_of_Potassium_K_in_Representative_Soil_Series_of_Ghana.

Yeboah, E., Asamoah, G., Kofi, B., Abunyewa, A.A., 2016. Effect of biochar type and rate of application on maize yield indices and water use efficiency on an Ultisol in Ghana. Energy Procedia 93, 14–18. https://doi.org/10.1016/j.egypro.2016.07.143.

Zingore, S., Delve, R.J., Nyamangara, J., Giller, K.E., 2007. Multiple benefits of manure: the key to maintenance of soil fertility and restoration of depleted sandy soils on African smallholder farms. Nutr. Cycl. Agroecosyst. 80 (3), 267–282. https://doi.org/10.1007/s10705-007-9142-2.

Zingore, S., Mutegi, J., Agesa, B., Tamene, L., Kihara, J., 2015. Soil degradation in sub-Saharan Africa and crop production options for soil rehabilitation. Better Crop. 99 (1), 24–26. https://cgspace.cgiar.org/handle/10568/68702.