Agricultural Systems

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

--Manuscript Draft--

Manuscript Number:	AGSY-D-22-00430R3
Article Type:	Research Paper
Keywords:	Maize yield variability, Fertilizer rate, QUEFTS, Linear Mixed Effects Model, Random Forest, Maize yield prediction
Corresponding Author:	Prem Bindraban UNITED STATES
First Author:	Anselme K. K. Kouame
Order of Authors:	Anselme K. K. Kouame
	Prem S. Bindraban
	Isaac N. Kissiedu
	Williams K. Atakora, PhD
	Khalil El Mejahed, PhD
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 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 R2 = R2m = 0.48) and environmental variables (R^2m = 0.43) contributed more to explaining yield variability, whereas in GS and TZ, fertilizer rates (R^2m = 0.35 in GS and 0.26 in TZ) and variety (R^2m = 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.
Suggested Reviewers:	Shamie Zingore, PhD Director for Research African Plant Nutrition Institute, African Plant Nutrition Institute s.zingore@apni.net Highly experienced agronomy scientists on African agriculture.
	Rebbie Harawa, PhD ICRISAT: International Crops Research Institute for the Semi-Arid Tropics

	R.Harawa@cgiar.org Knowledgeable scientist about fertilizer use in Africa.
	Bas Kempen, PhD Researcher, ISRIC World Soil Information bas.kempen@wur.nl One of the authorities in Machine learning methodologies.
	Marc Corbeels, PhD Researcher, IITA: International Institute of Tropical Agriculture m.corbeels@cgiar.org Knowledgeable researcher on agricultural production systems Africa.
Response to Reviewers:	All our responses are in the document "Responses to Reviewers"

Editors Agricultural Systems

Topic: Submission of research paper on factors that determine maize yield variability in Ghana.

5 April 2023

Dear Editors,

The use of fertilizers to increase crop yields in sub-Saharan Africa is generally accepted. Yet blanket fertilizer recommendations seem ineffective and are increasingly tailored to location-specific conditions. These recommendations are generally based on the chemical properties of the soil, but their effectiveness still seems quite low. Therefore, we analyzed the factors that determine the variability of maize yield in a context where the rate of fertilizer application per hectare is increasing. This information should encourage us to consider the most relevant factors in the methodologies to prioritize the actions to be taken.

Anselme K. K. Kouame, Isaac N. Kissiedu, Williams K. Atakora, Khalil El Mejahed, Prem S. Bindraban. IDENTIFYING DRIVERS FOR VARIABILITY IN MAIZE (*Zea mays L.*) YIELD IN GHANA: A META-REGRESSION APPROACH

With kind regards,

Prem Bindraban
International Fertilizer Development Center (IFDC), pbindraban@ifdc.org; Tel: +31624168617

Also, on behalf of Anselme Kouame, Isaac Kissiedu, Williams Atakora, Khalil El Mejahed

Highlights

Editors Agricultural Systems

Topic: Submission of research paper on factors that determine maize yield variability in Ghana.

26 February 2023

Dear Editors,

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 modeling showed that root zone depth was also a key factor in explaining maize yield variability in Ghana.

Anselme K. K. Kouame, Prem S. Bindraban, Isaac N. Kissiedu, Williams K. Atakora, Khalil El Mejahed, IDENTIFYING DRIVERS FOR VARIABILITY IN MAIZE (*Zea mays L.*) YIELD IN GHANA: A META-REGRESSION APPROACH

With kind regards,

Prem Bindraban
International Fertilizer Development Center (IFDC), pbindraban@ifdc.org; Tel: +31624168617

Also, on behalf of Anselme Kouame, Isaac Kissiedu, Williams Atakora, Khalil El Mejahed

Editors Agricultural Systems

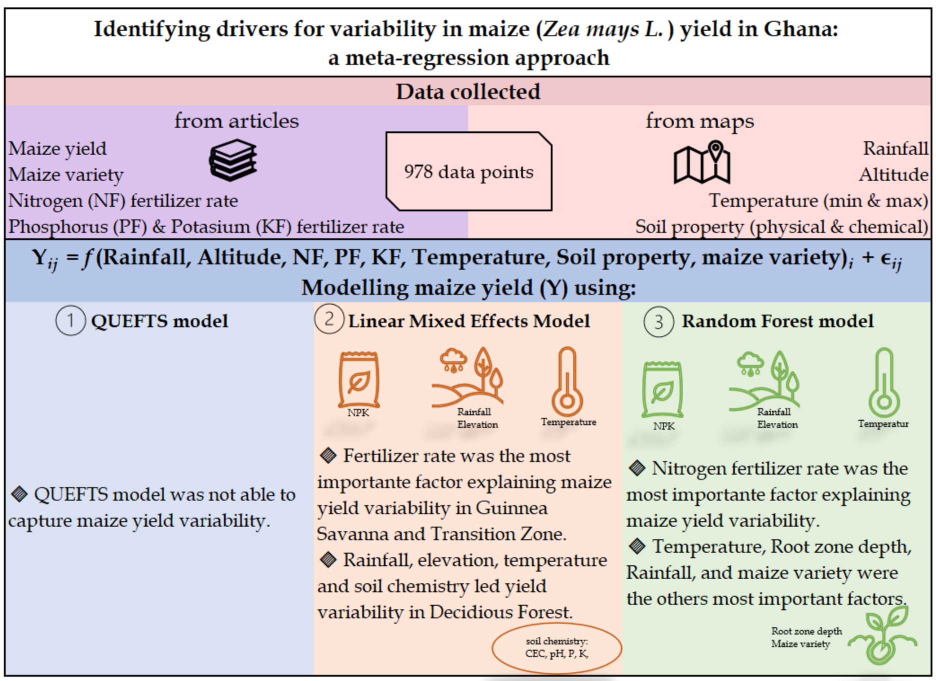
Topic: Submission of research paper on Identifying drives for variability in maize yield in Ghana.

18 August 2022

Dear Editors,

GRAPHICAL ABSTRACT

Anselme K. K. Kouame, Prem S. Bindraban, Isaac N. Kissiedu, Williams K. Atakora, Khalil El Mejahed. Identifying drivers for variability in maize (Zea mays L.) yield in Ghana: a meta-regression approach



With kind regards,

Prem Bindraban
International Fertilizer Development Center (IFDC), pbindraban@ifdc.org; Tel: +31624168617

Also, on behalf of Anselme Kouame, Isaac Kissiedu, Williams Atakora, Khalil El Mejahed

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 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 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.

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

Reviewer comments:

Reviewer #1:

The revised manuscript has an updated analysis and presentation of data, and comments of the reviewer have been addressed. Unfortunately, the revised manuscript contains an exorbitant number of editorial and factual errors in the presentation and description of findings. For example, most supplementary tables and some supplementary figures are not corrected numbered in the text. Many supplementary tables and figures cited in the manuscript do not appear in the supplementary material. In a number of instances, the variables mentioned in the text as significantly related to yield do not match the variables listed in tables as significantly related to yield. In a number of instances, cited relationships in the text do not match with relationships shown in supplementary figures or correlation matrices.

In addition, the Discussion section is extremely long. I believe a substantially shortened version of the Discussion section that focuses on key findings and messages would improve the manuscript.

Comment-1: I recommend reducing the Discussion section to about 60% of its current length and focusing the Discussion on the most important findings and messages.

- **Authors response-1**: The authors thank the reviewer for this feedback. Indeed, the authors have reduced the length of the discussion, which was 3407 words to 2226 words, a reduction of 35% in length of the old version of the discussion. The authors think that too much reduction of the current length of the discussion (new version) could alter the message conveyed by the article, and also fail to explain certain results which deserve to be explained, in view of the many meta-analysis done.

I urge the authors to please be more mindful in ensuring accuracy in their presentation and reporting of data and finalization of the manuscript. The following are some discrepancies and errors that need to be addressed.

Comment-2. Lines 236-237: “Table 1” should be “Supplementary Table S3”.

- **Authors response-2**: The authors thank the reviewer for his comments. The authors have changed “Table 1” to “Supplementary Table S3” (Line 238).

Comment-3. Line 241 and throughout the manuscript: "Supplementary Table S4" should be "Supplementary Table S3". From this point onward in the manuscript all the numbers of supplemental tables cited the text are not the correct numbers of the supplementary table. The manuscript must be revised to ensure table numbers cited in the text accurately match with the relevant table.

- **Authors response-3**: The authors agree with the reviewer and thank him for his comment. An error in the automatic table number update incremented the table numbers throughout the manuscript, causing these supplementary table numbering errors. Therefore, the authors have gone through the entire manuscript and corrected any citation errors in supplementary table numbers. (Line 242)

Comment-4: Lines 313, 317, 318, 323, and throughout the manuscript: "Supplementary Table S17" and beyond that are listed in the text and not included in the supplement. The cited supplementary tables must be included in the supplement.

- **Authors response-4**: The authors again agree with the reviewer. As mentioned in the previous author's response 3, the automatic table and figure number update incremented the supplementary table numbers throughout the manuscript, thus generating table numbers that did not exist in the supplementary table and figure file. Therefore, the authors have gone through the entire manuscript and corrected all citation errors of supplementary table and figure numbers.

Comment-5: Line 370: "Supplementary Figure S1A" should be "Supplementary Figure S7". The manuscript must be revised to ensure figure numbers cited in the text accurately match with the relevant figure.

- **Authors response-5**: The authors agree with the reviewer and thank him for his comment. Then the authors went through the entire manuscript and corrected citation errors in additional figure numbers. The authors replaced "Supplementary Figure S1A" cited at line 370 to "Supplementary Figure S 7 A". (Line 381)

Comment-6: Lines 496 and 498: 'Supplementary Figure S8' and 'Supplementary Figure S9' are not included in the supplement.

- **Authors response-6**: The authors again agree with the reviewer. The automatic figure number update incremented the figure numbers throughout the manuscript, thus generating figure numbers that did not exist in the supplemental figure file. Therefore, the authors went through the entire manuscript and corrected all citation errors of the supplementary figure numbers and adapted them to the result mentioned in the supplementary figure file.

<u>Comment-7:</u> Lines 535, 539, 545, 550, and 558: The cited supplementary figures are not included in the supplement. Cited supplementary figures must be included in the supplement.

- **Authors response-7**: The authors agree with the reviewer and thank him warmly for this comment. The supplementary figures cited in the manuscript, which do not appear in the supplementary figures file, have all been corrected throughout the manuscript.

<u>Comment-8:</u> Lines 261-262: The table or source for the cited 47%, 24%, and 33% is unclear and should be more clearly indicated.

- **Authors response-8**: The authors agree with the reviewer and thank him very much for this comment. The table number referring to quote percentages has been corrected and quoted correctly. (Line 265)

<u>Comment-9:</u> Lines 293-294: The source of the conclusion that variation in yield was greater for SOC should be cited because the relationship for yield and SOC is not reported in Supplementary Figure S3.

- **Authors response-9**: The authors thank the reviewer for this comment. The table referring to this claim is Table 2. It was therefore cited by the authors as recommended by the reviewer. (Line 301-302)

<u>Comment-10:</u> There are a number of instances where a variable mentioned in the text does not match with the variables listed in the cited table. The following are some examples of where revisions are required to ensure that variables cited in the text as related to yield match with the variables listed in the cited table.

<u>Comment-10-1</u> Lines 297-298: 'RootDEP' is listed in the text as significantly related to yield, but it is not included in Table 2 that is cited at the end of the sentence.

- **Authors response-10-1**: The authors thank the reviewer again for his comment. The table referring to "RootDEP" was the Supplementary Table S 9 . Therefore, the authors cited the table to which they referred at line 309.

<u>Comment-10-2:</u> Line 322: 'CLAY' and 'SOC' are listed in the text as significantly related to yield, but it is not included in Table 3 that is cited at the end of the sentence.

- **Authors response-10-2**: The authors thank the reviewer for his comment. The table referring to "CLAY" and "SOC" as being significantly related to yield was the Supplementary Table S 11. Therefore, the authors cited the Supplementary Table S 11 to which they referred at Line 333-334.

<u>Comment-10.3</u>: Line 332: 'MMet' and 'SOC' are listed in the text as significantly related to yield, but it is not included in Table 4 that is cited at the end of the sentence.

- **Authors response-10.3**: The authors thank the reviewer for his comment. The table referring to 'MMet' and 'SOC' as being significantly related to yield was the Supplementary Table S 12. Therefore, the authors cited the Supplementary Table S 12 to which they referred at Line 342-344.

<u>Comment-11:</u> Lines 312-313: The sentence indicates that KF is significantly associated with yield, but KF is not listed as a significant variable in Table 2 and the Discussion (lines 456-457) concludes that KF is not important. The manuscript should be revised to ensure the conclusions throughout the manuscript on the importance of KF are consistent with the results of data analyses.

- **Authors response-11**: The authors thank the reviewer for his comment. Indeed, the KF was not significantly related to maize yield in Table 2 but was in Supplementary S9. That justify why the authors said that KF was also significantly associated with maize yield at line 322-323.

<u>Comment-12</u>: Lines 345-346: This sentence stating that PF was not significantly related to yield is not consistent with cited Table 5 that shows PF was related to yield.

- **Authors response-12**: The authors thank the reviewer for this insightful comment. The authors modified line 354-356 and included PF as a factor significantly related to maize yield as shown in Table 5.

<u>Comment-13:</u> Line 378: Should 'NSE' be 'RMSE' as reported in Fig. 4?

- **Authors response-13**: The authors thank the reviewer for this insightful comment. No, the NSE must not be RMSE as shown in figure 4. However, this is not an error, the two do not have the same meaning and the same interpretation even if they both have the same figures.

Comment-14: Lines 433-434: This statement of a positive trend for yield and SOC is not supported by the correlation matrix or a graph in Supplementary Figure 3. In fact, this statement contradicts the correlation matrix in Supplementary Figure 3 that shows no significant correlation for yield and SOC.

- **Authors response-14**: The authors thank the reviewer for this insightful comment. Therefore, the reviewer is correct on this point. This assertion of a positive trend for yield and SOC is not supported by the correlation matrix shows by the Supplementary Figure 3. The author has therefore removed the SOC at this position and reworded the sentence according to what is. (Line 449-450)

Comment-15: Lines 440-441. This sentence is not accurate. Supplementary Figure 4 does not report a relationship for yield and pH and the reported relationship for CEC and yield is not significant.

- **Authors response-15**: The authors agree with the reviewer and thank him very much for this comment. The figure number referred to by the authors is Supplementary Figure S 3 c, d. The authors have therefore corrected the figure number cited in line.

Comment-16: Lines 448-450: This sentence does not accurately report the cited data in Supplementary Figure 6. The graph in Supplementary Figure 6 indicates the relationship with CEC was significant when all data were used.

Comment-17: Lines 453-454: This sentence does not accurately report data in Supplementary Figure 6 or 7. The data for CEC are not reported in Supplementary Figure 7 as indicated. And the relationship reported in Supplementary Figure 6 (when some CEC data are excluded) is neither positive or significant.

- **Authors response-16-17**: The authors agreed with the reviewer for these 2 comments. Indeed, the authors were referring instead to “Supplementary Graph S 3 a” and “Supplementary Graph S 3 e”, which showed a relationship between CEC and performance when all data points were included in the model. The authors therefore reworded the paragraph to cite the graph to which they were referring on lines 470-482.

Comment-18: Lines 475-476: This statement is not accurate because no significant relationship of yield and SOC is shown in Supplementary Figure 3, and it is unclear from Table 2 whether the relationship is positive or negative.

- **Authors response-18**: The authors agree with the reviewer and thank him very much for his comment. In fact, the authors revised this paragraph and cited the graphs to which it referred on lines 496-511. Indeed, the correlation matrix of "Supplementary 3 a" and Table 2 do not show this significant relationship between SOC and maize yield when all data points are considered. However, the "Supplementary Figure S 3 f "does show this positive and significant ($p<0.05$) relationship between SOC and maize yield on all data. Therefore "Supplementary Figure S 3 f " have been cited by authors (lines 496-511).

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.
[*] Corresponding author: pbindraban@ifdc.org (P.S. Bindraban).

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.

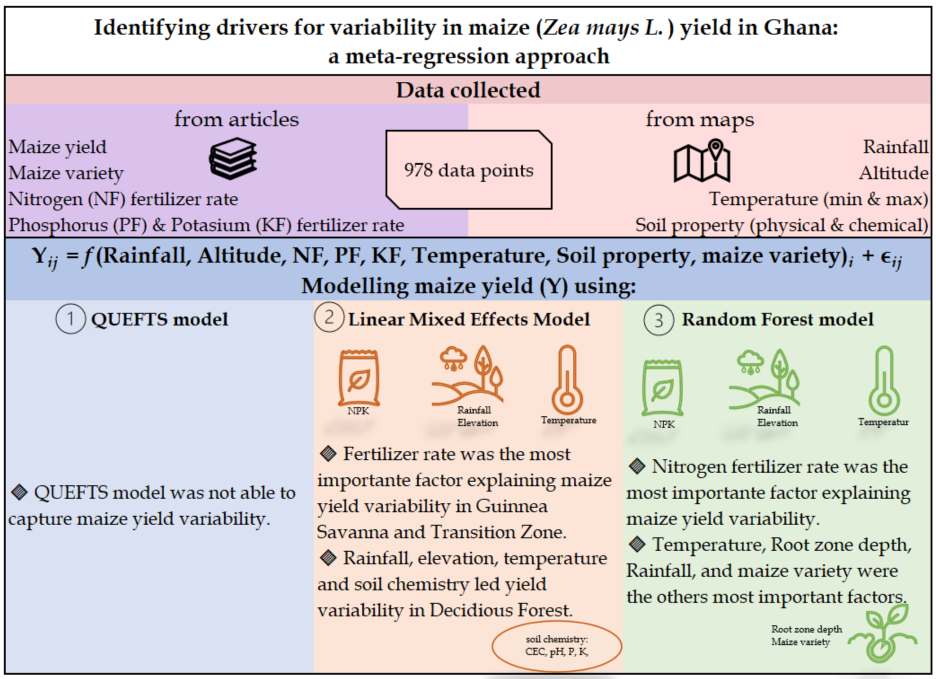
METHODS: The QUantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) model, the Linear Mixed Effects Model (LMM), and the Random Forest (RF) model were used to evaluate multiple effect sizes.

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 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.

Graphical abstract



Key words:

- Maize yield variability,
- Fertilizer rate,
- QUEFTS,
- Linear Mixed Effects Model,
- Random Forest,
- Maize yield prediction

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 modeling showed that root zone depth was also a key factor in explaining maize yield variability in Ghana.

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 hectares 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., 2021).

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 et al., 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 et al., 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. (2021) 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 more than 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 et al., 2012; Xu et al., 2013; Ren et al., 2015), and specifically recommended for use in Ghana as well (Wijayanto et al., 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 equation (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:

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

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) (Figure 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 (Figure 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

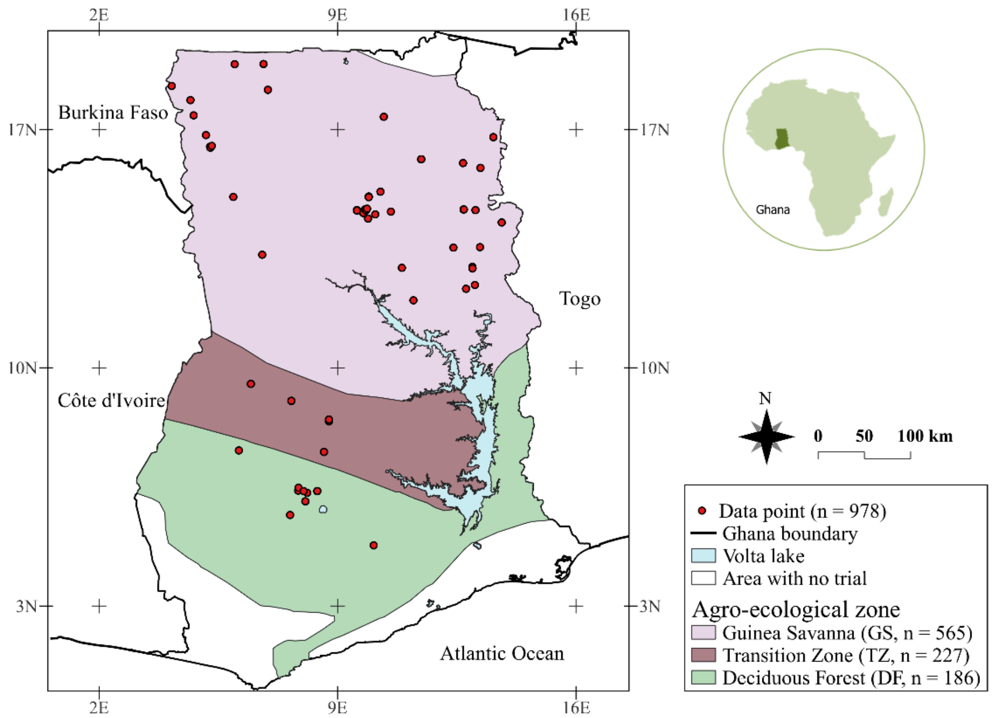


Figure 1: Geographic distribution of the experimental locations on a map of Ghana that shows the country's agro-ecological zones. Many experimental locations overlap.
Agroecological map source: (Antwi et al., 2014)

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 et al., 2017) database, at a spatial resolution of 1 km^2 and land elevation data with a resolution of 30 meters 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 et al., 2000)
Monthly precipitation	mm	MP [1]	(Fick et al., 2017)
Monthly mean temperature	°C	MMeT [2]	
Monthly min temperature	°C	MMiT [2]	
Monthly max temperature	°C	MMaT [2]	
Soil organic carbon	g kg^{-1}	SOC	(Hengl et al., 2015)
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	
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 et al. (2012) and Antwi et al. (2017) were used to replace the 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 et al. (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 Figure 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 et al., 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 less than 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 less than 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 equation 2, where Y_{ijk} is the *k*th observed value of yield formed by the *i*th level of treatment effect, the *j*th level of AEZ effect, the *ij*th level of interaction between treatment and AEZ, and the error (ϵ_{ijk}).

$$Y_{ijk} = Treatment_i + AEZ_j + (Treatment*AEZ)_{ij} + \epsilon_{ijk} \quad (2)$$

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 et al., 2010; McHugh, 2011).

One-way ANOVA equations 3, 4 and 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 et al., 2010). The one-way ANOVA equations with random effects were constructed as follows:

$$Y_{ij\_Treat} = Treatment_i + \epsilon_{ij} \tag{3}$$
$$Y_{ij\_AEZ} = AEZ_i + \epsilon_{ij} \tag{4}$$
$$Y_{ij\_Year} = Year_i + \epsilon_{ij} \tag{5}$$

where Y_{ij} represents the yields, Treatment_i is the effect of the *i*th treatment, AEZ_i is the effect of the *i*th AEZ, $Year_i$ is the effect of the *i*th 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 *i*th 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 *j*th AEZ) and $Year_k$ (random effect of the *i*th year) in LMM 6_{ENV}, 7_{SPP}, 8_{SCP}, 9_{VAR}, 10_{FER}, 11_{step}, and only by $Year_j$ (random effect of the *j*th 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:

$$(LMM\ 6_{ENV}) \quad Y_{ijk\_ENV} = f(ENV)_I + AEZ_j + Year_k + \epsilon_{ijk} \quad (6)$$
$$(LMM\ 7_{SPP}) \quad Y_{ijk\_SPP} = f(SPP)_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (7)$$
$$(LMM\ 8_{SCP}) \quad Y_{ijk\_SCP} = f(SCP)_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (8)$$
$$(LMM\ 9_{VAR}) \quad Y_{ijk\_VAR} = f(VAR)_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (9)$$
$$(LMM\ 10_{FER}) \quad Y_{ijk\_FER} = f(FER)_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (10)$$
$$(LMM\ 11_{step}) \quad Y_{ijk\_step} = f(ENV,\ SPP,\ SCP,\ VAR,\ FER)_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (11)$$
$$(LMM\ 12_{ENV\_AEZ}) \quad Y_{ij\_ENV\_AEZ} = f(ENV)_i + Year_j + \epsilon_{ij} \quad (12)$$
$$(LMM\ 13_{SPP\_AEZ}) \quad Y_{ij\_SPP\_AEZ} = f(SPP)_i + Year_j + \epsilon_{ij} \quad (13)$$
$$(LMM\ 14_{SCP\_AEZ}) \quad Y_{ij\_SCP\_AEZ} = f(SCP)_i + Year_j + \epsilon_{ij} \quad (14)$$
$$(LMM\ 15_{VAR\_AEZ}) \quad Y_{ij\_VAR\_AEZ} = f(VAR)_i + Year_j + \epsilon_{ij} \quad (15)$$
$$(LMM\ 16_{FER\_AEZ}) \quad Y_{ij\_FER\_AEZ} = f(FER)_i + Year_j + \epsilon_{ij} \quad (16)$$
$$(LMM\ 17_{step\_AEZ}) \quad Y_{ij\_AEZ\_step} = f(ENV,\ SPP,\ SCP,\ VAR,\ FER)_i + Year_j + \epsilon_{ij} \quad (17)$$

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 equation 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 et al., 2021).

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

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 et al., 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 et al., 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 , Figure 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, Figure 2 B).

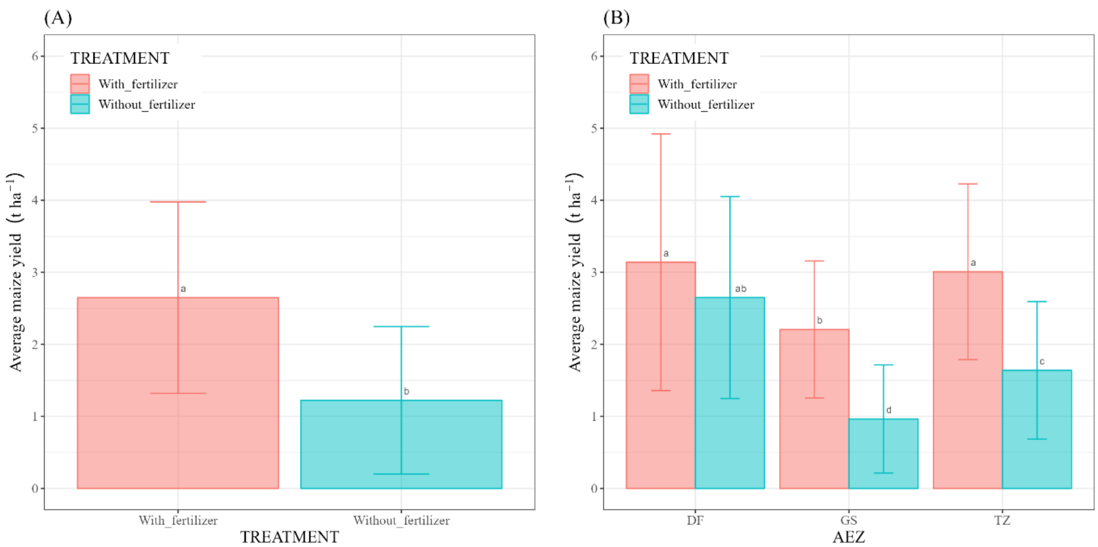


Figure 2: Average maize yield (t ha^{-1}) of treatments with fertilizer and without fertilizer (A) and the interaction between treatments and agro-ecological zone (AEZ) (B). DF (Deciduous Forest), GS (Guinea Savanna), and TZ (Transition Zone). Error bars are standard deviation bars. Different letters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

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 (Figure 3 A). Figure 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%], Figure 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 ($r = 0.43$, $p < 0.05$)

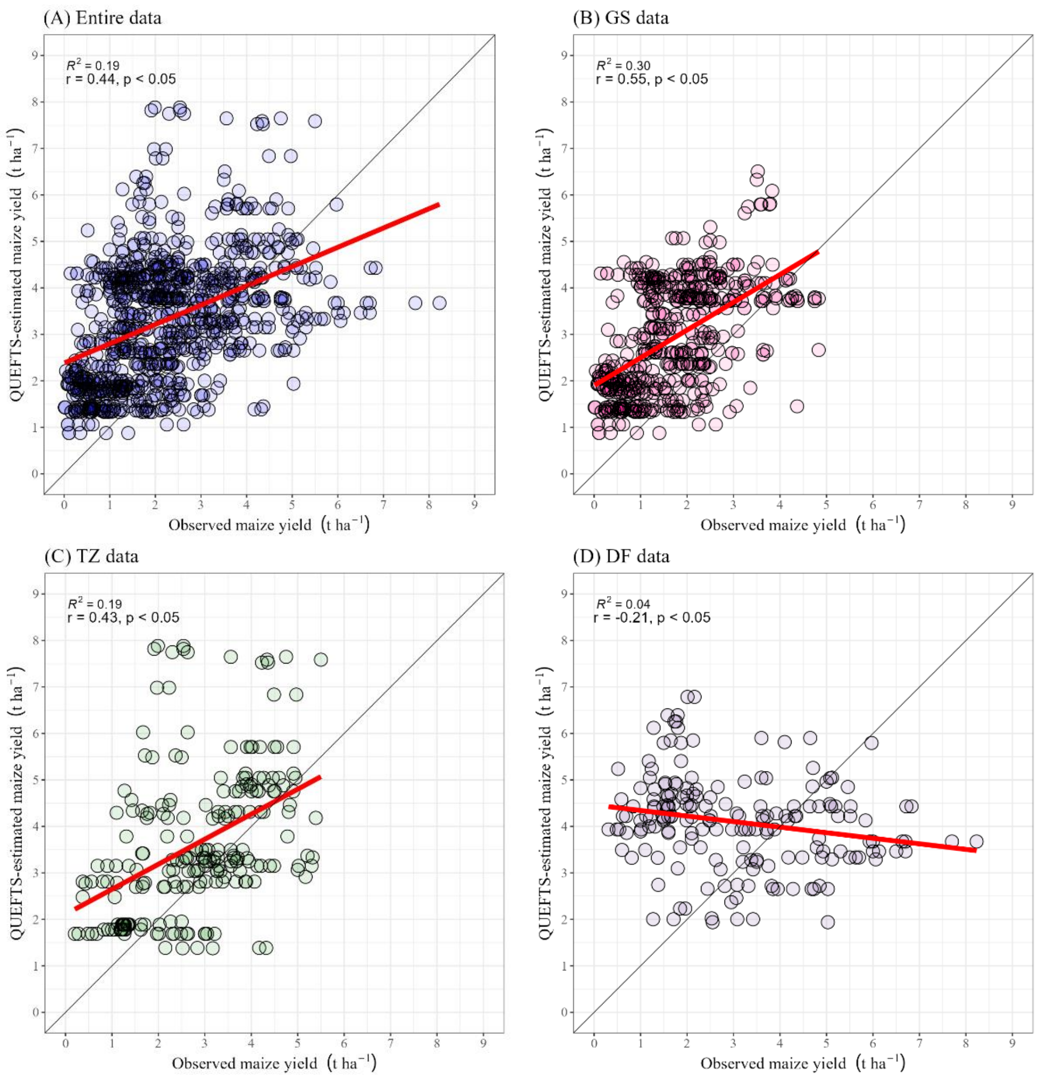


Figure 3: Relationship between observed and QUEFTS-estimated maize yield (t ha^{-1}) across all data points (A) and per AEZ (B, C, D). The R^2 indicates the portion of yield variability explained by soil chemical properties and fertilizers (confounded effect). The bold red line represents the linear 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.

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 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 Figure 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 Figure 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 overestimated 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 Figure 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 Figure S 3 a, c). Compared to P, K, pH, and CEC, the variation in yield due to SOC was greater (F value = 0.35) (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 Figure S 3 a).

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
LMM 11_{step}(†)	CLAY	5.4	1	959.9	12.3	< 0.05	0.26	0.53	0.17
	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			

PF	27.4	1	963.5	61.9	< 0.05

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

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 Figure S 3 a). A plot of maize yield on MP showed a significant ($p < 0.05$) and clear negative trend (Supplementary Figure 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 Figure S 4 a, b, e). The PF rate accounted for 48% of the R^2_m (0.39); 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 ($R^2_m = 0.4$), 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 ($R^2_m = 0.19$).

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
	ELV	4.6	1	462.3	9.1	< 0.05		
	SAND	3.7	1	350.4	7.9	< 0.05		
	CEC	6.7	1	556.0	13.2	< 0.05		
LMM 17_{step_GS}(†)	VARIETY	15.3	2	549.4	14.8	< 0.05	0.39	0.02
	NF	8.7	1	554.5	16.8	< 0.05		
	PF	28.6	1	551.1	55.5	< 0.05		

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

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, and SCP ($R^2_m = 0.24$).

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.

LMM analysis showed that the contribution of maize variety type in the variations in yield in the 3 AEZs was also significant ($p < 0.05$) (Table 3, 4, 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 Figure 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 (Figure 4A, Supplementary Figure 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.

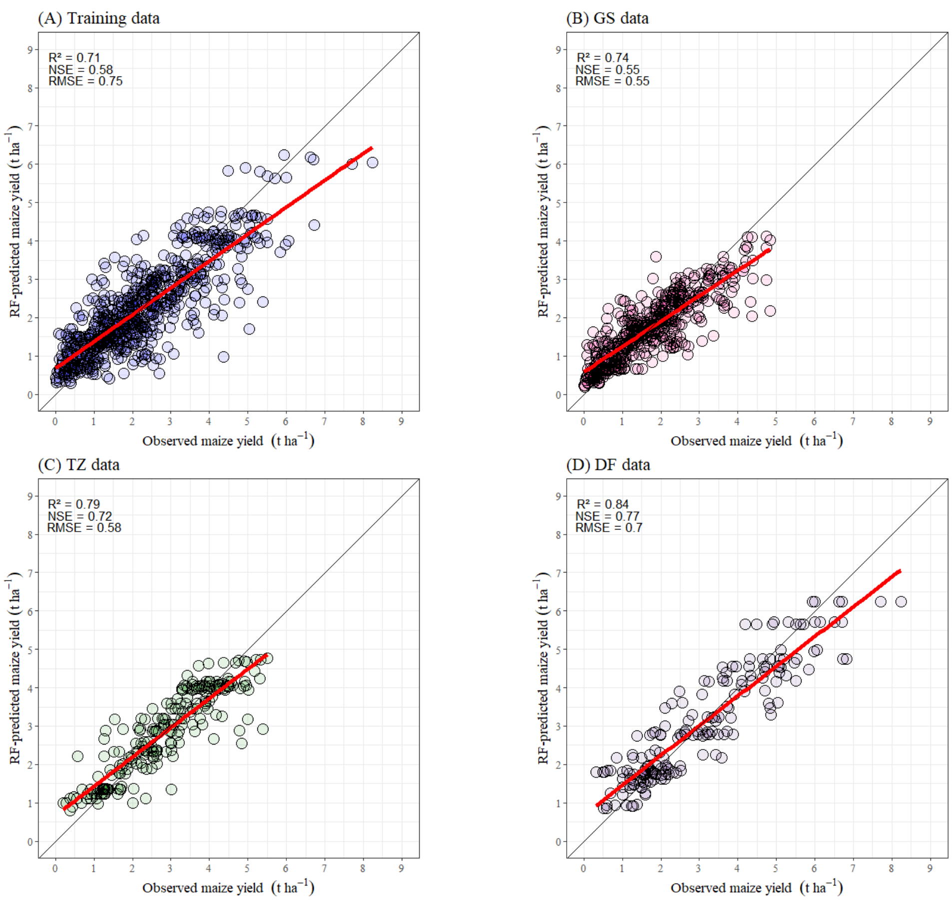


Figure 4: Relationship between observed maize yield and that predicted by the Random Forest (RF) model using 3 repeats of a 10-fold cross-validation approach. (A) The linear regression on 75% of the entire data point (training data point), (B) only the data point from GS, (C) only the data point from 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).

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 et al. (2007). Previous studies in Kenya, Benin, and Rwanda have also shown low precision and accuracy of QUEFTS maize yield predictions (Mulder, 2000; Onduru et al., 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 (Figure S 3 a), also identified by Braimoh et al. (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 Figure 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 Figure S 3 d). Excluding these data points turned the relation significantly ($p < 0.05$) positive (Supplementary Figure 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 Figure S 6 a). According to Yawson et al. (2011), Ghana's forest soils will 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 Figure 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 Figure 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 Table 3 and 5, in GS and DF respectively, PF had the highest (F-value = 60.31) and lowest (F-value = 4.99) 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 Figure S 7 A). A more specific analysis showed that this influence originated mostly from the DF data points (Supplementary Table S 13, Figure 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 Figure 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 et al. (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 Figure S 4 d). While the correlation coefficient was significant ($p < 0.05$), the distribution of data points in the Supplementary Figure 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, Figure S 3 a, b, S 5 b, S 6 a) and RF yield modelling (Supplementary Figure S 7 A) showed that environmental variables (MMeT, MP, and ELV) were related to maize yield variability, as also reported by Onduru et al. (2007), Mugwe et al. (2009), Fosu-Mensah et al. (2019), Kyei-Mensah et al. (2019), and Cudjoe et al. (2021). However, MP of the entire data point reveals a negative trend with yield (Supplementary Figure 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 et al. (2020) and Bagula et al. (2022).

The negative trend between MP and RootDEP support this logic and why high MP in some locations with low RootDEP did not result in high maize yields (Supplementary Figure 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 Figure S 5 a, 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 Figure 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 et al. (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 et al., 1986; Kravchenko et al., 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 et al. (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 (Figure 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 (μ = 1 .68 t ha-1, σ = 1) than in TZ (μ = 2.82 t ha-1, σ = 1.3) and DF (μ = 3.05 t ha-1, σ = 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 suggest 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 used 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.

Author contribution

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

Acknowledgement

This study received funding from the Université Mohammed VI Polytechnique and the OCP group, through the Fertilizer Research and 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.

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.

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. *International Journal of 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. *Applied and Environmental Soil Science*, *2012*, 870948. https://doi.org/10.1155/2012/870948

Adjei, V., & Kyerematen, R. (2018). Impacts of Changing Climate on Maize Production in the Transitional Zone of Ghana. *American Journal of Climate Change*, *07*(03), 463-476. https://doi.org/10.4236/ajcc.2018.73028

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. (2021). *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. (2021). Characterization of farmers and the effect of fertilization on maize yields in the Guinea Savannah, Sudan Savannah, and Transitional agroecological zones of Ghana. *EFB Bioeconomy Journal*, *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. *International Journal of 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 Journal of Soil Science*, *10*(03), 91-108. https://doi.org/10.4236/ojss.2020.103005

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 Research Journal of Agriculture and Food Science*, *5*(3), 133-140, Article DRJA43071532. http://directresearchpublisher.org/aboutjournal/drjafs

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. *Journal of Disaster Research*, *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 Geoscience*, *2*(1), 1201906. https://doi.org/10.1080/23312041.2016.1201906

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. *Quarterly Journal of the Royal Meteorological Society*. 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. *Journal of Plant Nutrition*, *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. *Journal of Agriculture and Food Research*, *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. *Journal of Statistical Software*, *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 A. Bationo, et al. (Eds.), *Improving the Profitability, Sustainability and Efficiency of Nutrients Through Site Specific Fertilizer Recommendations in West Africa Agro-Ecosystems* (Vol. 1, pp. 93-138). Springer International Publishing AG 2018. https://doi.org/10.1007/978-3-319-58789-9_7

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. *Advances in Agriculture*, *2021*, 1-8. https://doi.org/10.1155/2021/1765251

Bindraban, P. S., Stoorvogel, J. J., Jansen, D. M., Vlaming, J., & Groot, J. J. R. (2000). Land quality indicators for sustainable land management: proposed method for yield gap and soil nutrient balance. *Agriculture, Ecosystems and Environment*(81), 103-112. https://doi.org/10.1016/S0167-8809(00)00184-5

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 and management*, *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., & P.S., B. (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

Buah, S. S. J., Ibrahim, H., Derigubah, M., Kuzie, M., Segtaa, J. V., Bayala, J., Zougmore, R., & Ouedraogo, M. (2017). Tillage and fertilizer effect on maize and soybean yields in the Guinea savanna zone of Ghana. *Agriculture & Food Security*, *6*(1), 1-11. https://doi.org/10.1186/s40066-017-0094-8

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. *The Journal of Agricultural Science*, 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. *African Journal of Agricultural Research*, *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 Journal of Agricultural Science*, *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. *Experimental Agriculture*, *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. *Archives of Agronomy and Soil Science*, *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. 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

Ezui, K. S., Franke, A. C., Ahiabor, B. D. K., Tetteh, F. M., Sogbedji, J., Janssen, B. H., Mando, A., & Giller, K. E. (2017). Understanding cassava yield response to soil and fertilizer nutrient supply in West Africa. *Plant and Soil*, *420*(1-2), 331-347. https://doi.org/10.1007/s11104-017-3387-6

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, 81:583-583.*

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. *African Journal of Environmental Science Technology*, *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. *International Journal of Climatology*, *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 Science Society of America Journal*, *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 Journal of Applied Ecology*, *27*(1), 108 126-108 126. https://www.ajol.info/index.php/wajae/article/view/189216

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 Sensing Applications: Society and Environment*, *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* (pp. 176-194). Routledge. 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. *Environmental Research Letters*, *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. *Applied Sciences*, *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

IFDC. (2012). *Ghana Fertilizer Assessment*. International Fertilizer Development Center Muscle Shoals Alabama 35662 USA. http://ghana.countrystat.org/fileadmin/user_upload/countrystat_fenix/congo/docs/Ghana%20Fertilizer%20Need%20Assessment%20for%20USAID%20FtF%20AFAPIFDC%202012.pdf

ISRIC. *Africa SoilGrids nutrients*. ISRIC. Retrieved 12 Apr. 2021 from https://data.isric.org/geonetwork/srv/eng/catalog.search#/search

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. *Journal of Crop Improvement*, *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. *International Journal of Scientific Research in Agricultural Sciences*, *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 ? *Agronomy Journal*(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 & Agriculture*, *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. Journal of Statistical Software, 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. *Advances in Agriculture*, *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 degradation development*, *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-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. *Journal of Plant Sciences*, *6*(4), 165-173. https://doi.org/10.3923/jps.2011.165.173

MacCarthy, D. S., Vlek, P. L. G., & Fosu-Mensah, B. Y. (2012). The Response of Maize to N Fertilization in a Sub-Humid Region of Ghana: Understanding the Process Using a Crop Simulation Model. In J. Kihara, et al. (Eds.), *Improving Soil Fertility Recommendations in Africa using the Decision Support System for Agrotechnology Transfer (DSSAT)*. Springer Science + Business Media. https://doi.org/10.1007/978-94-007-2960-5_5

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

Mobilian, C., & Craft, C. B. (2021). Wetland Soils: Physical and Chemical Properties and Biogeochemical Processes. *Reference Module in Earth Systems and Environmental Sciences, Elsevier*. https://doi.org/10.1016/B978-0-12-819166-8.00049-9

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. *Experimental Agriculture*, *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., 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. *Journal of the Royal Society Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non- Gaussian data: a practical guide for biologists. *Biological Reviews*, *85*(4), 935-956. https://doi.org/10.1111/j.1469-185x.2010.00141.x

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, *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 Sensing*, *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. *International Journal of Geosciences*, *5*(7). https://doi.org/10.4236/ijg.2014.57060

Nyuor, B. A., Donkor, E., Aidoo, R., Buah, S. S., Naab, J., Nutsugah, S., Bayala, J., & Zougmoré, R. (2016). Economic Impacts of Climate Change on Cereal Production: Implications for Sustainable Agriculture in Northern Ghana. *Sustainability*, *8*(8). https://doi.org/10.3390/su8080724

O'Halloran, I. P., Kachanoski, R. G., & Stewart, J. W. B. (1985). Spatial variability of soil phosphorus as influenced by soil texture and management. *Canadian journal of soil science*, *65*(3), 475-487. https://doi.org/10.4141/cjss85-051

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. *Bioresource Technology*, *51*(2-3), 117-123. https://doi.org/10.1016/0960-8524(94)00103-8

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. *Tropical Science*, *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 and management*, *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 Science Society of America Journal*, *50*(2), 401-408. https://doi.org/10.2136/sssaj1986.03615995005000020029x

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. In *African Handbook of Climate Change Adaptation* (pp. 1721-1733). https://doi.org/10.1007/978-3-030-45106-6_144

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 African Journal of Applied Ecology*, *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

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. *Agricultural Systems*, *187*. https://doi.org/10.1016/j.agsy.2020.103016

Peprah, K. (2012). Rainfall and temperature correlation with crop yield: The case of Asunafo forest, Ghana. www.ijsr.net

Peterson, R. A. (2021). Finding Optimal Normalizing Transformations via bestNormalize. *The R Journal*, *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 Mining and Knowledge Discovery*, *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. *Agronomy Journal*, 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. *Journal of Agricultural Science*, *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. *Agronomy Journal*, *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 Crops Research*, *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. *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 Crops 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. *Experimental Agriculture*, 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. *Frontiers in Ecology the Environment*, *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. *Journal of Agronomy and Crop Science*, *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. *Nutrient Cycling in Agroecosystems*, *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*. I. I. o. T. Agriculture. https://www.researchgate.net/publication/335581532_Suitability_of_soils_for_cereal_cropping_in_northern_Ghana/link/5d6e6bbf299bf16522f2b9b6/download

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. *Agricultural Systems*, *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. *Agronomy Journal*, *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 and management*, *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 Crops Research*, *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, Journal of Agricultural Science*, *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. *Journal of Statistical Software*, *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 Crops Research*, *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 Journal of Agricultural and Biological Science*, *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. *Nutrient Cycling in Agroecosystems*, *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

Supplementary data

Appendix A (*Supplementary Table*)

Appendix B (*Supplementary Figure*)

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.
[*] Corresponding author: pbindraban@ifdc.org (P.S. Bindraban).

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 hectares 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., 2021).

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 et al., 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 et al., 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. (2021) 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 more than 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 et al., 2012; Xu et al., 2013; Ren et al., 2015), and specifically recommended for use in Ghana as well (Wijayanto et al., 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 equation (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:

$$\text{Maize yield} = \text{Environment}*\text{Soil}*\text{Variety}*\text{Fertilizer} \qquad (1)$$

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~~trials~~ 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) (Figure 1~~Figure 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

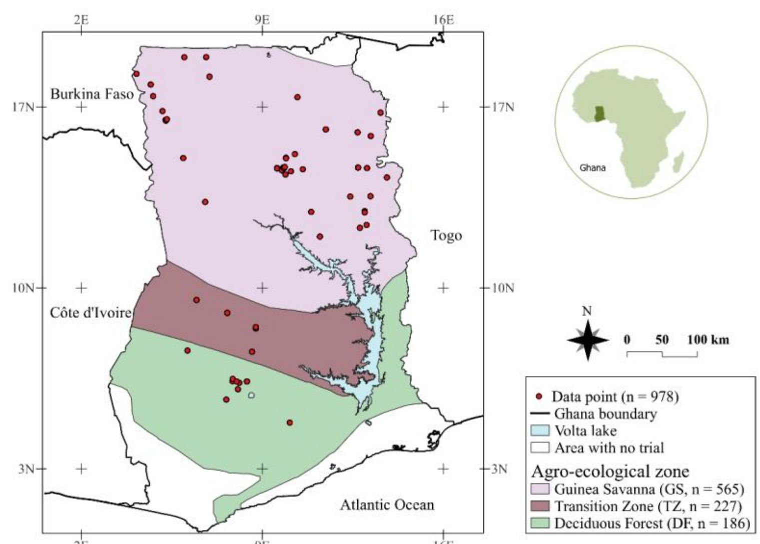


Figure 1: Geographic distribution of the experimental locations on a map of Ghana that shows the country's agro-ecological zones. Many experimental locations overlap.

Agroecological map source: (Antwi et al., 2014)

(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~~ 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 (Figure 1~~Figure 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~~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 et al., 2017) database, at a spatial resolution of 1 km^2 and land elevation data with a resolution of 30 meters 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 et al., 2000)
Monthly precipitation	mm	MP [1]	(Fick et al., 2017)
Monthly mean temperature	°C	MMeT [2]	
Monthly min temperature	°C	MMiT [2]	
Monthly max temperature	°C	MMaT [2]	
Soil organic carbon	$g\ kg^{-1}$	SOC	(Hengl et al., 2015)
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	
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.

Formatted Table

Formatted Table

2.5 QUEFTS model

QUEFTS was implemented in the R software, based on Sattaria et al. (2014). QUEFTS' calibration and validation parameters from Wijayanto et al. (2012) and Antwi et al. (2017) were used to replace the 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_4Ac extracted from the ISRIC maps were converted into P-Olsen and K-Mehlich3 extractable based on Sawyer et al. (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~~-value~~ < 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 ~~S2~~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 Figure 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~~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 ~~Table 1~~. The "vif()" function of the car R package (Fox et al., 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 less than 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 less than 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 equation 2, where Y_{ijk} is the *k*th observed value of yield formed by the *i*th level of treatment effect, the *j*th level of AEZ effect, the *ij*th level of interaction between treatment and AEZ, and the error (ϵ_{ijk}).

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

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 et al., 2010; McHugh, 2011).

One-way ANOVA equations 3, 4 and 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 et al., 2010). The one-way ANOVA equations with random effects were constructed as follows:

$$Y_{ij\_Treat} = Treatment_i + \epsilon_{ij} \quad (3)$$
$$Y_{ij\_AEZ} = AEZ_i + \epsilon_{ij} \quad (4)$$
$$Y_{ij\_Year} = Year_i + \epsilon_{ij} \quad (5)$$

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 ~~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 ~~covariate~~ 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}, ~~using the "step()" function~~ 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}) ~~in DF~~. 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:

$$(\text{LMM } 6_{\text{ENV}}) \quad Y_{ijk\_\text{ENV}} = f(\text{ENV}\cancel{\text{Environmental factors}})_I + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \qquad (6)$$

$$(\text{LMM } 7_{\text{SPP}}) \quad Y_{ijk\_\text{SPP}} = f(\text{SPP}\cancel{\text{Soil physical properties}})_i + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \qquad (7)$$

$$(\text{LMM } 8_{\text{SCP}}) \quad Y_{ijk\_\text{SCP}} = f(\text{SCP}\cancel{\text{Soil chemical properties}})_i + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \qquad (8)$$

$$(\text{LMM } 9_{\text{VAR}}) \quad Y_{ijk\_\text{VAR}} = f(\text{VAR}\cancel{\text{Maize variety type}})_i + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \qquad (9)$$

$$(\text{LMM } 10_{\text{FER}}) \quad Y_{ijk\_\text{FER}} = f(\text{FER}\cancel{\text{Fertilizer rate}})_i + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \qquad (10)$$

$$(\text{LMM } 11_{\text{step}}) \quad Y_{ijk\_\text{step}} = f(\text{ENV}\cancel{\text{Environment}}, \text{SPP}\cancel{\text{Soil}}, \text{SCP}, \text{VAR}\cancel{\text{Variety}}, \text{FER}\cancel{\text{Fertilizer}})_i + \text{AEZ}_j + \text{Year}_k + \epsilon_{ijk} \qquad (11)$$

$$(\text{LMM } 12_{\text{ENV\_AEZ}}) \quad Y_{ij\_\text{ENV\_AEZ}} = f(\text{ENV}\cancel{\text{Environmental factors}})_i + \text{Year}_j + \epsilon_{ij} \qquad (12)$$

$$(\text{LMM } 13_{\text{SPP\_AEZ}}) \quad Y_{ij\_\text{SPP\_AEZ}} = f(\text{SPP}\cancel{\text{Soil physical properties}})_i + \text{Year}_j + \epsilon_{ij} \qquad (13)$$

$$(\text{LMM } 14_{\text{SCP\_AEZ}}) \quad Y_{ij\_\text{SCP\_AEZ}} = f(\text{SCP}\cancel{\text{Soil chemical properties}})_i + \text{Year}_j + \epsilon_{ij} \qquad (14)$$

$$(\text{LMM } 15_{\text{VAR\_AEZ}}) \quad Y_{ij\_\text{VAR\_AEZ}} = f(\text{VAR}\cancel{\text{Maize variety type}})_i + \text{Year}_j + \epsilon_{ij} \qquad (15)$$

$$(\text{LMM } 16_{\text{FER\_AEZ}}) \quad Y_{ij\_\text{FER\_AEZ}} = f(\text{FER}\cancel{\text{Fertilizer rate}})_i + \text{Year}_j + \epsilon_{ij} \qquad (16)$$

$$(\text{LMM } 17_{\text{step\_AEZ}}) \quad Y_{ij\_\text{AEZ\_step}} = f(\text{ENV}\cancel{\text{Environment}}, \text{SPP}\cancel{\text{Soil}}, \text{SCP}, \text{-VAR}\cancel{\text{Variety}}, \text{FER}\cancel{\text{Fertilizer}})_i + \text{Year}_j + \epsilon_{ij} \qquad (17)$$

Formatted: Left

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 $_{\text{step\_DF}}$ fixed effect trend "$f()_i$". The predictors groups, i.e., ~~environmental factors (~~ENV~~)~~, ~~soil physical properties (~~SPP~~)~~, ~~soil chemical properties (~~SCP~~)~~, ~~maize variety (~~VAR~~)~~, and ~~fertilizer rate (~~FER~~)~~, involved in the fixed-effect statistics modellings, except for LMMs 11_{step} and $17_{\text{step\_AEZ}}$ and $\text{MLR}_{\text{step\_DF}}$, 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 ~~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*).

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

The measure of variable importance worked by calculating the increase in RF's prediction error after permuting the variables in equation 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 et al., 2021).

$$Y_{i\_\mathrm{rf}} = f(\mathrm{ENV}\cancel{\mathrm{Environment}}, \mathrm{SPP}\cancel{\mathrm{Soil}}, \mathrm{SCP}, \text{-}\mathrm{VAR}\ \cancel{\mathrm{Variety}}, \mathrm{FER}\cancel{\mathrm{Fertilizer}})_i + \epsilon_i \qquad (18)$$

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 et al., 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 et al., 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 ~~S3~~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 (~~Table 1~~ 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 ~~elevation~~ 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} (Supplementary~~-~~Table S 3~~Table S4~~). 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~~Table S5~~). 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 ~~Table S6~~). Tukey's HSD test found that

Field Code Changed

the mean yield was significantly different between fertilized and control data points ($p < 0.05$) (Supplementary Table S 5 ~~Table S7~~, Figure 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

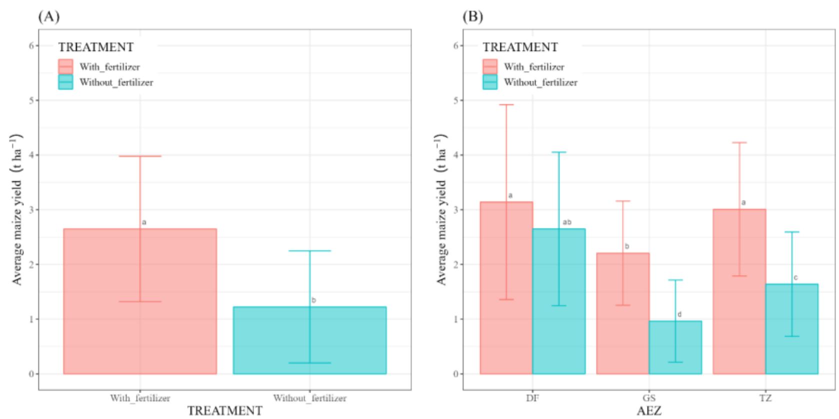


Figure 2: Average maize yield (t ha^{-1}) of treatments with fertilizer and without fertilizer (A) and the interaction between treatments and agro-ecological zone (AEZ) (B). DF (Deciduous Forest), GS (Guinea Savanna), and TZ (Transition Zone). Error bars are standard deviation bars. Different letters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

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 ~~Table S8~~, Figure 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~~7~~%, 20~~4~~%, and 33% of the total variation in yield, respectively (Supplementary Table S 6, S 7 and S 8 ~~Table S9, S10, and S11~~).

3.3 QUEFTS model yield estimated

When considering all the data points, the ~~estimated~~ average of estimated yield~~s~~ ~~using~~ from the QUEFTS model ~~and observed yields averaged~~was 3.3 t ha^{-1} ~~and 2.2 t ha^{-1}, respectively~~. 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 (Figure 3 A).

Figure 3~~Figure 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

Formatted: Indent: First line: 0"

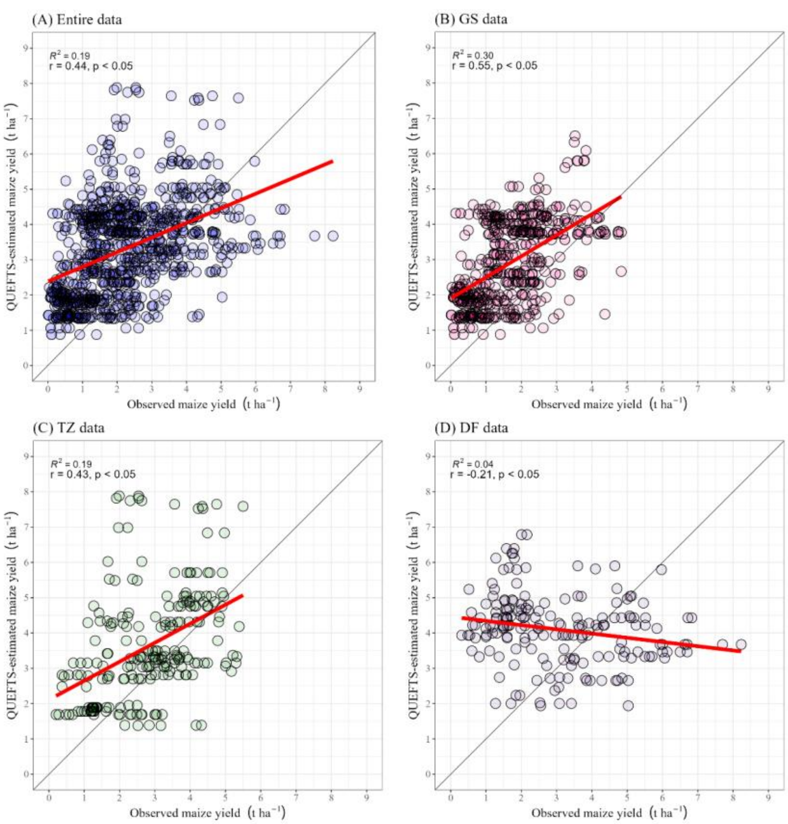


Figure 3: Relationship between observed and QUEFTS-estimated maize yield (t ha^{-1}) across all data points (A) and per AEZ (B, C, D). The R^2 indicates the portion of yield variability explained by soil chemical properties and fertilizers (confounded effect). The bold red line represents the linear 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.

yields in all 3 AEZs (R^2_{DF} [4%] < R^2_{TZ} [19%] < R^2_{GS} [30%], Figure 3~~Figure 3~~B, 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 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 Figure 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 Figure 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 overestimated 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~~Table S12~~). A correlation matrix and a plot of yield on the CEC did not show a clear trend (Supplementary Figure 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 Figure S 3 a, c). Compared to P, K, pH, and CEC, the variation in yield due to SOC was greater (F value = 0.35) (Table 2).

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
LMM 11_{step}(†)	CLAY	5.4	1	959.9	12.3	< 0.05	0.26	0.53	0.17
	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			

Formatted Table

Formatted: Subscript

VARIETY	53.8	2	960.4	60.9	< 0.05
NF	84.9	1	957.9	192.3	< 0.05
PF	27.4	1	963.5	61.9	< 0.05

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

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~~Table S13~~). 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 Figure S 3 a).

In the LMM analysis, MP was the only significant ($p < 0.05$) environmental predictor (ENV) of yield (Supplementary Table S 9~~Table S14~~), but the correlation matrix showed that MMeT and ELV were likewise significantly ($p < 0.05$) related to the variation in maize yield (Supplementary Figure S 3 a). A plot of maize yield on MP showed a significant ($p < 0.05$) and clear negative trend (Supplementary Figure S 3 b~~3b~~).

Maize variety type (VAR) was significantly ($p < 0.05$) related to yield (Table 2Table 2, Supplementary Table S 9~~Table S15~~). 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~~Table S16~~). Tukey's HSD test found that there was no significant difference in mean yield between OPV and hybrid ($p = 0.54$~~3~~), 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) ~~NF, PF, and KF~~ were significantly ($p < 0.05$) associated with maize yield variability (Table 2, Supplementary Table S 9~~Table S17~~). 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 ~~fertilizer rate~~ had the highest explanatory power ($R^2_m = 0.17$) of maize yield variability, followed by the SCP ~~chemical properties of the soil~~ ($R^2_m = 0.16$), the ~~variety of maize~~VAR ($R^2_m = 0.05$), ~~environmental factors~~ENV ($R^2_m = 0.03$) and finally the ~~physical properties of the soil~~SPP ($R^2_m = 0.02$) (Supplementary Table S 9~~Table S18~~). The ICC values in Table 2~~Table 2~~ and Supplementary Table S 9~~Table S19~~ 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~~Table S20~~). However, the contributions of ELV and CEC to the variations in maize yield in GS did not show a clear and significant trend (Supplementary Figure S 4 a, b, e). The PF rate accounted for 48% of the R^2_m (0.39); 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 ~~fertilizer rate~~FER, followed by ~~maize variety~~VAR (R^2_m = 0.4), ~~soil physical~~SPP and ~~chemical properties~~SCP (R^2_m = 0.03), and ~~environmental factors~~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~~Table 4~~, Supplementary Table S 12~~Table S21~~). The combined contribution of ELV, CLAY, VARIETY, NF, and PF explained 55~~1~~% 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 ~~environmental factors~~ENV (R^2_m = 0.21), then ~~maize variety~~VAR (R^2_m = 0.20), ~~soil chemical properties~~SCP (R^2_m = 0.19), and finally ~~soil physical properties~~SPP (R^2_m = 0.19).

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

Yield variable	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 equation (17) ~~(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 equation (17) ~~(17)~~ in the Transition Zone using stepwise linear mixed effects modelling.

In the MLR analysis, ELV, MP, SOC, CLAY, SAND, RootDEP, TOTP, P, K, CEC, VARIETY, ~~and~~ NF and PF were significant ($p < 0.05$) predictors of maize yield in DF, explaining 79% of maize yield variability (Table 5). ~~The PF rate was not significantly related to maize yield variability in Table 5 but weakly related to yield in the supplementary Table S22.~~ The ~~LMM~~ 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~~Table S23~~). 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 ~~soil chemical properties~~SCP ($R^2_m = 0.48$), followed by ~~environmental factors~~ENV ($R^2_m = 0.43$), ~~maize variety~~VAR, and ~~soil physical properties~~SCP ($R^2_m = 0.24$).

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.

LMM analysis showed that the contribution of maize variety type in the variations in yield in the 3 AEZs was also significant ($p < 0.05$) (Table 3, 4, 5~~Table 3, Table 4, and Table 5~~, Supplementary Table S 14, S 15, S 16~~Table S24, S25, S26~~). 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~~Table S27, S28, S29~~). Tukey's HSD test found that there was no significant difference in mean yield between OPV and hybrid in GS ($p = 0.25$~~3~~) and TZ ($p = 0.69$~~1~~), 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~~GS, TZ, and DF~~.

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 Figure ~~S1A~~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 (Figure 4 ~~Figure 4~~A, Supplementary Figure S 7~~2~~ 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.

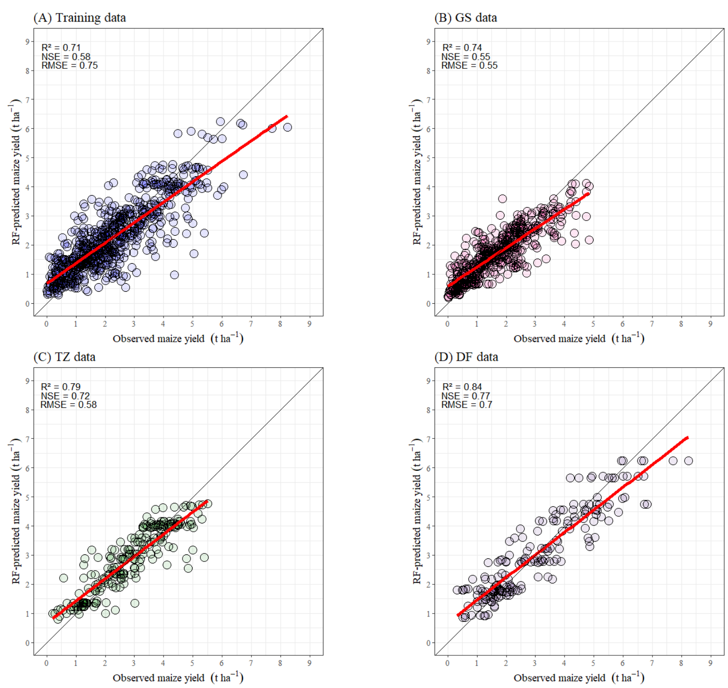


Figure 4: Relationship between observed maize yield and that predicted by the Ran dom Forest (RF) model using 3 repeats of a 10-fold cross-validation approach. (A) The linear regression on ~~the~~ 75% of the ~~full~~ entire data ~~set~~point (training data point), (B) only the data ~~set~~ point from GS, (C) only the data ~~set~~ point from TZ, and (D) only the data ~~set~~ point from DF. The solid red lines show the linear regressions fitted to the data ~~set~~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).

4 DISCUSSION

~~The difference in mean maize yield observed per Treatment*AEZ could be attributed to various biotic and abiotic constraints in the AEZs (Onduru et al., 2007; Mugwe et al., 2009). In TZ and DF, both the non-fertilized and fertilized trials performed better than in GS. This is because soil properties in DF and TZ were generally better than in GS. In addition, in GS, the climatic conditions were much more drastic; for example, the average monthly minimum (MMiT) and maximum (MMaT) temperatures were higher (Supplementary Table S30).~~

~~Data points in DF had high maize yields of 7-8 t ha^{-1}, which could be explained by the low MMiT (19°C) and MMaT (27°C) at the trial locations, sufficient MP during the crop growth period (520 mm), high ELV (390 m), and the deep RootDEP of 150 cm for storing the rainwater. However, several studies on plant breeding have shown that the agronomic potential of certain maize varieties, such as Obatanpa (OPV), is between 4 and 5 t ha^{-1} (Sallah et al., 2007; Adu et al., 2014; USAID/IFDC, 2015). The potential yields attributed to these varieties are averages from several research station trials (Sallah et al., 2007; Adu et al., 2014). Therefore, yields above these averages could be obtained in research trials under optimal agronomic practices and climatic conditions, which justifies why, in this study, the high yields of those maize trials were not considered outliers. In addition, Adu- Gyamfi et al. (2019) reported average yields of 6.5 t ha^{-1} with Obatanpa, well above its potential yield (according to research institutions), in GS which could be considered less fertile than TZ and DF given the soil and climatic conditions.~~

4.1 Soil chemical properties and fertilizer rates impacting~~controlling~~ yield variability

~~In general, the QUEFTS model weakly explained the spatial and temporal observed maize 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, t~~T~~he low spatial and temporal explanatory power of ~~the~~ QUEFTS ~~model~~ suggests that other ~~that~~ factors ~~other~~ than ~~the~~ soil chemical properties and fertilizer rates only, ~~to~~ ~~not considered in QUEFTS~~ contribute~~d~~ to the variability in maize yield as also reported by Debtanu et al. (2006) and ~~(~~Onduru et al.~~,~~ (2007). ~~Thus, the QUEFTS is unable to improve the understanding of native soil nutrient supply, and maize yield for the wide range of soil, climate, and management conditions in the 3 AEZs, as also reported by Debtanu et al. (2006).~~ Previous studies in Kenya, Benin, and Rwanda have

Formatted: Indent: First line: 0"

Field Code Changed

Field Code Changed

Formatted: Not Highlight

also shown low precision and accuracy of QUEFTS maize yield predictions (Mulder, 2000; Onduru et al., 2007; Breure et al., 2022). ~~However, the high yield variability ($R^2 > 50\%$) captured by QUEFTS in 2001 and 2008 confirms that the variability in yield is not only spatial but also temporal.~~ While~~In general, the~~ soil properties do not reveal~~show~~ strong temporal dynamics, ~~but the~~ rainfall and temperature in 2001 (MP = 502 - 698 mm, MMaT = 27 - 29°) and 2008 (MP = 709 mm, MMaT = 30°) were favorable~~, allowing optimal growth and development of the maize crops so that the observed yields were closer to what could be expected from fertilization and more in line with the presumed~~ better mimicking the presumed growth conditions in QUEFTS when its explanatory power was indeed highest ($R^2 > 50\%$). In addition~~ to the effect from climatic factor~~, the importance of soil physical properties, such as texture and soil depth, ~~which were identified to be important but~~ are also not considered in QUEFTS~~, likely also contributed to the poor explanatory power of yield variability by the QUEFTS~~.

~~The QUEFTS model explained more variability in GS than in the other 2 AEZs ($R^2_{GS} > R^2_{TZ} > R^2_{DF}$). Indeed, the model was calibrated and validated in the northern regions (i.e., the regions belonging to GS), which could explain why the model captured nearly 30% of the variability in GS compared to the other 2 AEZs. This also confirms the heterogeneity of soil chemical properties and the presence of a fertility gradient.~~

~~The model overestimated grain yields at almost all data point locations. This overestimation revealed bias in the model calibration. The accumulation and dilution coefficients and N, P, and K recovery efficiency (RE) reported by Antwi et al. (2017) and Wijayanto et al. (2012) need to be updated and adjusted considering the soil and climate specificity of the AEZs (Ezui et al., 2017). According to IFDC (2012), the RE for maize in Ghana is 0.50 kg N kg^{-1}, 0.35 kg P kg^{-1}, and 0.70 kg K kg^{-1}. Antwi et al. (2017) and Wijayanto et al. (2012) also obtained similar RE values; however, these RE values did not align with the RE in this study.~~

~~Contrary to the QUEFTS modelling,~~The LMM/MLR and RF models un~~re~~vei~~a~~led the importance of fertilizer rate (FER) and soil chemical properties (SCP) in maize yield variability as well, as corroborated by Braimoh and Vlek (2006). ~~According to the LMM analysis summarized in Table 2Table 2, O~~Of the nine ~~eleven 11~~ factors that were significantly ($p < 0.05$) associated with maize yield variability in the LMM analysis (Table 2), 5 ~~(or nearly half)~~ were related to the ~~soil chemical properties~~SCP. ~~SOC,~~The NF rate was found to be the most significant

Formatted: Superscript

($p < 0.05$) 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 (Figure S ~~3~~3 ~~a~~a), also identified by ~~. Indeed, according to~~ Braimoh et al. (2006)~~, maize yields will be less than 520 kg ha^{-1} under repeated cropping on soils with poor chemical properties.~~ Yeboah et al. (2016) and ~~also showed that improving soil chemical properties, such as TOTN, P, and K, CEC, and pH, had a pronounced effect on maize grain yield (~~Akolgo et al.~~,~~ (2020). ~~Therefore, knowledge about the condition of the soil's chemical properties and its evolution over time and space is necessary to promote land management practices to maintain and improve maize productivity and sustainable use of natural resources (Bindraban et al., 2000). However, a plot of yield on CEC and soil pH revealed an inverse trend (Supplementary Figure S 4c3 c, d).~~

~~Although~~ Ghanaian soils are generally not very acidic (Supplementary Table S ~~31~~3) and fairly optimal, however~~with~~ the slightly negative relation with maize yield likely result~~sing~~ from the narrow pH range and associations with other variables such as ~~, the ideal pH range for maize production is between 6 and 7 (Buah et al., 2017), and a reduction in pH will generally result in a negative trend for maize yield. However, because there was a positive and substantial association between pH and~~ ~~MMeT and~~ temperature (Supplementary Figure S~~5~~ 3 a, c, 5 a)~~, it is likely that this negative trend between pH and yield was caused by the influence of another variable, such as temperature, as the pH did not vary too much and the standard deviation was extremely small~~.

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 ~~CEC values for~~ 30 data points ~~situated~~ in the DF suggested an negative though not significant ($p = 0.09$) relation with yield when the entire data point is considered (Supplementary Figure S 3 d). ~~were much higher than average (Supplementary Figure S 26 e d), which accounted for the non-significant ($p = 0.09$) negative trend between CEC and yield when the entire data point is considered (Supplementary Figure S 3 d). According to Osei (1995), tropical soils like those in the DF generally have low CEC (typically 10 cmol$_{(+)}$ kg^{-1}), and as a result, these extreme values (83 cmol$_{(+)}$ kg^{-1}) influenced the trend of linear relationship between CEC and yield (Supplementary Figure S 3 e). After~~ ~~e~~Excluding the~~se~~se data points turned the relation ~~excessive CEC values from these 30 data points from the entire set of data points, the linear trend became~~ significantly ~~(~~($p < 0.05$~~)~~) ~~and~~ positive

(Supplementary Figure ~~S7e~~ S 3 e). ~~According to Adzawla, et al. (2021), among edaphic factors, CEC is an important one that is related to maize yield.~~

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 ~~and showed that the contribution of~~ KF ~~in~~ explaining 2.5-3.5% of maize yield variability ~~was only 2.5-3.5%~~. ~~These findings suggest that K is not a limiting nutrient for Ghana's maize production. Therefore, t~~The significant ($p < 0.05$) association of K and maize yield in Table 2~~Table 2~~ was probably due to the data points of the DF (Supplementary Table S 13~~Table S32~~, Supplementary Figure S 6 a). According to Yawson et al. (2011), Ghana's forest soils will require frequent and split KF applications since they have a low~~er~~ 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, ~~A~~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 ~~the 3~~ 2 AEZs (Supplementary Figure ~~S4a,~~ S 5 a, S 6 a). ~~This is explained by the fact that, in comparison to soils where poor yields were reported, t~~The bulk of high-yielding data points in ~~GS,~~ TZ and DF were found in soils with a comparatively lower SOC. Th~~is~~e negative trend ~~between SOC and yield in some AEZs data points, despite their inherently high soil SOC, indicates that~~ suggests yield variability to be ~~is~~ driven by other factors as well~~other than SOC~~. In research ~~led~~ by Kihara et al. (2016), Shehu et al. (2019) and Sileshi et al. (2022), maize fertilization trials ~~got poor yields~~ on soils with a relatively high~~larger~~ SOC level obtained poor yields, while~~ereas the trials that obtained the~~ highest yields were found ~~carried out~~ on soils with a low SOC content. Across ~~For t~~The entire set of data points, with a wider SOC range, ~~LMM yield variability modelling revealed a significant and positive trend of maize yield on SOC (Table 2,~~ Supplementary Figure S 3 f ~~a).~~ 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~~and an~~ application of NF in such soil ~~will~~ promoting~~e~~ ~~good plant nitrogen nutrition and thus contribute to a significant and positive influence of NF on~~ maize yield.

~~It was revealed that phosphorus should be considered while analyzing maize yield variability.~~ According to Table 3 and 5 ~~Table 3 and Table 5~~, in GS and DF respectively, PF had the highest (F-value = 60.31) and lowest (F-value = 4.99) contributions to maize yield variability. This might explain~~s~~ why lower rates of PF ~~(rates were an average of~~ 19 kg ha^{-1}~~-~~) are applied in the DF as opposed to 34 kg ha^{-1} in the GS. ~~Additionally, t~~The levels of soil P and TOTP in DF were greater than those in GS (Supplementary Table S 3~~Table S33~~), which may have prevented the PF rate in DF soils from strongly contributing to maize yield.

~~In this study, the NF rate was found to be the most significant (p < 0.05) contributor to the variation in maize yield in Ghana. The fact that the NF rate was substantially related to maize yield across the 3 AEZs revealed that N was the most limiting soil nutrient among the three macronutrients (N, P, and K). According to MacCarthy et al. (2012), maize yield modelling in a sub-humid region of Ghana performed better when N fertilization rates were higher.~~

4.2 Role of s~~S~~oil p~~P~~hysical characteristics ~~role i~~on yield variability

~~Among the variables in the soil physical characteristics group, o~~Only RootDEP demonstrated a significant ($p < 0.05$) positive trend with yield variation when the entire set of maize yield data points was ~~modelled~~ using in LMM. It was indeed~~even~~ identified by the RF model as the third most important variable, right after the NF rate and MMeT (Supplementary Figure S ~~8A~~7 A). ~~But a~~A more specific analysis showed that this influence ~~of RootDEP on maize yields of all the data points~~ originated mostly from the DF data points (Supplementary Table ~~S34~~S 13, Figure ~~S9a~~S 6 a) with deeper soils. ~~These results show how critical soil rooting depth is in maize production in Ghana. A study by~~ Guilpart et al. (2017)~~,~~ revealed this importance of RootDEP on ~~who used crop simulation to assess~~ the sensitivity of rainfed maize yields as function of ~~to~~ the water-holding capacity of the root zone~~, showed that a modest maize surplus could be produced in SSA, including Ghana, provided rooting depths are comparable to those of other important granaries, such as the United States Maize Belt and the South American Pampas (Leenaars et al., 2018)~~.

Sand and clay were found to be significantly ($p < 0.05$) associated with maize yield variability in both LMM and MLR modelling. ~~However, Ghanaian soil being much sandier (Supplementary Table S35~~S 3~~), especially in GS where there is high potential for maize production (Boullouz et al., 2022), it has low fertility and less SOC content (Kanton et al., 2016; Bationo et al., 2018).~~

~~Mobilian et al. (2021) reported that soil texture influences other soil properties, such as bulk density, water holding capacity, permeability, and porosity. For example, soils that are predominantly composed of sand particles have high permeability and low water holding capacity compared to soils with higher clay content.~~ While a negative relation of sand content is expected with yield due to various associated soil properties like low water-holding capacity, ~~This implies that~~ the positive ~~and significant~~ trend ~~between sand and maize yield~~ (Supplementary Figure S 3 a, S 4 a, c, and S 6 a, g) ~~on sandy soils~~ 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 et al., (1995), Zingore et al. (2007), Uzoma et al. (2011) and~~;~~ Frimpong et al., 2021~~; Obi et al., (1995); Zingore et al. (2017),~~ the application of organic fertilizers such as cow and poultry manure ~~biochar~~ on sandy soil ~~is not only beneficial for crop growth, but also~~ significantly improves the physicochemical properties of ~~the coarse~~ sandy soils ~~. Indeed, the application of organic fertilizer tends to improve the soil SOC content, which is~~ favorably affecting ~~e for good~~ maize growth ~~(Frimpong et al., 2021). The same result was found in Nigeria by Obi et al. (1995), who showed that application of poultry manure significantly improved maize yield in severely degraded sandy soil. Zingore et al. (2007) also reported that manure is a key nutrient resource in the tropics, especially on poorly buffered sandy soils, because of its multiple benefits to soil fertility~~. 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.

~~According to our analyses, t~~The contribution of clay content to the variations in maize yield was significant (Table 2~~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 ~~improves~~ameliorates soil physical characteristics such as enhancing water holding capacity. ~~However, i~~In GS, however, a sizable downward ~~upward~~ trend in the maize yield with clay was observed (Supplementary Figure S 4 d). While ~~Although~~ the correlation coefficient was~~seems~~ significant ($p < 0.05$), the distribution of data points in the Supplementary Figure S 4 d does clearly underline ~~not seem to make~~ this negative trend~~very clear~~, likely for similar reasons as indicated for sand. ~~In a study,~~ Sileshi et al. (2022) reports~~showed the sam~~similar~~e~~ results, and ~~. Similarly,~~

Njoroge et al. (2018) reported no significant difference in clay content between sites with high and low yield under ~~when~~ NPK application ~~fertilizer was used and those with low yield~~ in western Kenya.

4.3 Contribution of environmental factors to yield variation

~~Climate has always been central to agricultural production in Ghana (Peprah, 2012; Nyuor et al., 2016; Adjei et al., 2018).~~ The analysis of LMM (Table 2~~Table 2~~, Supplementary Table ~~S36~~S 9, ~~S37~~S 11, ~~S38~~S 12, ~~S39~~S 13, ~~and~~ Figure S~~10~~ 3 a, b~~,c~~, S~~11~~ 5 b, S 6 a) and RF yield modelling (Supplementary Figure S~~12~~ 7 ~~B~~A) showed that environmental variables (MMeT, MP, and ELV) were related to maize yield variability, as also reported by ~~(~~Onduru et al. (2007), Mugwe et al. (2009), Fosu-Mensah et al.~~,~~ (2019~~;)~~), Kyei-Mensah et al.~~,~~ (2019~~;)~~), and Cudjoe et al.~~,~~ (2021). However, ~~the contribution of~~ MP of ~~to~~ the entire data ~~set~~point reveals a negative trend with ~~point~~ yield ~~variation showed a negative pattern~~ (Supplementary Figure S~~13~~ 3 a, b). I~~ndeed, i~~n the DF~~locations~~ where very high yields of ~~(~~7-8 t ha^{-1}~~)~~ were reported~~obtained~~, the amount of rainfall was sufficient, ~~al~~though lower compared to other areas with lower yields.~~, was sufficient~~ 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 et al.~~,~~ (2020) and ~~;~~Bagula et al.~~,~~ (2022). ~~to fill the soil reservoir where deep RootDEP (150 mm) and high ELV (390 m) were observed, so that water would not be a limiting factor for maize growth. In addition, a~~The negative trend between MP and RootDEP support this logic and ~~could also help explain~~ why high MP in some locations ~~(~~with low RootDEP~~)~~ did not ~~tend to~~ result in high maize yields (Supplementary Figure S~~14~~ 6 a, h~~i~~). ~~Furthermore, optimal rainfall in low-yielding areas confirms our view that variables other than rainfall, which had a significant relationship with yield, also play an important role in maize variability.~~

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 Figure S~~15~~ 5 a, 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 ~~increase in maize yield~~ in TZ may have resulted from ~~coinciding with increasing temperatures may be due to other non-climatic factors, such as the use of~~ improved high-potential, drought-tolerant variety and fertilizer rate (Atiah et al., 2021). Drought-tolerant maize varieties in some trials,

Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed

such as GH 110, Mamaba, and Akposoe, could have performed relatively well~~thrived~~ at higher temperatures~~, so lower levels of rainfall could contribute modestly to maize yields~~.

ELV, in addition to MP and MMeT, was also a significant factor in explaining maize yield variability. In fact, the correlation matrices (Supplementary Figure S~~16~~ 3 a, S 4 a and S 5~~17~~a) 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 et al. (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 et al., 1986; Kravchenko et al., 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 et al. (2021) evaluated the influence of different NF rates on maize yield and found differential ~~that~~ varietal~~ies performed differently~~ responses~~depending on the rate NF rate~~, as did ~~(~~Adu et al.~~,~~ (2014) in Ghana, and Abera et al. (2017) in Ethiopia. ~~The use of different types of maize varieties with different genetic potential and drought resistance, could induce highly variable yields depending on AEZ and fertilizer application (Adu et al., 2014). Similar outcomes were also documented in Ethiopia by Abera et al. (2017). Indeed, the use of different types of maize varieties coupled with different fertilizer rates (also highly variable) has undoubtedly contributed to maize yields variability (van Loon et al., 2019; Bawa et al., 2021). For example,~~ 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~~ variety~~) and 14-28 days after planting for hybrid varieties to achieve 4.7 t ha^{-1} and 6.5 t ha^{-1}, respectively. T~~However, t~~he coupled contribution of environmental factors and fertilizer rates could explain the aberrant responses of varieties between the AEZ's.~~lack of significant difference between the mean yield of hybrids and OPV in GS and TZ, the significant difference between the mean yield of hybrids and OPV in DF, and the significant difference between the mean yield of hybrids and OPV with "Others" in GS, TZ, and DF.~~ ~~In fact, a~~A study conducted in DF and TZ by Kpotor et al. (2014) also revealed

Field Code Changed

Formatted: Not Highlight

Formatted: Not Highlight

Field Code Changed

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 (Figure 4)~~Figure 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 (μ = 1 .68 t ha-1, σ = 1) than in TZ (μ = 2.82 t ha-1, σ = 1.3) and DF (μ = 3.05 t ha-1, σ = 2).

QUEFTS is a model based on empirical processes and internal ~~curve~~ 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 ~~explaining~~ 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. ~~This is why a~~ Our multi-model approach ~~was used in this study~~ reveals this added value to unravel driving factors of maize yield. ~~, particularly given that van Loon et al. (2019) claimed that identifying the variables influencing maize yield in Ghana is much more difficult because yields are influenced by interdependent and frequently drastically different climates, soil, and management variables.~~

Formatted: Heading 2

Formatted: Indent: First line: 0.2"

Formatted: Superscript

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 ~~the fertilizer rate,~~ 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 ~~instead guided the~~ explain~~at~~ed most~~ion~~ of variability. In TZ and GS, ~~it was rather~~ NF and ~~N~~PF drove ~~that led the~~ yield explanation. This suggest that ~~may imply~~ 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 ~~high~~ used of ~~variability~~ different types of maize varieties ~~of the variety (OPV, hybrid, or local)~~ 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.

Author contribution

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

Acknowledgement

This study received funding from the Université Mohammed VI Polytechnique and the OCP group, through the Fertilizer Research and 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.

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.

Supplementary data

Appendix A (*Supplementary Table*)

Appendix B (*Supplementary Figure*)

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. *International Journal of 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. *Applied and Environmental Soil Science*, *2012*, 870948. https://doi.org/10.1155/2012/870948

Adjei, V., & Kyerematen, R. (2018). Impacts of Changing Climate on Maize Production in the Transitional Zone of Ghana. *American Journal of Climate Change*, *07*(03), 463-476. https://doi.org/10.4236/ajcc.2018.73028

~~Adu- Gyamfi, R., Agyin- Birikorang, S., Tindjina, I., Ahmed, S. M., Twumasi, A. D., Avornyo, V. K., & Singh, U. (2019). One- Time Fertilizer Briquettes Application for Maize Production in Savanna Agroecologies of Ghana. *Agronomy Journal*, *111*(6), 3339-3350. https://doi.org/10.2134/agronj2019.04.0292~~

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. (2021). *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. (2021). Characterization of farmers and the effect of fertilization on maize yields in the Guinea Savannah, Sudan Savannah, and Transitional agroecological zones of Ghana. *EFB Bioeconomy Journal*, *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. *International Journal of 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 Journal of Soil Science*, *10*(03), 91-108. https://doi.org/10.4236/ojss.2020.103005

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 Research Journal of Agriculture and Food Science*, *5*(3), 133-140, Article DRJA43071532. http://directresearchpublisher.org/aboutjournal/drjafs

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. *Journal of Disaster Research*, *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 Geoscience*, *2*(1), 1201906. https://doi.org/10.1080/23312041.2016.1201906

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. *Quarterly Journal of the Royal Meteorological Society*. 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. *Journal of Plant Nutrition*, *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. *Journal of Agriculture and Food Research*, *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. *Journal of Statistical Software*, *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 A. Bationo, et al. (Eds.), *Improving the Profitability, Sustainability and Efficiency of Nutrients Through Site Specific Fertilizer Recommendations in West Africa Agro-Ecosystems* (Vol. 1, pp. 93-138). Springer International Publishing AG 2018. https://doi.org/10.1007/978-3-319-58789-9_7

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. *Advances in Agriculture*, *2021*, 1-8. https://doi.org/10.1155/2021/1765251

Bindraban, P. S., Stoorvogel, J. J., Jansen, D. M., Vlaming, J., & Groot, J. J. R. (2000). Land quality indicators for sustainable land management: proposed method for yield gap and soil

Field Code Changed

nutrient balance. *Agriculture, Ecosystems and Environment*(81), 103-112. https://doi.org/10.1016/S0167-8809(00)00184-5

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 and management*, *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., & P.S., B. (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

Buah, S. S. J., Ibrahim, H., Derigubah, M., Kuzie, M., Segtaa, J. V., Bayala, J., Zougmore, R., & Ouedraogo, M. (2017). Tillage and fertilizer effect on maize and soybean yields in the Guinea savanna zone of Ghana. *Agriculture & Food Security*, *6*(1), 1-11. https://doi.org/10.1186/s40066-017-0094-8

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. *The Journal of Agricultural Science*, 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. *African Journal of Agricultural Research*, *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 Journal of Agricultural Science*, *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. *Experimental Agriculture*, *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

Formatted: Dutch (Netherlands)

Field Code Changed

Debtanu, M., Das, D. K., & Pathak, H. (2006). Simulation of fertilizer requirement for irrigated wheat in eastern India using the QUEFTS model. *Archives of Agronomy and Soil Science*, *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. 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

Ezui, K. S., Franke, A. C., Ahiabor, B. D. K., Tetteh, F. M., Sogbedji, J., Janssen, B. H., Mando, A., & Giller, K. E. (2017). Understanding cassava yield response to soil and fertilizer nutrient supply in West Africa. *Plant and Soil*, *420*(1-2), 331-347. https://doi.org/10.1007/s11104-017-3387-6

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, 81:583-583.*

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. *African Journal of Environmental Science Technology*, *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. *International Journal of Climatology*, *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 Science Society of America Journal*, *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 Journal of Applied Ecology*, *27*(1), 108 126-108 126. https://www.ajol.info/index.php/wajae/article/view/189216

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 Sensing Applications: Society and Environment*, *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* (pp. 176-194). Routledge. 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. *Environmental Research Letters*, *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. *Applied Sciences*, *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

IFDC. (2012). *Ghana Fertilizer Assessment*. International Fertilizer Development Center Muscle Shoals Alabama 35662 USA. http://ghana.countrystat.org/fileadmin/user_upload/countrystat_fenix/congo/docs/Ghana%20Fertilizer%20Need%20Assessment%20for%20USAID%20FtF%20AFAPIFDC%202012.pdf

ISRIC. *Africa SoilGrids nutrients*. ISRIC. Retrieved 12 Apr. 2021 from https://data.isric.org/geonetwork/srv/eng/catalog.search#/search

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. *Journal of Crop Improvement*, *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.)

Formatted: Dutch (Netherlands)

Field Code Changed

Formatted: Dutch (Netherlands)

Formatted: Dutch (Netherlands)

Genotypes under Different Levels of Nitrogen Fertilization across Two Agro-Ecological Zones in Ghana. *International Journal of Scientific Research in Agricultural Sciences*, *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 ? *Agronomy Journal*(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 & Agriculture*, *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. Journal of Statistical Software, 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. *Advances in Agriculture*, *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 degradation development*, *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-e10853. https://doi.org/10.19053/01211129.v29.n54.2020.10853

Field Code Changed

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

Field Code Changed

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. *Journal of Plant Sciences*, *6*(4), 165-173. https://doi.org/10.3923/jps.2011.165.173

MacCarthy, D. S., Vlek, P. L. G., & Fosu-Mensah, B. Y. (2012). The Response of Maize to N Fertilization in a Sub-Humid Region of Ghana: Understanding the Process Using a Crop Simulation Model. In J. Kihara, et al. (Eds.), *Improving Soil Fertility Recommendations in Africa using the Decision Support System for Agrotechnology Transfer (DSSAT)*. Springer Science + Business Media. https://doi.org/10.1007/978-94-007-2960-5_5

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

Mobilian, C., & Craft, C. B. (2021). Wetland Soils: Physical and Chemical Properties and Biogeochemical Processes. *Reference Module in Earth Systems and Environmental Sciences, Elsevier*. https://doi.org/10.1016/B978-0-12-819166-8.00049-9

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. *Experimental Agriculture*, *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., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R² and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non- Gaussian data: a practical guide for biologists. *Biological Reviews*, *85*(4), 935-956. https://doi.org/10.1111/j.1469-185x.2010.00141.x

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, *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 Sensing*, *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. *International Journal of Geosciences*, *5*(7). https://doi.org/10.4236/ijg.2014.57060

Nyuor, B. A., Donkor, E., Aidoo, R., Buah, S. S., Naab, J., Nutsugah, S., Bayala, J., & Zougmoré, R. (2016). Economic Impacts of Climate Change on Cereal Production: Implications for Sustainable Agriculture in Northern Ghana. *Sustainability*, *8*(8). https://doi.org/10.3390/su8080724

O'Halloran, I. P., Kachanoski, R. G., & Stewart, J. W. B. (1985). Spatial variability of soil phosphorus as influenced by soil texture and management. *Canadian journal of soil science*, *65*(3), 475-487. https://doi.org/10.4141/cjss85-051

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. *Bioresource Technology*, *51*(2-3), 117-123. https://doi.org/10.1016/0960-8524(94)00103-8

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. *Tropical Science*, *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 and management*, *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 Science Society of America Journal*, *50*(2), 401-408. https://doi.org/10.2136/sssaj1986.03615995005000020029x

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. In *African Handbook of Climate Change Adaptation* (pp. 1721-1733). https://doi.org/10.1007/978-3-030-45106-6_144

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 African Journal of Applied Ecology*, *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

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. *Agricultural Systems*, *187*. https://doi.org/10.1016/j.agsy.2020.103016

Peprah, K. (2012). Rainfall and temperature correlation with crop yield: The case of Asunafo forest, Ghana. www.ijsr.net

Peterson, R. A. (2021). Finding Optimal Normalizing Transformations via bestNormalize. *The R Journal*, *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 Mining and Knowledge Discovery*, *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. *Agronomy Journal*, 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. *Journal of Agricultural Science*, *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. *Agronomy Journal*, *93*(3), 577-583. https://doi.org/10.2134/agronj2001.933577x

Field Code Changed

Formatted: Dutch (Netherlands)

Field Code Changed

Field Code Changed

~~Sallah, P. Y. K., Twumasi-Afriyie, S., Ahenkora, K., Asiedu, E. A., Obeng-Antwi, E., Osei-Yeboah, S., Frimpong-Manso, P. P., Ankomah, A., & Dzah, B. D. (2007). Agronomic potentials of quality protein maize hybrids developed in Ghana. *Journal of Agricultural Science and Applications*, *40*(1), 81-89. https://doi.org/10.4314/gjas.v40i1.2157~~

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 Crops Research*, *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. *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 Crops 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. *Experimental Agriculture*, 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. *Frontiers in Ecology the Environment*, *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. *Journal of Agronomy and Crop Science*, *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. *Nutrient Cycling in Agroecosystems*, *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*. I. I. o. T. Agriculture. https://www.researchgate.net/publication/335581532_Suitability_of_soils_for_cereal_cropping_in_northern_Ghana/link/5d6e6bbf299bf16522f2b9b6/download

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. *Agricultural Systems*, *192*. https://doi.org/10.1016/j.agsy.2021.103181

Formatted: Dutch (Netherlands)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

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. *Agronomy Journal*, *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

~~USAID/IFDC. (2015). Seed Guide-Recommended Commercial Maize, Rice and Soybean Varieties Available for Northern Ghana. In U. S. A. f. I. Development (Ed.), (pp. 28): Feed the Future Ghana Agriculture Technology Transfer Project.~~

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 and management*, *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 Crops Research*, *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, Journal of Agricultural Science*, *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. *Journal of Statistical Software*, *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 Crops Research*, *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 Journal of Agricultural and Biological Science*, *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. *Nutrient Cycling in Agroecosystems*, *80*(3), 267-282. https://doi.org/10.1007/s10705-007-9142-2

Field Code Changed

Formatted: Space After: 8 pt

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

Supplementary Material Data

Click here to access/download
Supplementary Material
Supplementary Table.xlsx

Editors Agricultural Systems

Topic: Conflict of interest research paper submitted to Agricultural Systems.

5 April 2023

Dear Editors,

There is no conflict of interest regarding this submission as declared by all authors. The research has been carried out in the FERARI research and implementation program (https://ifdc.org/projects/fertilizer-research-and-responsible-implementation-ferari/) that is financially being supported by the involved institutions with their own contributions and by the Université Mohammed VI Polytechnique and the OCP group.

Anselme K. K. Kouame, Prem S. Bindraban, Isaac N. Kissiedu, Williams K. Atakora, Khalil El Mejahed. IDENTIFYING DRIVERS FOR VARIABILITY IN MAIZE (*Zea mays L.*) YIELD IN GHANA: A META-REGRESSION APPROACH

We look forward to your review and to the publication of the paper.

With kind regards,

Prem Bindraban
International Fertilizer Development Center (IFDC), pbindraban@ifdc.org; Tel: +31624168617

Also, on behalf of Anselme Kouame, Isaac Kissiedu, Williams Atakora, Khalil El Mejahed