

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:	<p>CONTEXT : Maize is the main cereal crop in Ghana, but its production is adversely affected by various biotic and abiotic factors.</p> <p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p>
Suggested Reviewers:	<p>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.</p> <p>Rebbie Harawa, PhD ICRISAT: International Crops Research Institute for the Semi-Arid Tropics</p>

	<p>R.Harawa@cgiar.org Knowledgeable scientist about fertilizer use in Africa.</p>
	<p>Bas Kempen, PhD Researcher, ISRIC World Soil Information bas.kempen@wur.nl One of the authorities in Machine learning methodologies.</p>
	<p>Marc Corbeels, PhD Researcher, IITA: International Institute of Tropical Agriculture m.corbeels@cgiar.org Knowledgeable researcher on agricultural production systems Africa.</p>
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,

A handwritten signature in black ink, appearing to read "Prem Bindraban".

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 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,

A handwritten signature in black ink, appearing to read 'Prem Bindraban', written over a light blue horizontal line.

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

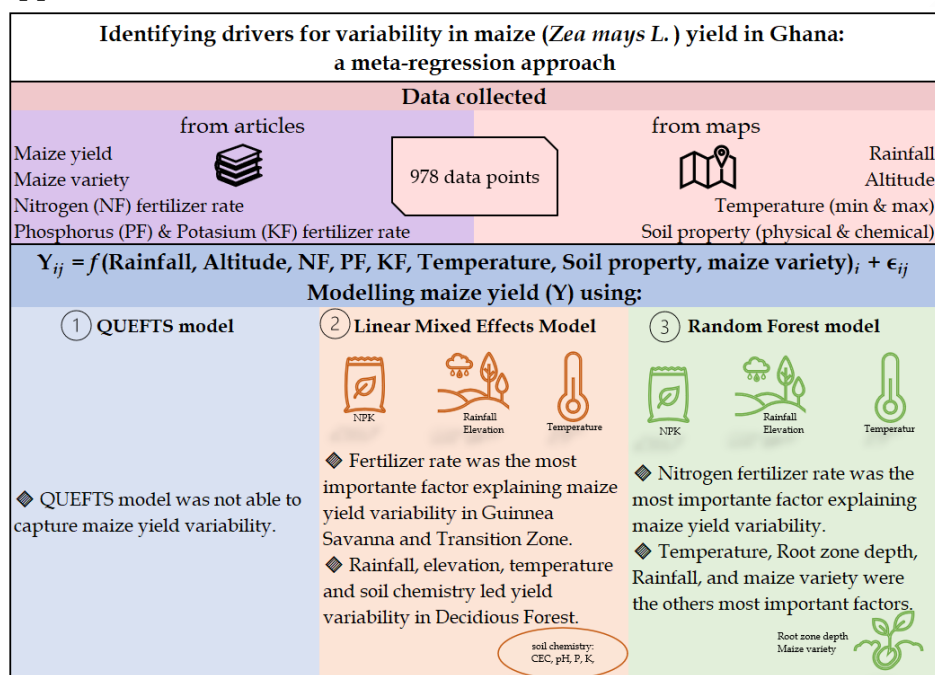
Topic: Submission of research paper on Identifying drives for variability in maize yield in Ghana.

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.

Comment-7: 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.

Comment-8: 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)

Comment-9: 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)

Comment-10: 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.

Comment-10-1 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.

Comment-10-2: 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.

Comment-10.3: 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.

Comment-11: 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.

Comment-12: 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.

Comment-13: 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).

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

2 Anselme K. K. Kouame ^a, Prem S. Bindraban ^{a,*}, Isaac N. Kissiedu ^a, Williams K. Atakora ^a, Khalil El Mejahed ^b

3 ^a International Fertilizer Development Center (IFDC), Muscle Shoals, AL 35662, USA.

4 ^b Mohammed VI Polytechnic University (UM6P), AgroBioSciences, Lot 660, Hay Moulay Rachid, 43150 Ben Guerir, Morocco.

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

6 **Abstract**

7 *CONTEXT:* Maize is the main cereal crop in Ghana, but its production is adversely affected by
8 various biotic and abiotic factors.

9 *OBJECTIVE:* This study aimed to highlight the factors related to maize yield variability. To this
10 end, yields from 978 data points within 3 agro-ecological zones (AEZs) were used in crop-based
11 and statistical modelling.

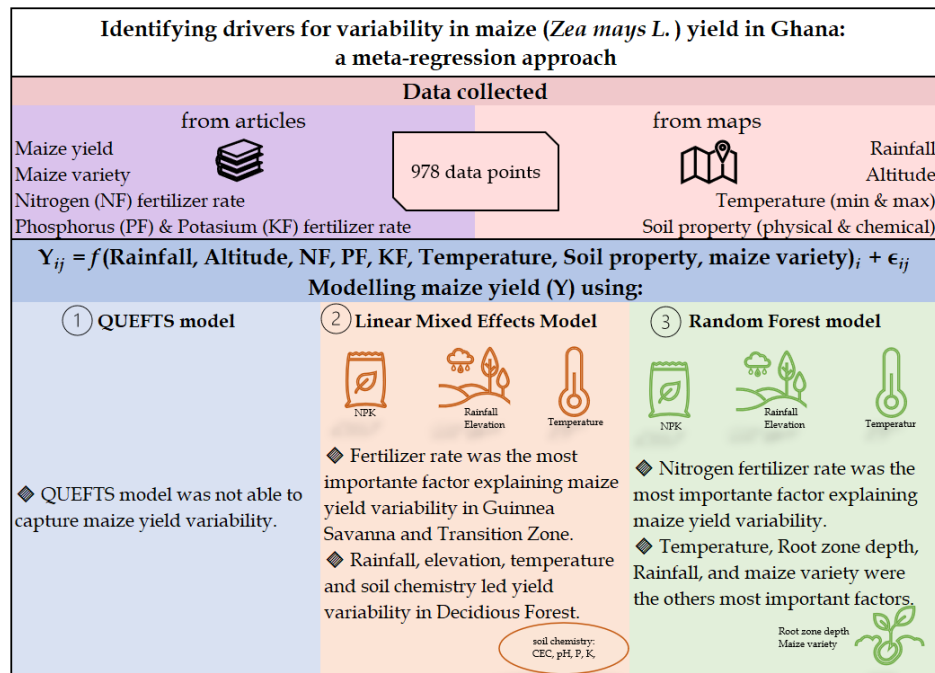
12 *METHODS:* The QUAntitative Evaluation of the Fertility of Tropical Soils (QUEFTS) model, the
13 Linear Mixed Effects Model (LMM), and the Random Forest (RF) model were used to evaluate
14 multiple effect sizes.

15 *RESULTS AND CONCLUSIONS:* Analyzing an entire set of yield data points with QUEFTS, and
16 LMM explained 19%, and 26% of yield variability, respectively. Considering all data points in the
17 RF model, nitrogen fertilizer (NF) rate, temperature, root zone depth, rainfall, and variety
18 accounted for 27%, 15%, 13%, 10%, and 9% of yield variation, respectively. In Guinea Savanna
19 (GS), Transition Zone (TZ), and Deciduous Forest (DF), QUEFTS explained 30%, 20%, and 4% of
20 yield variability, respectively. LMM, however, explained 47%, 51%, and 79% of yield variability
21 in those AEZs. LMM showed that the phosphorus fertilizer (PF) rate was very important and
22 exceeded the importance of the NF rate in GS. LMM showed also that yield variability was
23 significantly related to maize variety at the AEZ scale. In DF, soil chemistry (marginal $R^2 = R^2_m =$
24 0.48) and environmental variables ($R^2_m = 0.43$) contributed more to explaining yield variability,
25 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
26 GS and 0.20 in TZ) played a much larger role. In GS, TZ, and DF, the RF model explained 74%,
27 79%, and 84% of the variance in yield, respectively. These findings suggest low impact of
28 fertilization on yield on the inherently fertile soils in the DF, while fertilization drives yield
29 increase in the less fertile TZ and GS AEZs. We may conclude that QUEFTS was unable to capture
30 yield variability and, according to RF and LMM analysis, the NF rate was the most important

31 factor in explaining yield variability in the data. It can also be concluded that the factors
 32 responsible for yield variability are AEZ dependent.

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

36 Graphical abstract



37 Key words:

- 38 • Maize yield variability,
- 39 • Fertilizer rate,
- 40 • QUEFTS,
- 41 • Linear Mixed Effects Model,
- 42 • Random Forest,
- 43 • Maize yield prediction

44 Highlights:

- 45 • The factors influencing maize yield variability in the agro-ecological zones of Ghana are still unclear.
- 46
- 47 • Overall, nitrogen fertilizer rate was the most important factor explaining maize yield variability.
- 48
- 49 • In the Guinea Savanna and Transition Zone, fertilizer rates and maize variety determined the variability in yield.
- 50

- 51 • In the Deciduous Forest, the environmental factors and soil chemistry were predominant
52 in explaining the yield variability.
- 53 • Random forest modeling showed that root zone depth was also a key factor in explaining
54 maize yield variability in Ghana.

55 **1 INTRODUCTION**

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

65 However, farmers face changing and increasingly unpredictable weather conditions,
66 drastically reducing soil fertility, and typically use local or inbred crop varieties. [Bationo et al.](#)
67 [\(2018\)](#) reported that soil nutrient depletion rates of about 35 kg N, 4 kg P, and 20 kg K per hectare
68 are worrisome and prevalent in all agro-ecological zones (AEZs) in Ghana, with nitrogen (N) and
69 phosphorus (P) being the most deficient nutrients ([Zingore et al., 2015](#)). As a result, yields
70 obtained by smallholder farmers are far below the potentially attainable yields, hampering
71 agricultural production and jeopardizing economic development and food security ([Adzawla, et](#)
72 [al., 2021](#)).

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

80 Despite current low crop productivity, Ghana could intensify production and significantly
81 close the current yield gaps of major cereals ([Bationo et al., 2018](#); [van Loon et al., 2019](#)), since it

82 has been estimated that only about 20%, on average of the potential maize yield is being achieved
83 across the country (GYGA, 2021). For example, addressing nutrient deficiencies by applying
84 fertilizer alone would help to reduce the maize yield gap to 50% of the attainable yield (Mueller
85 et al., 2012). However, Adzawla, et al. (2021) and SRID/MoFA (2021) have reported that the maize
86 yield is around 2 t ha⁻¹ even with the application of nitrogen (N), phosphorus (P), and potassium
87 (K) compound fertilizers. Subsequently, Bua et al. (2020) found that maize yield is highly variable,
88 with yields ranging from a mere 500 kg ha⁻¹ to more than 8 t ha⁻¹. This large yield variability
89 depresses farmers' incentive and ability to purchase fertilizers in subsequent seasons (Njoroge,
90 2019).

91 It is important, therefore, to identify the key drivers for the observed variabilities in maize
92 yield, which can be done using model-based approaches. Various studies around the world have
93 shown that the application of system models (Wallach et al., 2018) can be useful in determining
94 and prioritizing the relative importance of factors that contribute to yield variability (Jeong et al.,
95 2016; Lamos-Díaz et al., 2020; Nevavuori et al., 2020; Paudel et al., 2021; Timsina et al., 2021).

96 Models such as the QUantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) have
97 been advocated by several studies for estimating crop yield (Tabi et al., 2007; Tittonell et al., 2008;
98 Wijayanto et al., 2012; Xu et al., 2013; Ren et al., 2015), and specifically recommended for use in
99 Ghana as well (Wijayanto et al., 2012; Antwi et al., 2017). QUEFTS was developed for maize and
100 considers soil chemical properties (pH, organic carbon, extractable and total phosphorus,
101 exchangeable potassium, and organic nitrogen) and fertilizer application as the input variables
102 (Janssen et al., 1990; Sattaria et al., 2014). The model assumes that all other production factors are
103 optimal and does not consider the maize variety, soil physical properties, or climatic variables.
104 But in real life, maize grain yield could be expressed by equation (1) according to Giller et al.
105 (2013), as the result of "Variety", biophysical "Environment" (temperature and precipitation),
106 "Soil" (acidity, texture, root depth, limiting nutrients), and management, including mineral
107 fertilizers interactions:

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

108 Therefore, it would be useful to complement the QUEFTS model with statistical modelling as
109 part of further research to determine how these variables not included in the QUEFTS relate to
110 maize yield in Ghana (Kihara et al., 2016; van Loon et al., 2019; Atiah et al., 2021). Thus, QUEFTS

111 was utilized in conjunction with the statistical models Linear Mixed Effects Model (LMM) and
112 Random Forest (RF).

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

117 **2 MATERIALS AND METHODS**

118 **2.1 Study area**

119 Maize trials were conducted in three of Ghana's AEZs located between 5° and 15° East and 4° and
120 16° North: DF (n = 186), TZ (n = 227), and GS (n = 565) (Figure 1). The rainy season defined the
121 planting periods of the trials conducted by researchers. Indeed, a large part of the GS has a mono-
122 modal rainy season ([Kranjac-Berisavljevic et al., 1999](#)) and the average monthly temperatures
123 varied between 27°C and 36°C ([Ghansah et al., 2018](#); [Darko et al., 2019](#)). Conversely, TZ and DF
124 have 2 rainy seasons; the major season lasts from March to July and a minor season occurs from
125 September to November, with June registering the highest rainfall ([Nkrumah et al., 2014](#)). The
126 average temperatures in these AEZs are between 24°C and 34°C ([Ghansah et al., 2018](#); [Darko et](#)
127 [al., 2019](#)).

128 **2.2 Yield data**

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

134 **2.3 Fertilizer data**

135 Fertilizer treatments were heterogeneous and included organic fertilizers only, organic fertilizers
136 in combination with inorganic fertilizers, and inorganic fertilizers only. Organic fertilizers,
137 including cow dung; poultry, goat, and sheep droppings; compost; town waste; biochar and palm
138 bunch ash, were all converted into N, P₂O₅, and K₂O rates ([Fening et al., 2009](#); [Adjei-Nsiah, 2012](#);
139 [Kanton et al., 2016](#); [Badu et al., 2019](#)). The lowest and highest rates of nitrogen fertilizer (NF) were

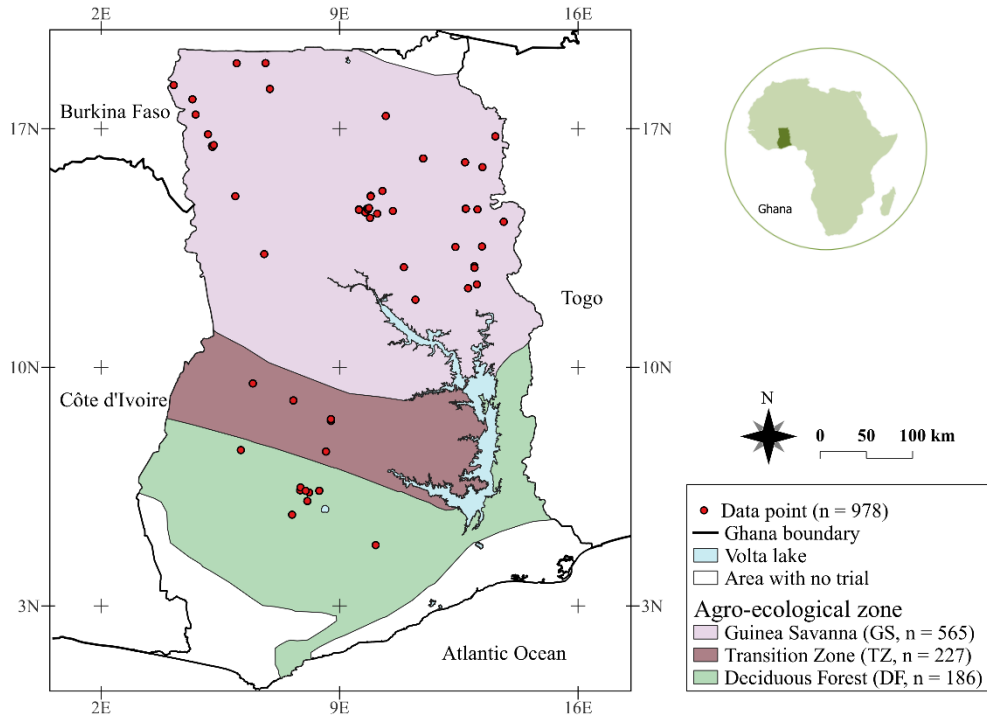


Figure 1: Geographic distribution of the experimental locations on a map of Ghana that shows the country's agro-ecological zones. Many experimental locations overlap.

Agroecological map source: [\(Antwi et al., 2014\)](#)

140 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
 141 and highest rates of P₂O₅ (PF) and K₂O (KF) were 3 kg ha⁻¹ and 90 kg ha⁻¹, with common rates of
 142 20 kg ha⁻¹, 40 kg ha⁻¹, and 60 kg ha⁻¹. To assess the fertilizer contribution to yield, data points were
 143 categorized into 2 treatments: "with_fertilizer" and "without_fertilizer."

144 2.4 Soil and environmental data

145 Table 1 presents the continuous variables used in the models. Soil chemical and physical
 146 properties data were obtained from the African SoilGrids ([ISRIC](#)), at 250 m of resolution for the 0
 147 - 30 cm topsoil. Rainfall and temperature datasets are from the WorldClim ([Fick et al., 2017](#))
 148 database, at a spatial resolution of 1 km² and land elevation data with a resolution of 30 meters
 149 from Shuttle Radar Topography Mission (SRTM) Digital Terrain Elevation.

150 The geographic coordinates of each data point were overlaid on the maps to extract soil and
 151 environmental data. Mean, minimum, and maximum temperatures and total rainfall for each
 152 data point were obtained from aggregated monthly measurements according to the

153 recommended rainy season for planting maize in Ghana, on the basis that maize was harvested
 154 no later than 120 days after the date of sowing ([Adu et al., 2014](#)).

155 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⁾	
Monthly mean temperature	°C	MMeT ⁽²⁾	(Fick et al., 2017)
Monthly min temperature	°C	MMiT ⁽²⁾	
Monthly max temperature	°C	MMaT ⁽²⁾	
Soil organic carbon	g kg ⁻¹	SOC	
Total nitrogen	g kg ⁻¹	TOTN	
Total phosphorus	mg kg ⁻¹	TOTP	
Extractable phosphorus	mg kg ⁻¹	P	
Extractable potassium	mmol kg ⁻¹	K	
Soil pH	-	pH	(Hengl et al., 2015)
Cation exchange capacity	mmol ₍₊₎ kg ⁻¹	CEC	
Sand content	%	SAND	
Clay content	%	CLAY	
Silt content	%	SILT	
Root zone depth	cm	RootDEP	

156 ⁽¹⁾ Precipitation is the sum for the estimated maize growing season, ⁽²⁾ Minimum, maximum, and mean temperatures are the mean for
 157 the estimated maize growing season.

158 2.5 QUEFTS model

159 QUEFTS was implemented in the R software, based on [Sattaria et al. \(2014\)](#). QUEFTS' calibration
 160 and validation parameters from [Wijayanto et al. \(2012\)](#) and [Antwi et al. \(2017\)](#) were used to
 161 replace the [Janssen et al. \(1990\)](#)' default parameterization. The maximum yield was set to 10 t ha⁻¹.
 162 Soil-available P in P-Mehlich3 and the exchangeable K in K-NH₄Ac extracted from the ISRIC
 163 maps were converted into P-Olsen and K-Mehlich3 extractable based on [Sawyer et al. \(1999\)](#)'
 164 pedo-transfer functions. The performance of the QUEFTS model was assessed using the
 165 coefficient of determination (R²) ([Krause et al., 2005](#)), through a linear regression between
 166 QUEFTS-estimated and observed yield, and the significance of correlation (r) was determined
 167 based on a p < 0.05.

168 2.6 Linear Mixed Effects Model

169 Prior to statistical modelling, data points were subjected to the Shapiro-Wilk normality test and
 170 were not found to be normally distributed (Supplementary Table S 2). Therefore, maize yield was
 171 transformed and normalized to meet the assumptions of homoscedasticity and homogeneity of

172 analysis of variance (ANOVA) and LMM errors variance (Supplementary Figure S 1). To evaluate
173 the effectiveness of the normalization technique applied, the “bestNormalize()” function of the
174 bestNormalize R package ([Peterson, 2021](#)) was used. Thus, the out-of-sample method via 10-fold
175 cross-validation with 3 repeats was performed to estimate the Pearson P-statistic (normality
176 statistic). The transformation technique was selected according to the calculated value of the
177 normality statistic.

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

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

192 The R function “TukeyHSD()” was used to perform the Tukey-Kramer test, and the
193 multcompView R package ([Graves et al., 2019](#)) was used to compact the letter display to indicate
194 significant differences between AEZ, Treatment, and Treatment*AEZ subgroups following
195 equation 2, where Y_{ijk} is the k th observed value of yield formed by the i th level of treatment effect,
196 the j th level of AEZ effect, the ij th level of interaction between treatment and AEZ, and the error
197 (ϵ_{ijk}).

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

198 The means of the maize yield were significantly different at $p < 0.05$. Tukey’s Honest
199 Significant Difference (TukeyHSD) multiple comparison analysis method tests were used because

200 there were unequal data point sizes among AEZ and Treatments subgroups ([Marusteri et al.,](#)
201 [2010](#); [McHugh, 2011](#)).

202 One-way ANOVA equations 3, 4 and 5 were used to measure the granularity of the random
203 effect (i.e., Intraclass Correlation Coefficient – ICC) by Treatment, AEZ, and Year on maize
204 yields, respectively. The ICC was calculated as the ratio of the variance between the random effect
205 and the model’s total variance ([Nakagawa et al., 2010](#)). The one-way ANOVA equations with
206 random effects were constructed as follows:

$$Y_{ij_Treat} = Treatment_i + \epsilon_{ij} \quad (3)$$

$$Y_{ij_AEZ} = AEZ_i + \epsilon_{ij} \quad (4)$$

$$Y_{ij_Year} = Year_i + \epsilon_{ij} \quad (5)$$

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

209 Since the objective was to determine if the variability in maize yield was related and could be
210 explained by the variables listed in Table 1 and maize variety type, a LMM was chosen. According
211 to [Dargie et al. \(2022\)](#), LMM accounts for sample size imbalance and confounding effects of
212 uncontrolled variables, as in our case. So, maize yield was modelled in three majors components:
213 a linear function with a fixed effect trend “ $f()_i$ ” with i th level of the explanatory variable
214 (environmental factors – ENV, soil physical properties – SPP, soil chemical properties – SCP,
215 maize variety type – VAR, fertilizer rate – FER), where the intercept was allowed to vary by
216 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},
217 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},
218 14_{SCP_AEZ}, 15_{VAR_AEZ}, and 16_{FER_AEZ}, with an error term (ϵ_{ijk} or ϵ_{ij}) denoting the small-scale
219 fluctuations around “ $f()_i$ ”. The LMMs 11_{step} and 17_{step_AEZ} were optimized by backward selection
220 of variables using the lmerTest R package ([Kuznetsova et al., 2017](#)) “step()” function. At the level
221 of each AEZ, LMM 17_{step_AEZ} was used to reveal the AEZ’s variables that were linked to maize
222 yield variability. As a result, $Y_{ij_AEZ_step}$ represented the maize yield modelled in a specific and
223 known AEZ (i.e., GS, TZ, or DF). However, stepwise LMM 17_{step_AEZ}, in DF, showed that the year’s
224 ICC was 0%. Therefore, LMM 17_{step_AEZ} was converted into a Multiple Linear Regression (MLR
225 _{step_DF}) formed with the trend function “ $f()_i$ ” and the error term (ϵ_{ij}). To avoid confusing models,
226 GS, TZ, or DF was assigned as a subscript to those LMMs or MLRs that referred to yield modelling

227 in a particular AEZ, and all LMM where the AEZs were not assigned as a subscript referred to
 228 the yield modelling across the entire set of data points. The LMMs were constructed as follows:

$$(LMM\ 6_{ENV}) \quad Y_{ijk_ENV} = f(ENV)_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (6)$$

$$(LMM\ 7_{SPP}) \quad Y_{ijk_SPP} = f(SPP)_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (7)$$

$$(LMM\ 8_{SCP}) \quad Y_{ijk_SCP} = f(SCP)_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (8)$$

$$(LMM\ 9_{VAR}) \quad Y_{ijk_VAR} = f(VAR)_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (9)$$

$$(LMM\ 10_{FER}) \quad Y_{ijk_FER} = f(FER)_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (10)$$

$$(LMM\ 11_{step}) \quad Y_{ijk_step} = f(ENV, SPP, SCP, VAR, FER)_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (11)$$

$$(LMM\ 12_{ENV_AEZ}) \quad Y_{ij_ENV_AEZ} = f(ENV)_i + Year_j + \epsilon_{ij} \quad (12)$$

$$(LMM\ 13_{SPP_AEZ}) \quad Y_{ij_SPP_AEZ} = f(SPP)_i + Year_j + \epsilon_{ij} \quad (13)$$

$$(LMM\ 14_{SCP_AEZ}) \quad Y_{ij_SCP_AEZ} = f(SCP)_i + Year_j + \epsilon_{ij} \quad (14)$$

$$(LMM\ 15_{VAR_AEZ}) \quad Y_{ij_VAR_AEZ} = f(VAR)_i + Year_j + \epsilon_{ij} \quad (15)$$

$$(LMM\ 16_{FER_AEZ}) \quad Y_{ij_FER_AEZ} = f(FER)_i + Year_j + \epsilon_{ij} \quad (16)$$

$$(LMM\ 17_{step_AEZ}) \quad Y_{ij_AEZ_step} = f(ENV, SPP, SCP, VAR, FER)_i + Year_j + \epsilon_{ij} \quad (17)$$

229 Parameter estimation for the variance components in the LMMs was done with the REstricted
 230 Maximum Likelihood (REML) approach ([Searle et al., 1992](#)) using the “lmer()” function in the
 231 lme4 R package ([Bates et al., 2015](#)). To assess the significance of LMM and MLR fixed effects and
 232 also to accommodate imbalances in data point sizes among some predictors, an ANOVA table
 233 with F-tests and p-values, using Kenward-Roger’s method for denominator degrees-of-freedom
 234 and F-statistic, was used ([Spilke et al., 2005](#)). R² statistics for mixed-effects model from [Nakagawa](#)
 235 [et al. \(2017\)](#)’ were used to assess the goodness of fit of the LMMs and MLR_{step_DF} fixed effect trend
 236 “f()”. The predictors groups, i.e., ENV, SPP, SCP, VAR, and FER, involved in the fixed-effect
 237 statistics modellings, except for LMMs 11_{step} and 17_{step_AEZ} and MLR_{step_DF}, were compared using
 238 the marginal R² (R²_m) to explain the maize yield variability related to the trend function (f) and
 239 also to rank the predictor groups according to their explanatory power.

240 2.7 Random Forest model

241 The RF regression model was trained to (i) identify, prioritize, and rank the variables in Table 1
 242 that were most important in explaining the variability of maize yields across AEZs, and (ii)
 243 predict maize yields within each AEZ based on the most important factors highlighted in (i). The
 244 measure of variable importance worked by calculating the increase in RF’s prediction error after
 245 permuting the variables in equation 18. RF is a non-parametric modelling machine learning
 246 technique that has been gaining popularity in agricultural data analysis; it is resistant to outliers

247 and can better handle both straightforward linear and complex nonlinear associations compared
248 to LMM/MLR ([Han et al., 2021](#)).

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

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

255 Variable importance ranking and yield prediction were done using the Ranger R package
256 ([Wright et al., 2017](#)). To prevent the model from being overfitted, the number of variables was
257 reduced for RF modelling (by reducing noise), thus a 10-fold cross-validation with 3 repeats was
258 used to enhance the effectiveness of the variable elimination strategy. This was implemented
259 using the “rfe()” and “rfeControl()” functions of the caret R package. The “rfe()” function applies
260 a backward selection process to find the optimal combination of variables that were most relevant
261 in predicting yield. The RF model fitted on the reduced variables was fine-tuned using a grid
262 search approach to select the best hyperparameters for the model, thus the “mtry” varied between
263 1 and 8, the “nodesize” ranged from 1 to 100 at an interval of 5, and the “ntree” ranged from 1 to
264 500 at an interval of 5.

265 RF model performance was evaluated using the coefficient of determination (R^2) ([Krause et al.,](#)
266 [2005](#)), accuracy using the root mean square error (RMSE) ([Krause et al., 2005](#)), and efficiency using
267 the Nash-Sutcliffe coefficient (NSE) ([Nash et al., 1970](#)). The linear regression between the
268 observed maize yield and that predicted by the RF was visualized using a 1:1 plot. All data
269 manipulation, model simulations, analysis, and visualization were performed in
270 RStudio®software ([RStudio Team, 2022](#)).

271 **3 RESULTS**

272 **3.1 Cropping system characteristics**

273 In soils across AEZs, sand predominated over silt and clay, which were present in roughly equal
274 amounts (Supplementary Table S 3). DF soils contained by far the greatest levels of soil organic

275 carbon (SOC), total phosphorus (TOTP), and cation exchange capacity (CEC) than TZ and GS
276 soils. Across the AEZs, potassium (K) and phosphorus (P) were generally present in comparable
277 amounts but at levels relatively below optimal values for P ([Kugbe et al., 2019](#); [Daniel et al., 2021](#))
278 and at a good level for K ([Antwi et al., 2016](#)). In addition, soil pH was around 6 and fluctuated
279 only very slightly from one AEZ to another ([Bationo et al., 2018](#)). Generally, the data points with
280 the deepest root zones (RootDEP) on average were observed in soils of the TZ, and the data points
281 with the shallowest soils were in the GS (Supplementary Table S 3), as also reported by [Bationo](#)
282 [et al. \(2018\)](#).

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

287 Across AEZs, the average rate of nitrogen fertilizer (NF), phosphorus fertilizer (PF), and
288 potassium fertilizer (KF) was 67 kg ha⁻¹, 30 kg ha⁻¹, and 29 kg ha⁻¹, respectively. The highest
289 average NF rate was in DF at 80 kg ha⁻¹ (Supplementary Table S 3). On the other hand, the average
290 PF (34 kg ha⁻¹) and KF (33 kg ha⁻¹) rates were higher in GS.

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

296 **3.2 Maize yield characteristics**

297 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,
298 similar to those reported by [SRID/MoFA \(2021\)](#). The main effect of treatments was significant
299 and large ($p < 0.05$) (Supplementary Table S 5). Tukey’s HSD test found that the mean yield was
300 significantly different between fertilized and control data points ($p < 0.05$) (Supplementary Table
301 S 5, Figure 2 A).

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

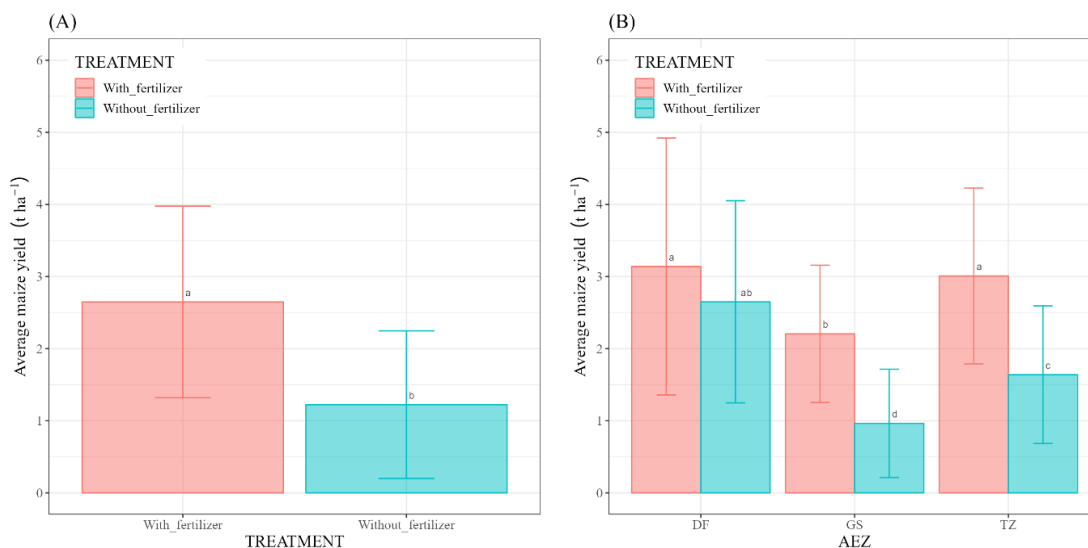


Figure 2: Average maize yield ($t\ ha^{-1}$) of treatments with fertilizer and without fertilizer (A) and the interaction between treatments and agro-ecological zone (AEZ) (B). DF (Deciduous Forest), GS (Guinea Savanna), and TZ (Transition Zone). Error bars are standard deviation bars. Different letters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

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

312 3.3 QUEFTS model yield estimated

313 When considering all the data points, the average estimated yield from the QUEFTS model was
 314 3.3 $t\ ha^{-1}$. A linear regression between observed and QUEFTS-estimated yields, based on the entire
 315 data points involved in the simulation (size effects of pH, SOC, P, TOTP, TOTN, K, NF, NP, and
 316 KF were confounded), revealed that soil chemical properties and fertilizer rate explained only
 317 19% ($r = 0.44$, $p < 0.05$) of total yield variability (Figure 3 A). Figure 3 shows the linear regression
 318 between observed and QUEFTS-estimated yields in GS, TZ, and DF (size effects of pH, SOC, P,

319 TOTP, TOTN, K, NF, NP, and KF were confounded in each AEZ). Indeed, there were weak
 320 correlations between observed and QUEFTS-estimated yields in all 3 AEZs (R^2_{DF} [4%] < R^2_{TZ}
 321 [19%] < R^2_{GS} [30%], Figure 3B, C, D). In DF, observed and QUEFTS-estimated yields were
 322 negatively correlated ($r = -0.21$, $p < 0.05$), but were positively correlated in GS ($r = 0.55$, $p < 0.05$)
 323 and TZ ($r = 0.43$, $p < 0.05$)

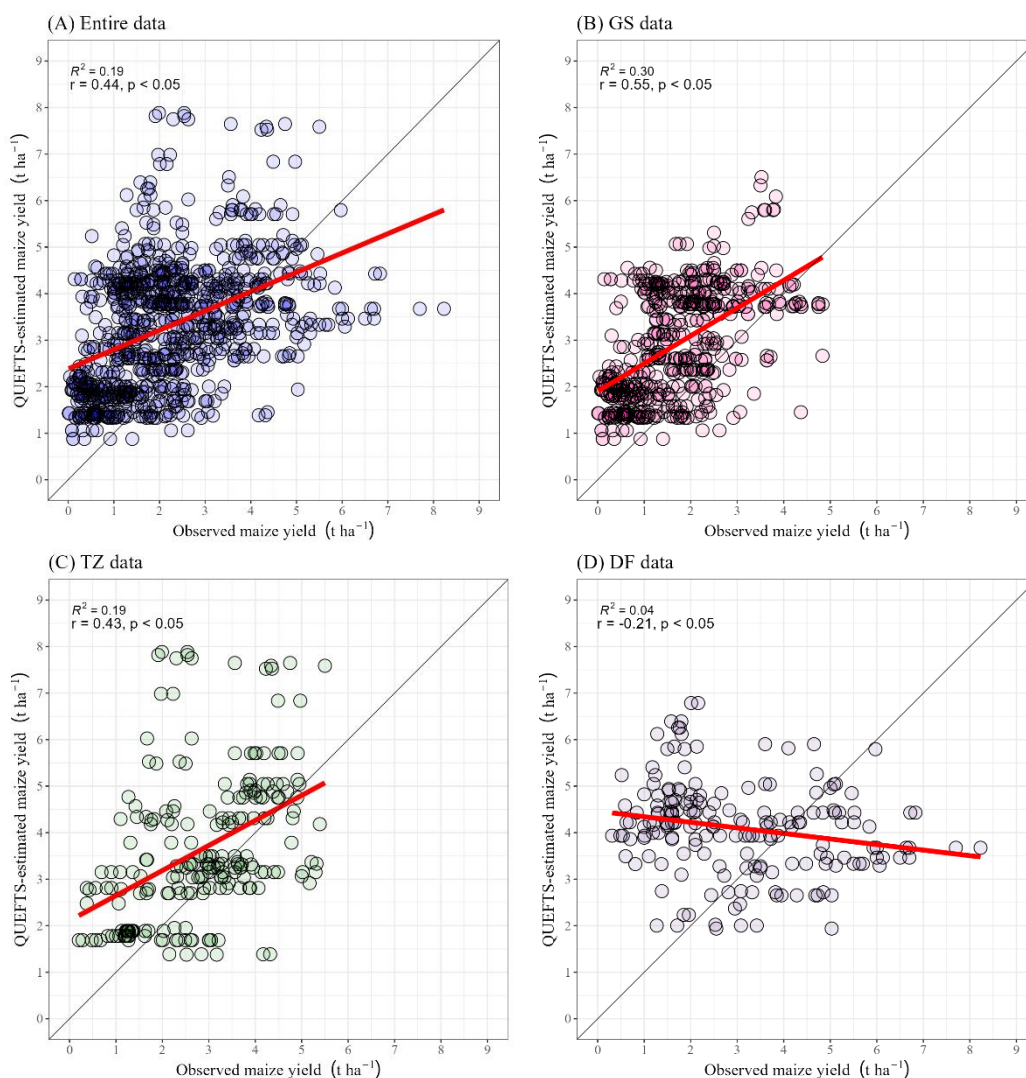


Figure 3: Relationship between observed and QUEFTS-estimated maize yield (t ha⁻¹) across all data points (A) and per AEZ (B, C, D). The R^2 indicates the portion of yield variability explained by soil chemical properties and fertilizers (confounded effect). The bold red line represents the linear regression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

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

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

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

335 3.4 Modelling maize yield variability across all data points

336 Variability in maize yield was greatly associated with soil chemical properties (SCP) across the 3
 337 AEZs. In the LMM, yield varied significantly ($p < 0.05$) with SOC, P, and K (Table 2Table 2,
 338 Supplementary Table S 9). A correlation matrix and a plot of yield on the CEC did not show a
 339 clear trend (Supplementary Figure S 3 a, d). In addition, pH was not a significant predictor of
 340 yield in LMM 11_{step} but was significantly ($p < 0.05$) related to yield in LMM 8_{SCP}, and a plot of
 341 yield on pH also revealed a clear and negative trend (Supplementary Figure S 3 a, c). Compared
 342 to P, K, pH, and CEC, the variation in yield due to SOC was greater (F value = 0.35) (Table 2).

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

348 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
----	------	---	-------	------	--------

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

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

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

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

369 **3.5 Maize yield determinants for each agro-ecological zone**

370 In GS, LMM demonstrated that maize yield significantly ($p < 0.05$) varied with ELV, CLAY, SOC,
 371 VARIETY, NF, PF, SAND, and CEC (Table 3Table 3, Supplementary Table S 11). However, the
 372 contributions of ELV and CEC to the variations in maize yield in GS did not show a clear and
 373 significant trend (Supplementary Figure S 4 a, b, e). The PF rate accounted for 48% of the R^2_m
 374 (0.39); therefore, GS was found to have a larger significant yield fluctuation due to PF rate than
 375 the other AEZs. Random effects (i.e., year) accounted for 2 – 32% of the total variation in maize
 376 yield in GS. The highest explanation power of yield variability ($R^2_m = 0.35$) in GS was shown by
 377 FER, followed by VAR ($R^2_m = 0.4$), SPP and SCP ($R^2_m = 0.03$), and ENV ($R^2_m = 0.02$).

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

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

Model	Variable	Sum Sq	DF	Den DF	F value	Pr(>F)	R^2_m	$\frac{ICC}{Year}$
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		

386 (†) Maize yield modelled with equation (17) in Guinea Savanna using stepwise linear mixed effects modelling.

387 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	$\frac{ICC}{Year}$
LMM 17 _{step_TZ} (†)	ELV	8.2	1	6.1	23.1	< 0.05	0.55	0.37
	CLAY	19.5	1	25.9	54.8	< 0.05		
	VARIETY	37.6	2	143.3	52.7	< 0.05		
	NF	44.9	1	215.4	125.9	< 0.05		
	PF	4.2	1	215.9	11.7	< 0.05		

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

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

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

Model	Variable	Sum Sq	DF	F value	Pr(>F)	R ²
MLR _{step_DF} ^(†)	ELV	1.6	1	7.4	< 0.05	0.79
	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	
	TOTP	2.9	1	13.8	< 0.05	
	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	

400 (†) Maize yield modelled in Deciduous Forest using stepwise multiple linear regression model.

401 LMM analysis showed that the contribution of maize variety type in the variations in yield in
 402 the 3 AEZs was also significant ($p < 0.05$) (Table 3, 4, 5, Supplementary Table S 14, S 15, S 16). A
 403 one-way ANOVA revealed that there was a significant ($p < 0.05$) difference in mean yield between
 404 at least two variety types in GS, TZ, and DF (Supplementary Table S 14, S 15, S 16). Tukey’s HSD
 405 test found that there was no significant difference in mean yield between OPV and hybrid in GS
 406 ($p = 0.25$) and TZ ($p = 0.69$), but there was significant difference in mean yield in DF ($p < 0.05$).
 407 There was a significant ($p < 0.05$) difference in mean yield between OPV and “Others” variety
 408 types in GS, TZ, and DF, and also between hybrid and “Others” variety types in the 3 AEZs.

409 3.6 Modelling maize yield using Random Forest

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

415 The R² ranged between 0.71 (training data) and 0.75 (testing data), the RMSE ranged between
 416 750 kg ha⁻¹ (training data) and 700 kg ha⁻¹ (testing data), and the NSE ranged between 0.58
 417 (training data) and 0.64 (testing data), indicating that the RF model performed well in yield
 418 prediction (Figure 4A, Supplementary Figure S 7 B). Based on the importance of the variables in

419 explaining yield using the training data points, maize yield prediction was made in each AEZ
 420 (Figure 4 Figure 4 B, C, D). In comparison to TZ and DF, the RF model's performance in GS was
 421 somewhat subpar. Indeed, in GS, RF-predicted maize yield showed the lowest values of NSE =
 422 0.55 and $R^2 = 0.74$.

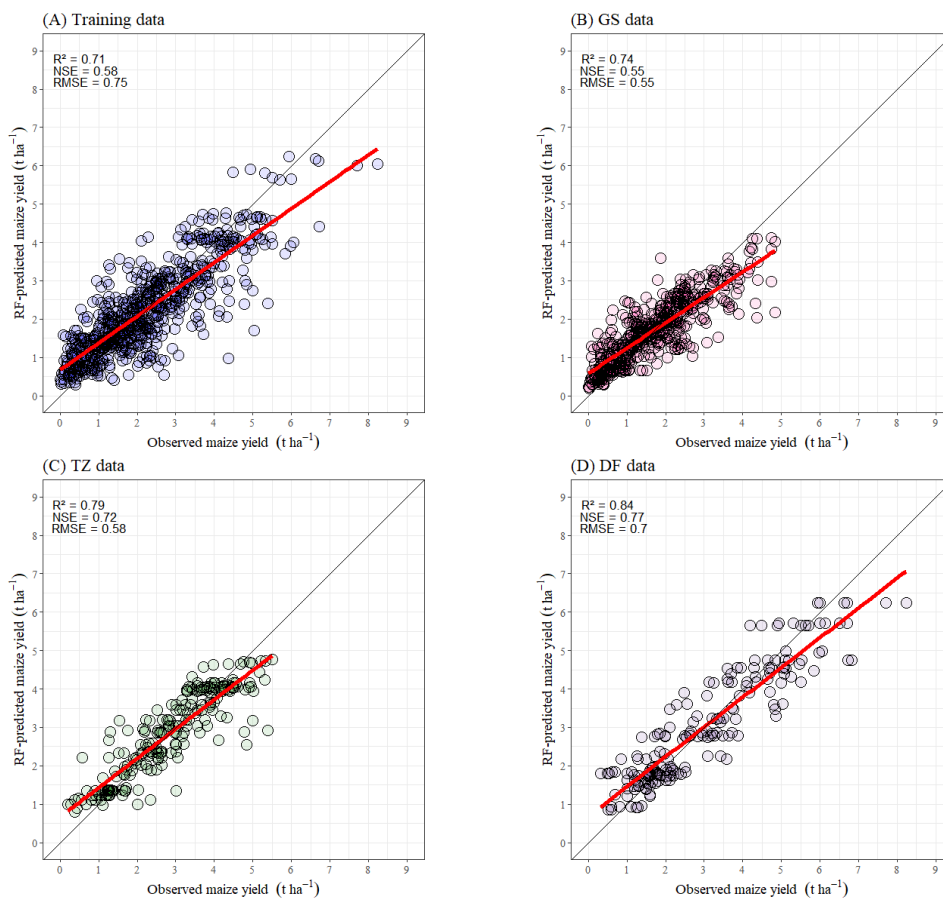


Figure 4: Relationship between observed maize yield and that predicted by the Random Forest (RF) model using 3 repeats of a 10-fold cross-validation approach. (A) The linear regression on 75% of the entire data point (training data point), (B) only the data point from GS, (C) only the data point from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

423 4 DISCUSSION

424 4.1 Soil chemical properties and fertilizer rates impacting yield variability

425 The QUEFTS model was calibrated and validated for the northern regions (i.e., the GS) and
 426 explained more variability there than in the other 2 AEZs. Yet, the low spatial and temporal
 427 explanatory power of QUEFTS suggests that other factors than soil chemical properties and

428 fertilizer rates only, contribute to the variability in maize yield as also reported by [Debtanu et al.](#)
429 [\(2006\)](#) and [Onduru et al. \(2007\)](#). Previous studies in Kenya, Benin, and Rwanda have also shown
430 low precision and accuracy of QUEFTS maize yield predictions ([Mulder, 2000](#); [Onduru et al.,](#)
431 [2007](#); [Breure et al., 2022](#)). While soil properties do not reveal strong temporal dynamics, rainfall,
432 and temperature in 2001 (MP = 502 - 698 mm, MMaT = 27 - 29°) and 2008 (MP = 709 mm, MMaT
433 = 30°) were favorable better mimicking the presumed growth conditions in QUEFTS when its
434 explanatory power was indeed highest ($R^2 > 50\%$). In addition, the importance of soil physical
435 properties, such as texture and soil depth, are also not considered in QUEFTS.

436 The LMM/MLR and RF models unveiled the importance of fertilizer rate (FER) and soil
437 chemical properties (SCP) in maize yield variability as well, as corroborated by [Braithmoh and Vlek](#)
438 [\(2006\)](#). Of the nine factors that were significantly ($p < 0.05$) associated with maize yield variability
439 in the LMM analysis (Table 2), 5 were related to the SCP. The NF rate was found to be the most
440 significant ($p < 0.05$) contributor to the variation in maize yield across the 3 AEZs in Ghana.
441 Extractable P and K demonstrated a positive trend in the main effect among these soil chemical
442 variables (Figure S 3 a), also identified by [Braithmoh et al. \(2006\)](#), [Yeboah et al. \(2016\)](#) and [Akolgo](#)
443 [et al. \(2020\)](#).

444 Ghanaian soils are generally not very acidic (Supplementary Table S 3) and fairly optimal,
445 however the slightly negative relation with maize yield likely results from the narrow pH range
446 and associations with other variables such as temperature (Supplementary Figure S 3 a, c, 5 a).

447 With tropical soils like in the DF having typical CEC values of around $10 \text{ cmol}_{(+) } \text{ kg}^{-1}$ ([Osei,](#)
448 [1995](#)), the high values ($83 \text{ cmol}_{(+) } \text{ kg}^{-1}$) of the 30 data points in the DF suggested an negative
449 though not significant ($p = 0.09$) relation with yield when the entire data point is considered
450 Supplementary Figure S 3 d). Excluding these data points turned the relation significantly ($p <$
451 0.05) positive (Supplementary Figure S 3 e).

452 The LMM analyses revealed that compared to NF and PF rates, the KF rate was not a strong
453 predictor of maize yield in any of the AEZs. LMM and RF modelling of yields with all data points
454 also confirmed this result, with KF explaining 2.5-3.5% of maize yield variability. The significant
455 ($p < 0.05$) association of K and maize yield in Table 2 was probably due to the data points of the
456 DF (Supplementary Table S 13, Supplementary Figure S 6 a). According to [Yawson et al. \(2011\)](#),
457 Ghana's forest soils will require frequent and split KF applications since they have a low capacity

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

463 We observed significant but downward linear trends in the main effect for SOC in 2 AEZs
464 (Supplementary Figure S 5 a, S 6 a). The bulk of high-yielding data points in TZ and DF were
465 found in soils with a comparatively lower SOC. This negative trend suggests yield variability to
466 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
468 yields, while highest yields were found on soils with a low SOC content. Across the entire set of
469 data points, with a wider SOC range, Supplementary Figure S 3 f revealed a significant ($p < 0.05$)
470 positive trend of maize yield on SOC. [Logah et al. \(2011\)](#) and many other scholars found the same
471 result ([Lal, 2006](#); [Solomon et al., 2016](#); [Owoade et al., 2020](#); [Owoade et al., 2021](#)). Therefore, SOC
472 is used as an indicator of soil fertility ([Tittonell et al., 2008](#)), with application of NF in such soil
473 promoting maize yield.

474 According to Table 3 and 5, in GS and DF respectively, PF had the highest (F-value = 60.31)
475 and lowest (F-value = 4.99) contributions to maize yield variability. This might explain why lower
476 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
477 P and TOTP in DF were greater than those in GS (Supplementary Table S 3), which may have
478 prevented the PF rate in DF soils from strongly contributing to maize yield.

479 **4.2 Role of soil physical characteristics on yield variability**

480 Only RootDEP demonstrated a significant ($p < 0.05$) positive trend with yield variation when the
481 entire set of maize yield data points was using in LMM. It was indeed identified by the RF model
482 as the third most important variable, right after the NF rate and MMeT (Supplementary Figure S
483 7 A). A more specific analysis showed that this influence originated mostly from the DF data
484 points (Supplementary Table S 13, Figure S 6 a) with deeper soils. [Guilpart et al. \(2017\)](#) revealed
485 this importance of RootDEP on the sensitivity of rainfed maize yields as function of the water-
486 holding capacity of the root zone.

487 Sand and clay were found to be significantly ($p < 0.05$) associated with maize yield variability
488 in both LMM and MLR modelling. While a negative relation of sand content is expected with
489 yield due to various associated soil properties like low water-holding capacity, the positive trend
490 (Supplementary Figure S 3 a, S 4 a, c, and S 6 a, g) could be related to the influence of organic
491 fertilizers (poultry and cow manure) in this study ([Dapaah et al., 2008](#); [Quansah, 2010](#); [Adjei-
492 Nsiah, 2012](#); [Kanton et al., 2016](#); [Badu et al., 2019](#)). According to [Obi et al. \(1995\)](#), [Zingore et al.
493 \(2007\)](#), [Uzoma et al. \(2011\)](#) and [Frimpong et al. \(2021\)](#), the application of organic fertilizers such
494 as cow and poultry manure on sandy soil significantly improves the physicochemical properties
495 of sandy soils favorably affecting maize growth. Moreover, the GS has a higher potential for
496 maize production ([Boullouz et al., 2022](#)) that can be better exploited with appropriate fertilization
497 and water management practices; organic manure being one of these favorable practices.

498 The contribution of clay content to the variations in maize yield was significant (Table 2). Clay
499 plays an important role in the supply, retention, and fixation of many macronutrients and
500 micronutrients in the soil that improve maize crop nutrition ([O'Halloran et al., 1985](#); [Batjes, 2011](#);
501 [Florence et al., 2017](#)), and ameliorates soil physical characteristics such as enhancing water
502 holding capacity. In GS, however, a sizable downward trend in the maize yield with clay was
503 observed (Supplementary Figure S 4 d). While the correlation coefficient was significant ($p <$
504 0.05), the distribution of data points in the Supplementary Figure S 4 d does clearly underline this
505 negative trend, likely for similar reasons as indicated for sand. [Sileshi et al. \(2022\)](#) reports similar
506 results, and [Njoroge et al. \(2018\)](#) reported no significant difference in clay content between sites
507 with high and low yield under NPK application in western Kenya.

508 **4.3 Contribution of environmental factors to yield variation**

509 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
510 a) and RF yield modelling (Supplementary Figure S 7 A) showed that environmental variables
511 (MMeT, MP, and ELV) were related to maize yield variability, as also reported by [Onduru et al.
512 \(2007\)](#), [Mugwe et al. \(2009\)](#), [Fosu-Mensah et al. \(2019\)](#), [Kyei-Mensah et al. \(2019\)](#), and [Cudjoe et
513 al. \(2021\)](#). However, MP of the entire data point reveals a negative trend with yield
514 (Supplementary Figure S 3 a, b). In the DF where very high yields of 7-8 t ha⁻¹ were reported, the
515 amount of rainfall was sufficient, though lower compared to other areas with lower yields. Their
516 larger water holder capacity of the deeper soils (RootDEP of 150 mm) made sufficient water
517 available for maize growth, as also reported by [Durodola et al. \(2020\)](#) and [Bagula et al. \(2022\)](#).

518 The negative trend between MP and RootDEP support this logic and why high MP in some
519 locations with low RootDEP did not result in high maize yields (Supplementary Figure S 6 a, h).

520 [Baffour-Ata et al. \(2021\)](#) reported that temperature was significantly and positively related to
521 maize yield in some regions of Ghana. This result was also observed in our analysis in TZ
522 (Supplementary Figure S 5 a, b). However, according to [Bationo et al. \(2018\)](#), high air
523 temperatures, high light levels, and heat-trapping sandy soils combine to make the local
524 environment too hot for good plant growth. Therefore, the positive relation found in TZ may
525 have resulted from improved high-potential, drought-tolerant variety and fertilizer rate ([Atiah et
526 al., 2021](#)). Drought-tolerant maize varieties in some trials, such as GH 110, Mamaba, and Akposoe,
527 could have performed relatively well at higher temperatures.

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

536 **4.4 Yield variability related to maize variety**

537 Overall, the results showed that maize variety type was significantly ($p < 0.05$) related to maize
538 yield variability, as found by [Kpotor et al. \(2014\)](#) also. [Bawa et al. \(2021\)](#) evaluated the influence
539 of different NF rates on maize yield and found differential varietal responses, as did [Adu et al.
540 \(2014\)](#) in Ghana, and [Abera et al. \(2017\)](#) in Ethiopia. [Afreh et al. \(2022\)](#) reported that, depending
541 on the interaction between maize variety and NF rate, farmers would need to apply their NF
542 during planting for Omankwa (OPV) and 14-28 days after planting for hybrid varieties to achieve
543 4.7 t ha⁻¹ and 6.5 t ha⁻¹, respectively. The coupled contribution of environmental factors and
544 fertilizer rates could explain the aberrant responses of varieties between the AEZ's. A study
545 conducted in DF and TZ by [Kpotor et al. \(2014\)](#) also revealed substantial interactions between
546 variety, NF rate, and trial site, which were very important in determining maize yield.

547 **4.5 Model performance**

548 Crop yield is intricately influenced by a variety of genetic, environmental, and management
549 factors as well as their interactions. For those involved in agriculture, the ability to precisely
550 predict crop yield in a variety of geographic settings with changing environmental circumstances
551 is becoming more and more crucial ([Wang, 2021](#)). However, the identifying of variables
552 influencing maize yield (in Ghana) is highly complex because yields are influenced by
553 interdependent and frequently drastically different climates, soil, and management variables (see
554 also [van Loon et al. \(2019\)](#)). Model cross-validation showed that RF performed well in maize yield
555 prediction across AEZs (Figure 4). The small value of RMSE observed in GS in yield prediction,
556 compared to TZ and DF, could be explained by the low mean yield observed (μ) and the standard
557 deviation (σ) less dispersed in GS ($\mu = 1.68 \text{ t ha}^{-1}$, $\sigma = 1$) than in TZ ($\mu = 2.82 \text{ t ha}^{-1}$, $\sigma = 1.3$) and
558 DF ($\mu = 3.05 \text{ t ha}^{-1}$, $\sigma = 2$).

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

572 **5 CONCLUSION**

573 Increasing maize yields to meet food demand while increasing farmers' profitability remains a
574 major challenge for Ghanaian farming systems. The overall objective of this study was to
575 characterize rainfed maize yield variation, understand the sources of variability, and predict
576 maize yield using robust statistical and modelling tools. The high variability in yield to fertilizer

577 application, both within and across AEZs and years, reflects a high degree of heterogeneity in soil
578 characteristics, environmental factors, maize varieties, and growing conditions at various spatial
579 and temporal scales. This study provided new information on maize yield variability in Ghana.
580 When the entire set of data points was involved in the yield modelling, LMM/MLR and RF
581 models showed that the NF rates were the most important factors explaining the maize yield
582 variability in Ghana. RF revealed also that the second most important factor was MMeT, followed
583 by RootDEP and MP. Since yield variability was significantly related to AEZ, in DF, soil chemical
584 properties and environmental factors explained most of variability. In TZ and GS, NF and PF
585 drove yield explanation. This suggest that the inherently high soil fertility in DF overrules the
586 importance of fertilizers, while fertilizers drive yield increase in the less-fertile TZ and GS. In all
587 3 AEZs, the used of different types of maize varieties also played an important role in the overall
588 yield variability observed. Importantly, our multi-model approach that combines advanced
589 statistical methods, crop-soil modelling, and machine learning, reveals its ability to identify
590 drivers for yield despite the huge complexity of the production system.

591 **Author contribution**

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

596 *Acknowledgement*

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

602 **Declaration of Competing Interest**

603 The authors declare that they have no conflict of interest.

604 **Data availability**

605 The authors do not have permission to share data.

606 References

- 607 Abera, T., Debele, T., & Wegary, D. (2017). Effects of Varieties and Nitrogen Fertilizer on Yield
608 and Yield Components of Maize on Farmers Field in Mid Altitude Areas of Western
609 Ethiopia. *International Journal of Agronomy*, 2017, 1-13.
610 <https://doi.org/10.1155/2017/4253917>
- 611 Adjei-Nsiah, S. (2012). Response of Maize (*Zea mays* L.) to Different Rates of Palm Bunch Ash
612 Application in the Semi-deciduous Forest Agro-ecological Zone of Ghana. *Applied and*
613 *Environmental Soil Science*, 2012, 870948. <https://doi.org/10.1155/2012/870948>
- 614 Adjei, V., & Kyerematen, R. (2018). Impacts of Changing Climate on Maize Production in the
615 Transitional Zone of Ghana. *American Journal of Climate Change*, 07(03), 463-476.
616 <https://doi.org/10.4236/ajcc.2018.73028>
- 617 Adu, G. B., Abdulai, M. S., Alidu, H., Nustugah, S. K., Buah, S. S., Kombiok, J. M., Obeng-Antwi,
618 K., Abudulai, M., & Etwire, P. M. (2014). Recommended Production Practices for Maize
619 in Ghana. <https://doi.org/10.13140/2.1.4376.3527>
- 620 Adzawla, W., Atakora, W. K., Gouzaye, A., & Bindraban, P. S. (2021). *Crop Yield and Fertilizer Use*
621 *Among Farmers in Guinea Savannah and Transitional Zones of Ghana*. [https://ifdc.org/wp-](https://ifdc.org/wp-content/uploads/2021/07/IFDC-FERARI-Research-Report-No-5-Final.pdf)
622 [content/uploads/2021/07/IFDC-FERARI-Research-Report-No-5-Final.pdf](https://ifdc.org/wp-content/uploads/2021/07/IFDC-FERARI-Research-Report-No-5-Final.pdf)
- 623 Adzawla, W., Atakora, W. K., Kissiedu, I. N., Martey, E., Etwire, P. M., Gouzaye, A., & Bindraban,
624 P. S. (2021). Characterization of farmers and the effect of fertilization on maize yields in
625 the Guinea Savannah, Sudan Savannah, and Transitional agroecological zones of Ghana.
626 *EFB Bioeconomy Journal*, 1. <https://doi.org/10.1016/j.bioeco.2021.100019>
- 627 Afreh, D. N., Afari, M. A. B., Adjei, R. R., Sarfo Boateng, A., Santo, K. G., Abdulai, M., & Popovic,
628 V. (2022). Response of Two Maize (*Zea mays* L.) Varieties to Times of NPK (15-15-15)
629 Fertilizer Application. *International Journal of Agronomy*, 2022, 1-7.
630 <https://doi.org/10.1155/2022/7186913>
- 631 Akolgo, G. A., Kemausuor, F., Awafo, E. A., Amankwah, E., Atta-Darkwa, T., Essandoh, E. O.,
632 Bart-Plange, A., & Maia, C. M. B. d. F. (2020). Biochar as a Soil Amendment Tool: Effects
633 on Soil Properties and Yield of Maize and Cabbage in Brong-Ahafo Region, Ghana. *Open*
634 *Journal of Soil Science*, 10(03), 91-108. <https://doi.org/10.4236/ojss.2020.103005>
- 635 Antwi, A., Duker, A., Fosu, M., & Abaidoo, R. C. (2017). Simulation of major soil nutrients
636 requirement for maize production using the QUEFTS model in the Northern region of
637 Ghana. *Direct Research Journal of Agriculture and Food Science*, 5(3), 133-140, Article
638 DRJA43071532. <http://directresearchpublisher.org/aboutjournal/drjafs>
- 639 Antwi, E. K., Boakye-Danquah, J., Asabere, S. B., Yiran, G. A. B., Loh, S. K., Awere, K. G., Abagale,
640 F. K., Asubonteng, K. O., Attua, E. M., & Owusu, A. B. (2014). Land use and landscape
641 structural changes in the ecoregions of Ghana. *Journal of Disaster Research*, 9(4), 452-467.
- 642 Antwi, M., Duker, A. A., Fosu, M., Abaidoo, R. C., & Pirasteh, S. (2016). Geospatial approach to
643 study the spatial distribution of major soil nutrients in the Northern region of Ghana.
644 *Cogent Geoscience*, 2(1), 1201906. <https://doi.org/10.1080/23312041.2016.1201906>
- 645 Atiah, W. A., Amekudzi, L. K., Akum, R. A., Quansah, E., Antwi- Agyei, P., & Danuor, S. K.
646 (2021). Climate variability and impacts on maize (*Zea mays*) yield in Ghana, West Africa.
647 *Quarterly Journal of the Royal Meteorological Society*. <https://doi.org/10.1002/qj.4199>
- 648 Badu, E., Kaba, J. S., Abunyewa, A. A., Dawoe, E. K., Agbenyega, O., & Barnes, R. V. (2019).
649 Biochar and inorganic nitrogen fertilizer effects on maize (*Zea mays* L.) nitrogen use and
650 yield in moist semi-deciduous forest zone of Ghana. *Journal of Plant Nutrition*, 42(19), 2407-
651 2422. <https://doi.org/10.1080/01904167.2019.1659347>

- 652 Baffour-Ata, F., Antwi-Agyei, P., Nkiaka, E., Dougill, A. J., Anning, A. K., & Kwakye, S. O. (2021).
653 Effect of climate variability on yields of selected staple food crops in northern Ghana.
654 *Journal of Agriculture and Food Research*, 6. <https://doi.org/10.1016/j.jafr.2021.100205>
- 655 Bagula, E. M., Majaliwa, J.-G. M., Basamba, T. A., Mondo, J.-G. M., Vanlauwe, B., Gabiri, G.,
656 Tumuhairwe, J.-B., Mushagalusa, G. N., Musinguzi, P., Akello, S., Egeru, A., & Tenywa,
657 M. M. (2022). Water Use Efficiency of Maize (*Zea mays* L.) Crop under Selected Soil and
658 Water Conservation Practices along the Slope Gradient in Ruzizi Watershed, Eastern D.R.
659 Congo. *Land*, 11(10). <https://doi.org/10.3390/land11101833>
- 660 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using
661 lme4. *Journal of Statistical Software*, 67(1), 1 - 48. <https://doi.org/10.18637/jss.v067.i01>
- 662 Bationo, A., Fening, J. O., & Kwaw, A. (2018). Assessment of Soil Fertility Status and Integrated
663 Soil Fertility Management in Ghana. In A. Bationo, et al. (Eds.), *Improving the Profitability,
664 Sustainability and Efficiency of Nutrients Through Site Specific Fertilizer Recommendations in
665 West Africa Agro-Ecosystems* (Vol. 1, pp. 93-138). Springer International Publishing AG
666 2018. https://doi.org/10.1007/978-3-319-58789-9_7
- 667 Batjes, N. H. (2011). *Global Distribution of Soil Phosphorus Retention Potential* (ISRIC Report 2011/06,
668 42 pp, Issue. <https://library.wur.nl/WebQuery/wurpubs/fulltext/185289>
- 669 Bawa, A., & Tang, Y. (2021). Yield and Growth Response of Maize (*Zea mays* L.) to Varietal and
670 Nitrogen Application in the Guinea Savanna Agro-Ecology of Ghana. *Advances in
671 Agriculture*, 2021, 1-8. <https://doi.org/10.1155/2021/1765251>
- 672 Bindraban, P. S., Stoorvogel, J. J., Jansen, D. M., Vlaming, J., & Groot, J. J. R. (2000). Land quality
673 indicators for sustainable land management: proposed method for yield gap and soil
674 nutrient balance. *Agriculture, Ecosystems and Environment*(81), 103-112.
675 [https://doi.org/10.1016/S0167-8809\(00\)00184-5](https://doi.org/10.1016/S0167-8809(00)00184-5)
- 676 Bonilla-Cedrez, C., Chamberlin, J., & Hijmans, R. J. (2021). Fertilizer and grain prices constrain
677 food production in sub-Saharan Africa. *Nature Food*, 2(10), 766-772.
678 <https://doi.org/10.1038/s43016-021-00370-1>
- 679 Boullouz, M., Bindraban, P. S., Kissiedu, I. N., Kouame, A. K. K., Devkota, K. P., & Atakora, W.
680 K. (2022). An integrative approach based on crop modeling and geospatial and statistical
681 analysis to quantify and explain the maize (*Zea mays*) yield gap in Ghana. *Front. Soil Sci.*
682 2:1037222. <https://doi.org/10.3389/fsoil.2022.1037222>
- 683 Braimoh, A. K., & Vlek, P. L. G. (2006). Soil quality and other factors influencing maize yield in
684 northern Ghana. *Soil use and management*, 22(2), 165-171. <https://doi.org/10.1111/j.1475-2743.2006.00032.x>
- 685
686 Breure, M. S., Kempen, B., & Hoffland, E. (2022). Spatial predictions of maize yields using
687 QUEFTS - A comparison of methods. *Geoderma*, 425.
688 <https://doi.org/10.1016/j.geoderma.2022.116018>
- 689 Bua, S., El-Mejahed, K., MacCarthy, D., Adogoba, D. S., Kissiedu, I. N., Atakora, W. K., Fosu, M.,
690 & P.S., B. (2020). *Yield responses of maize to fertilizers in Ghana* (IFDC FERARI n°2, Issue.
691 IFDC. <https://ifdc.org/wp-content/uploads/2020/09/FERARI-Policy-Brief-2-Yield-Responses-of-Maize-to-Fertilizers-in-Ghana.pdf>
- 692
693 Buah, S. S. J., Ibrahim, H., Derigubah, M., Kuzie, M., Segtaa, J. V., Bayala, J., Zougmore, R., &
694 Ouedraogo, M. (2017). Tillage and fertilizer effect on maize and soybean yields in the
695 Guinea savanna zone of Ghana. *Agriculture & Food Security*, 6(1), 1-11.
696 <https://doi.org/10.1186/s40066-017-0094-8>

- 697 Cicin-Sain, I. (2018). *Population growth and food: A systems perspective*. Retrieved 01 July 2021 from
698 <https://sustainablefoodtrust.org/articles/population-growth-and-food-a-systems->
699 [perspective/](https://sustainablefoodtrust.org/articles/population-growth-and-food-a-systems-perspective/)
- 700 Cooper, P. J. M. (1979). The association between altitude, environmental variables, maize growth
701 and yields in Kenya. *The Journal of Agricultural Science*, 93(3), 635-649. [https://doi.org/](https://doi.org/10.1017/S0021859600039058)
702 [10.1017/S0021859600039058](https://doi.org/10.1017/S0021859600039058).
- 703 Cudjoe, G. P., Antwi-Agyei, P., & Gyampoh, B. A. (2021). The Effect of Climate Variability on
704 Maize Production in the Ejura-Sekyedumase Municipality, Ghana. *Climate*, 9(10).
705 <https://doi.org/10.3390/cli9100145>
- 706 Daniel, K. A. N., Alice, A. A., & Thomas, A.-G. (2021). Response of maize (*Zea mays* L.) to foliar
707 and soil applied fertilizers in the Semi-deciduous forest zone of Ghana. *African Journal of*
708 *Agricultural Research*, 17(8), 1114-1122. <https://doi.org/10.5897/AJAR2021.15561>
- 709 Dapaah, H. K., Ennin, S. A., & Asafu-Agyei, J. N. (2008). Combining inorganic fertilizer with
710 poultry manure for sustainable production of quality protein maize in Ghana. *Ghana*
711 *Journal of Agricultural Science*, 41(1). <https://doi.org/10.4314/gjas.v41i1.46144>
- 712 Dargie, S., Girma, T., Chibsa, T., Kassa, S., Boke, S., Abera, A., Haileselassie, B., Addisie, S.,
713 Amsalu, S., Haileselassie, M., Soboka, S., Abera, W., & Weldesemayat, S. G. (2022).
714 Balanced fertilization increases wheat yield response on different soils and agroecological
715 zones in Ethiopia. *Experimental Agriculture*, 58.
716 <https://doi.org/10.1017/s0014479722000151>
- 717 Darko, D., Adjei, K. A., Odai, S. N., Obuobie, E., Asmah, R., & Trolle, D. (2019). Recent climate
718 trends for the Volta Basin in West Africa. *Weather*, 14(51).
719 <https://doi.org/10.1002/wea.3303>
- 720 Debtanu, M., Das, D. K., & Pathak, H. (2006). Simulation of fertilizer requirement for irrigated
721 wheat in eastern India using the QUEFTS model. *Archives of Agronomy and Soil Science*,
722 52(4), 403-418. <https://doi.org/10.1100/tsw.2006.43>
- 723 Diao, X., Hazell, P., Kolavalli, S., & Resnick, D. (2019). Ghana's economic and agricultural
724 transformation: Past performance and future prospects. 277.
725 <https://doi.org/10.1093/oso/9780198845348.001.0001>
- 726 Durodola, O. S., & Mourad, K. A. (2020). Modelling Maize Yield and Water Requirements under
727 Different Climate Change Scenarios. *Climate*, 8(11). <https://doi.org/10.3390/cli8110127>
- 728 Essel, B., Abaidoo, R. C., Opoku, A., & Ewusi-Mensah, N. (2020). Economically Optimal Rate for
729 Nutrient Application to Maize in the Semi-deciduous Forest Zone of Ghana. *J Soil Sci Plant*
730 *Nutr*, 20(4), 1703-1713. <https://doi.org/10.1007/s42729-020-00240-y>
- 731 Ezui, K. S., Franke, A. C., Ahiabor, B. D. K., Tetteh, F. M., Sogbedji, J., Janssen, B. H., Mando, A.,
732 & Giller, K. E. (2017). Understanding cassava yield response to soil and fertilizer nutrient
733 supply in West Africa. *Plant and Soil*, 420(1-2), 331-347. [https://doi.org/10.1007/s11104-](https://doi.org/10.1007/s11104-017-3387-6)
734 [017-3387-6](https://doi.org/10.1007/s11104-017-3387-6)
- 735 FAO/OECD. (2018). *Food Security and Nutrition: Challenges for Agriculture and the Hidden Potential*
736 *of Soil*. FAO. <http://www.fao.org/3/CA0917EN/ca0917en.pdf>
- 737 Farr, T. G., & Kobrick, M. (2000). Shuttle Radar Topography Mission produces a wealth of data.
738 *Eos Trans. AGU*, 81:583-583.
- 739 Fening, J. O., Yeboah, E., Gyapong, T. A., & Gaizie, E. (2009). On farm evaluation of the
740 contribution of three green manures to maize yield in the semi-deciduous forest zone of
741 Ghana. *African Journal of Environmental Science Technology*, 3(9).
742 <https://www.ajol.info/index.php/ajest/article/view/46071>

743 Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1- km spatial resolution climate surfaces for
744 global land areas. *International Journal of Climatology*, 37(12), 4302-4315.
745 <https://doi.org/10.1002/joc.5086>

746 Florence, A., Ransom, M., & Mengel, D. (2017). Potassium Fixation by Oxidized and Reduced
747 Forms of Phyllosilicates. *Soil Science Society of America Journal*, 81(5), 1247-1255.
748 <https://doi.org/10.2136/sssaj2016.12.0420>

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

753 Fox, J., & Weisberg, S. (2019). An R Companion to Applied Regression, Third edition. Sage,
754 Thousand Oaks CA. . <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.

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

758 Ghansah, B., Forkuo, E. K., Osei, E. F., Appoh, R. K., Asare, M. Y., & Kluste, N. A. B. (2018).
759 Mapping the spatial distribution of small reservoirs in the White Volta Sub-basin of
760 Ghana. *Remote Sensing Applications: Society and Environment*, 9, 107-115.
761 <https://doi.org/10.1016/j.rsase.2017.12.003>

762 Giller, K. E., Franke, A. C., Abaidoo, R., Baijukya, F., Bala, A., Boahen, S., Dashiell, K., Kantengwa,
763 S., Sanginga, J.-M., Sanginga, N., Simmons, A., & Turner, A. (2013). N2Africa: putting
764 nitrogen fixation to work for smallholder farmers in Africa. In *Agro-ecological intensification
765 of agricultural systems in the African highlands* (pp. 176-194). Routledge.
766 <https://doi.org/10.4324/9780203114742>

767 Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J.,
768 Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food security: the challenge of feeding
769 9 billion people. *Sci. (80-.)*. 327(5967), 812-818. <https://doi.org/10.1126/science.1185383>

770 Graves, S., Piepho, H.-P., & Selzer, L. (2019). *Visualizations of Paired Comparisons: Package
771 'multcompView'*. <https://cran.r-project.org/web/packages/multcompView/index.html>

772 Guilpart, N., Grassini, P., van Wart, J., Yang, H., van Ittersum, M. K., van Bussel, L. G. J., Wolf, J.,
773 Claessens, L., Leenaars, J. G. B., & Cassman, K. G. (2017). Rooting for food security in Sub-
774 Saharan Africa. *Environmental Research Letters*, 12(11), 1-8. <https://doi.org/10.1088/1748-9326/aa9003>

775

776 GYGA. (2021). *Agriculture in Ghana*. University of Nebraska-Lincoln and Wageningen University
777 & Research. Retrieved 24th March 2021 from <https://www.yieldgap.org/ghana>

778 Han, S., & Kim, H. (2021). Optimal Feature Set Size in Random Forest Regression. *Applied Sciences*,
779 11(8). <https://doi.org/10.3390/app11083428>

780 Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., Sila, A.,
781 MacMillan, R. A., Mendes de Jesus, J., Tamene, L., & Tondoh, J. E. (2015). Mapping Soil
782 Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current
783 Predictions. *PLoS One*, 10(6), 1-26. <https://doi.org/10.1371/journal.pone.0125814>

784 IFDC. (2012). *Ghana Fertilizer Assessment*. International Fertilizer Development Center Muscle
785 Shoals Alabama 35662 USA.
786 [http://ghana.countrystat.org/fileadmin/user_upload/countrystat_fenix/congo/docs/
787 Ghana%20Fertilizer%20Need%20Assessment%20for%20USAID%20FtF%20AFAPIFDC%
788 202012.pdf](http://ghana.countrystat.org/fileadmin/user_upload/countrystat_fenix/congo/docs/Ghana%20Fertilizer%20Need%20Assessment%20for%20USAID%20FtF%20AFAPIFDC%202012.pdf)

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

791 Janssen, B. H., Guiking, F. C. T., van der Eijk, D., Smaling, E. M. A., Wolf, J., & van Reuler, H.
792 (1990). A system for quantitative evaluation of the fertility of tropical soils (QUEFTS).
793 *Geoderma*, 46, 299-318. [https://doi.org/10.1016/0016-7061\(90\)90021-Z](https://doi.org/10.1016/0016-7061(90)90021-Z)

794 Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim,
795 K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random Forests for Global and
796 Regional Crop Yield Predictions. *PLoS One*, 11(6), e0156571.
797 <https://doi.org/10.1371/journal.pone.0156571>

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

801 Kanton, R. A. L., Prasad, P. V. V., Mohammed, A. M., Bidzakin, J. K., Ansoba, E. Y., Asungre, P.
802 A., Lamini, S., Mahama, G., Kusi, F., & Sugri, I. (2016). Organic and Inorganic Fertilizer
803 Effects on the Growth and Yield of Maize in a Dry Agro-Ecology in Northern Ghana.
804 *Journal of Crop Improvement*, 30(1), 1-16. <https://doi.org/10.1080/15427528.2015.1085939>

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

809 Kpotor, P., Akromah, R., Ewool, M. B., Kena, A. W., Owusu-Adjei, E., & Tuffour, H. O. (2014).
810 Assessment of the Relative Yielding Abilities and Stability of Maize (*Zea mays* L.)
811 Genotypes under Different Levels of Nitrogen Fertilization across Two Agro-Ecological
812 Zones in Ghana. *International Journal of Scientific Research in Agricultural Sciences*, 1(7), 128-
813 141. <https://doi.org/10.12983/ijrsas-2014-p0128-0141>

814 Kranjac-Berisavljevic, G., Bayorbor, T. B., Abdulai, A. S., Obeng, F., Blench, R. M., Turton, C. N.,
815 Boyd, C., & Drake, E. (1999). Rethinking natural resource degradation in semi-arid Sub-
816 Saharan Africa: the case of semi-arid Ghana. *University for Development Studies, Tamale*
817 *ODI, London*.

818 Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for
819 hydrological model assessment. *Adv. Geosci.*(5), 89-97. [https://doi.org/10.5194/adgeo-5-](https://doi.org/10.5194/adgeo-5-89-2005)
820 [89-2005](https://doi.org/10.5194/adgeo-5-89-2005)

821 Kravchenko, A. N., & Robertson, G. P. (2007). Can Topographical and Yield Data Substantially
822 Improve Total Soil Carbon Mapping by Regression Kriging ? *Agronomy Journal*(99), 12-17.
823 <https://doi.org/10.2134/agronj2005.0251>

824 Kugbe, J. X., Kombat, R., Atakora, W., & Tejada Moral, M. (2019). Secondary and micronutrient
825 inclusion in fertilizer formulation impact on maize growth and yield across northern
826 Ghana. *Cogent Food & Agriculture*, 5(1). <https://doi.org/10.1080/23311932.2019.1700030>

827 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear
828 Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-23.
829 <https://doi.org/10.18637/jss.v082.i13>

830 Kyei-Mensah, C., Kyerematen, R., & Adu-Acheampong, S. (2019). Impact of Rainfall Variability
831 on Crop Production within the Worobong Ecological Area of Fanteakwa District, Ghana.
832 *Advances in Agriculture*, 2019, 1-7. <https://doi.org/10.1155/2019/7930127>

833 Lal, R. (2006). Enhancing crop yields in the developing countries through restoration of the soil
834 organic carbon pool in agricultural lands. *Land degradation development*, 17(2), 197-209.
835 <https://doi.org/10.1002/ldr.696>

- 836 Lamos-Díaz, H., Puentes-Garzón, D. E., & Zarate-Caicedo, D. A. (2020). Comparison Between
837 Machine Learning Models for Yield Forecast in Cocoa Crops in Santander, Colombia.
838 *Revista Facultad de Ingeniería*, 29(54), e10853-e10853.
839 <https://doi.org/10.19053/01211129.v29.n54.2020.10853>
- 840 Leenaars, J. G. B., Claessens, L., Heuvelink, G. B. M., Hengl, T., Ruiperez Gonzalez, M., van
841 Bussel, L. G. J., Guilpart, N., Yang, H., & Cassman, K. G. (2018). Mapping rootable depth
842 and root zone plant-available water holding capacity of the soil of sub-Saharan Africa.
843 *Geoderma*, 324, 18-36. <https://doi.org/10.1016/j.geoderma.2018.02.046>
- 844 Logah, V., Ewusi-Mensah, N., & Tetteh, F. K. M. (2011). Soil Organic Carbon and Crop Yield
845 under Different Soil Amendments and Cropping Systems in the Semi-deciduous Forest
846 Zone of Ghana. *Journal of Plant Sciences*, 6(4), 165-173.
847 <https://doi.org/10.3923/jps.2011.165.173>
- 848 MacCarthy, D. S., Vlek, P. L. G., & Fosu-Mensah, B. Y. (2012). The Response of Maize to N
849 Fertilization in a Sub-Humid Region of Ghana: Understanding the Process Using a Crop
850 Simulation Model. In J. Kihara, et al. (Eds.), *Improving Soil Fertility Recommendations in
851 Africa using the Decision Support System for Agrotechnology Transfer (DSSAT)*. Springer
852 Science + Business Media. https://doi.org/10.1007/978-94-007-2960-5_5
- 853 Marusteri, M., & Bacarea, V. (2010). Comparing groups for statistical differences: how to choose
854 the right statistical test? *Biochemia medica*, 20(1), 15-32. <https://hrcak.srce.hr/file/73801>
- 855 McHugh, M. L. (2011). Multiple comparison analysis testing in ANOVA. *Biochemia medica*, 21(3),
856 203-209. <https://www.biochemia-medica.com/en/journal/21/3/10.11613/BM.2011.029>
- 857 Mobilian, C., & Craft, C. B. (2021). Wetland Soils: Physical and Chemical Properties and
858 Biogeochemical Processes. *Reference Module in Earth Systems and Environmental Sciences*,
859 Elsevier. <https://doi.org/10.1016/B978-0-12-819166-8.00049-9>
- 860 Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., & Foley, J. A. (2012).
861 Closing yield gaps through nutrient and water management. *Nature*, 490(7419), 254-257.
862 <https://doi.org/10.1038/nature11420>
- 863 Mugwe, J., Mugendi, D., Kungu, J., & Muna, M. M. (2009). Maize Yields Response to Application
864 of Organic and Inorganic Input under on-Station and on-Farm Experiments in Central
865 Kenya. *Experimental Agriculture*, 45(1), 47-59.
866 <https://doi.org/10.1017/S0014479708007084>
- 867 Mulder, I. (2000). *Soil Fertility: QUEFTS and Farmer's Perceptions* (Vol. Working Paper No 30). IIED.
868 https://www.researchgate.net/publication/228856865_Soil_fertility_QUEFTS_and_farmers%27_perceptions
- 870 Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and
871 intra-class correlation coefficient from generalized linear mixed-effects models revisited
872 and expanded. *Journal of the Royal Society Interface*, 14(134), 20170213.
873 <https://doi.org/10.1098/rsif.2017.0213>
- 874 Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non- Gaussian data: a
875 practical guide for biologists. *Biological Reviews*, 85(4), 935-956.
876 <https://doi.org/10.1111/j.1469-185x.2010.00141.x>
- 877 Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I – A
878 discussion of principles. *Journal of hydrology*, 10(3), 282-290.
879 https://en.wikipedia.org/wiki/Nash-Sutcliffe_model_efficiency_coefficient
- 880 Nevavuori, P., Narra, N., Linna, P., & Lipping, T. (2020). Crop Yield Prediction Using
881 Multitemporal UAV Data and Spatio-Temporal Deep Learning Models. *Remote Sensing*,
882 12(23), 4000. <https://www.mdpi.com/2072-4292/12/23/4000>

- 883 Njoroge, K. S. (2019). *Explaining variability in maize yield responses to nutrient applications in*
884 *smallholder farms of western Kenya* (Publication Number 978-94-6395-164-7) Wageningen
885 University & Research]. The Netherlands. <https://doi.org/10.18174/503185>
- 886 Njoroge, R., Otinga, A. N., Okalebo, J. R., Pepela, M., & Merckx, R. (2018). Maize (*Zea mays* L.)
887 Response to Secondary and Micronutrients for Profitable N, P and K Fertilizer Use in
888 Poorly Responsive Soils. *Agronomy*, 8(4). <https://doi.org/10.3390/agronomy8040049>
- 889 Nkrumah, F., Klutse, N. A. B., Adukpo, D. C., Owusu, K., & Quagraine, K. A. (2014). Rainfall
890 variability over Ghana: model versus rain gauge observation. *International Journal of*
891 *Geosciences*, 5(7). <https://doi.org/10.4236/ijg.2014.57060>
- 892 Nyuor, B. A., Donkor, E., Aidoo, R., Buah, S. S., Naab, J., Nutsugah, S., Bayala, J., & Zougmore, R.
893 (2016). Economic Impacts of Climate Change on Cereal Production: Implications for
894 Sustainable Agriculture in Northern Ghana. *Sustainability*, 8(8).
895 <https://doi.org/10.3390/su8080724>
- 896 O'Halloran, I. P., Kachanoski, R. G., & Stewart, J. W. B. (1985). Spatial variability of soil
897 phosphorus as influenced by soil texture and management. *Canadian journal of soil science*,
898 65(3), 475-487. <https://doi.org/10.4141/cjss85-051>
- 899 Obi, M. E., & Ebo, P. O. (1995). The effects of organic and inorganic amendments on soil physical
900 properties and maize production in a severely degraded sandy soil in southern Nigeria.
901 *Bioresource Technology*, 51(2-3), 117-123. [https://doi.org/10.1016/0960-8524\(94\)00103-8](https://doi.org/10.1016/0960-8524(94)00103-8)
- 902 Onduru, D. D., & Du Preez, C. C. (2007). Spatial and temporal aspects of agricultural
903 sustainability in the semi-arid tropics: a case study in Mbeere district, Eastern Kenya.
904 *Tropical Science*, 47(3), 134-148. <https://doi.org/10.1002/ts.207>
- 905 Osei, B. A. (1995). Effects of different lime application rates and time on some chemical properties
906 of an acid soil in Ghana. *Soil use and management*, 11, 25-29.
907 <https://doi.org/10.1111/j.1475-2743.1995.tb00491.x>
- 908 Ovalles, F. A., & Collins, M. E. (1986). Soil- landscape relationships and soil variability in north
909 central Florida. *Soil Science Society of America Journal*, 50(2), 401-408.
910 <https://doi.org/10.2136/sssaj1986.03615995005000020029x>
- 911 Owoade, F. M., Adiku, S. G. K., Atkinson, C. J., & MacCarthy, D. S. (2021). Differential Impact of
912 Land Use Types on Soil Productivity Components in Two Agro-ecological Zones of
913 Southern Ghana. In *African Handbook of Climate Change Adaptation* (pp. 1721-1733).
914 https://doi.org/10.1007/978-3-030-45106-6_144
- 915 Owoade, F. M., Adiku, S. G. K., Atkinson, C. J., MacCarthy, D. S., Kumahor, S. K., & Kolawole,
916 G. O. (2020). Location and Land use effects on Soil Carbon Accretion and Productivity in
917 the Coastal Savanna Agro-ecological Zone of Ghana. *West African Journal of Applied*
918 *Ecology*, 28(2), 1-13.
919 <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUK>
920 [EwijkM_d2cD5AhVNLewKHfCfCYgQFnoECAIQAQ&url=https%3A%2F%2Fwww.ajol.](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUK)
921 [info%2Findex.php%2Fwajae%2Farticle%2Fview%2F202673%2F191158&usg=AOvVaw1](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUK)
922 [WdXQq6X2x3K-dU3roFeYD](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUK)
- 923 Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N.
924 (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, 187.
925 <https://doi.org/10.1016/j.agry.2020.103016>
- 926 Peprah, K. (2012). Rainfall and temperature correlation with crop yield: The case of Asunafo
927 forest, Ghana. www.ijsr.net
- 928 Peterson, R. A. (2021). Finding Optimal Normalizing Transformations via bestNormalize. *The R*
929 *Journal*, 13(1), 310-329. <https://doi.org/10.32614/RJ-2021-041>

930 Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for
931 random forest. *WIREs Data Mining and Knowledge Discovery*, 9(3).
932 <https://doi.org/10.1002/widm.1301>

933 Quansah, G. W. (2010). *Effect of organic and inorganic fertilizers and their combinations on the growth
934 and yield of maize in the semi-deciduous forest zone of Ghana* [Msc, Kwame Nkrouma
935 University of Science and Technology]. Kumsi, Ghana.
936 <http://csirspace.csirgh.com/handle/123456789/2379>

937 Rahman, A. N., Larbi, A., Kotu, B., Asante, M. O., Akakpo, D. B., Mellon- Bedi, S., & Hoeschle-
938 Zeledon, I. (2021). Maize-legume strip cropping effect on productivity, income, and
939 income risk of farmers in northern Ghana. *Agronomy Journal*, 1-12.
940 <https://doi.org/10.1002/agj2.20536>

941 Ren, T., Zou, J., Wang, Y., Li, X. K., Cong, R. H., & Lu, J. W. (2015). Estimating nutrient
942 requirements for winter oilseed rape based on QUEFTS analysis. *Journal of Agricultural
943 Science*, 154(3), 425-437. <https://doi.org/10.1017/S0021859615000301>

944 RStudio Team. (2022). RStudio: Integrated Development for R. . *RStudio (2021.09.2)*, PBC, Boston,
945 MA URL <http://www.rstudio.com/>.

946 Sadras, V. O., & Calvino, P. A. (2001). Quantification of grain yield response to soil depth in
947 soybean, maize, sunflower, and wheat. *Agronomy Journal*, 93(3), 577-583.
948 <https://doi.org/10.2134/agronj2001.933577x>

949 Sattaria, S. Z., van Ittersuma, M. K., Bouwmanb, A. F., Smitd, A. L., & Janssen, B. H. (2014). Crop
950 yield response to soil fertility and N, P, K inputs in different environments: Testing and
951 improving the QUEFTS model. *Field Crops Research*, 157, 35-46.
952 <https://doi.org/10.1016/j.fcr.2013.12.005>

953 Sawyer, J. E., & Mallarino, A. P. (1999). Differentiating and understanding the Mehlich 3, Bray,
954 and Olsen soil phosphorus tests. *Agronomy Conference Proceedings and Presentations*, 12.
955 <https://dr.lib.iastate.edu/handle/20.500.12876/4349>

956 Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. John Wiley & Sons.
957 <https://doi.org/10.1002/9780470316856>

958 Shehu, B. M., Lawan, B. A., Jibrin, J. M., Kamara, A. Y., Mohammed, I. B., Rurinda, J., Zingore, S.,
959 Craufurd, P., Vanlauwe, B., Adam, A. M., & Merckx, R. (2019). Balanced nutrient
960 requirements for maize in the Northern Nigerian Savanna: Parameterization and
961 validation of QUEFTS model. *Field Crops Res*, 241, 107585.
962 <https://doi.org/10.1016/j.fcr.2019.107585>

963 Sileshi, G. W., Kihara, J., Tamene, L., Vanlauwe, B., Phiri, E., & Jama, B. (2022). Unravelling causes
964 of poor crop response to applied N and P fertilizers on African soils. *Experimental
965 Agriculture*, 1-17. <https://doi.org/10.1017/S0014479721000247>

966 Solomon, D., Lehmann, J., Fraser, J. A., Leach, M., Amanor, K., Frausin, V., Kristiansen, S. M.,
967 Millimouno, D., & Fairhead, J. (2016). Indigenous African soil enrichment as a climate-
968 smart sustainable agriculture alternative. *Frontiers in Ecology the Environment*, 14(2), 71-76.
969 <https://doi.org/10.1002/fee.1226>

970 Spilke, J., Piepho, H. P., & Hu, X. (2005). Analysis of unbalanced data by mixed linear models
971 using the MIXED procedure of the SAS system. *Journal of Agronomy and Crop Science*,
972 191(1), 47-54. <https://doi.org/10.1111/j.1439-037X.2004.00120.x>

973 SRID/MoFA. (2021). *Agriculture in Ghana: Facts and Figures (2020)*.
974 https://srid.mofa.gov.gh/sites/default/files/Agriculture%20In%20Ghana%20Facts%20%26%20Figures_%202020%20FINAL.pdf

- 976 Tabi, F. O., Diels, J., Ogunkunle, A. O., Iwuafor, E. N. O., Vanlauwe, B., & Sanginga, N. (2007).
977 Potential nutrient supply, nutrient utilization efficiencies, fertilizer recovery rates and
978 maize yield in northern Nigeria. *Nutrient Cycling in Agroecosystems*, 80(2), 161-172.
979 <https://doi.org/10.1007/s10705-007-9129-z>
- 980 Tetteh, F., Larbi, A., Nketia, K. A., Senayah, J. K., Hoeschle-Zeledon, I., & Abdulrahman, N. (2016).
981 *Suitability of soils for cereal cropping in Northern Ghana*. I. I. o. T. Agriculture.
982 [https://www.researchgate.net/publication/335581532_Suitability_of_soils_for_cereal_c](https://www.researchgate.net/publication/335581532_Suitability_of_soils_for_cereal_cropping_in_northern_Ghana/link/5d6e6bbf299bf16522f2b9b6/download)
983 [ropping_in_northern_Ghana/link/5d6e6bbf299bf16522f2b9b6/download](https://www.researchgate.net/publication/335581532_Suitability_of_soils_for_cereal_cropping_in_northern_Ghana/link/5d6e6bbf299bf16522f2b9b6/download)
- 984 Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable
985 intensification of agriculture. *Proc Natl Acad Sci U S A*, 108(50), 20260-20264.
986 <https://doi.org/10.1073/pnas.1116437108>
- 987 Timsina, J., Dutta, S., Devkota, K. P., Chakraborty, S., Neupane, R. K., Bishta, S., Amgain, L. P.,
988 Singh, V. K., Islam, S., & Majumdar, K. (2021). Improved nutrient management in cereals
989 using Nutrient Expert and machine learning tools: Productivity, profitability and nutrient
990 use efficiency. *Agricultural Systems*, 192. <https://doi.org/10.1016/j.agsy.2021.103181>
- 991 Tiftonell, P., Corbeels, M., van Wijk, M. T., Vanlauwe, B., & Giller, K. E. (2008). Combining
992 Organic and Mineral Fertilizers for Integrated Soil Fertility Management in Smallholder
993 Farming Systems of Kenya: Explorations Using the Crop-Soil Model FIELD. *Agronomy*
994 *Journal*, 100(5), 1511-1526. <https://doi.org/10.2134/agronj2007.0355>
- 995 USAID. (2022). *Agriculture and food security*. USAID. Retrieved 18/02/2022 from
996 <https://www.usaid.gov/ghana/agriculture-and-food-security>
- 997 Uzoma, K. C., Inoue, M., Andry, H., Fujimaki, H., Zahoor, A., & Nishihara, E. (2011). Effect of
998 cow manure biochar on maize productivity under sandy soil condition. *Soil use and*
999 *management*, 27(2), 205-212. <https://doi.org/10.1111/j.1475-2743.2011.00340.x>
- 1000 van Loon, M. P., Adjei-Nsiah, S., Descheemaeker, K., Akotsen-Mensah, C., van Dijk, M., Morley,
1001 T., van Ittersum, M. K., & Reidsma, P. (2019). Can yield variability be explained?
1002 Integrated assessment of maize yield gaps across smallholders in Ghana. *Field Crops*
1003 *Research*, 236, 132-144. <https://doi.org/10.1016/j.fcr.2019.03.022>
- 1004 Wallach, D., Makowski, D., Jones, J. W., & Brun, F. (2018). *Working with dynamic crop models*.
1005 Academic Press-Elsevier. <https://doi.org/10.1016/C2016-0-01552-8>
- 1006 Wang, L. (2021). Data Driven Explanation of Temporal and Spatial Variability of Maize Yield in
1007 the United States. *Front Plant Sci*, 12, 701192. <https://doi.org/10.3389/fpls.2021.701192>
- 1008 Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
1009 Retrieved from <https://ggplot2.tidyverse.org>.
- 1010 Wijayanto, Y., & Prastyanto, E. (2012). A study of using QUEFTS model for establishing site
1011 specific fertilizer recommendation in maize on the basis of farmer fields. *AGRIVITA*,
1012 *Journal of Agricultural Science*, 33(3), 273-278.
1013 [https://media.neliti.com/media/publications/53171-EN-a-study-of-using-quefts-](https://media.neliti.com/media/publications/53171-EN-a-study-of-using-quefts-model-for-establ.pdf)
1014 [model-for-establ.pdf](https://media.neliti.com/media/publications/53171-EN-a-study-of-using-quefts-model-for-establ.pdf)
- 1015 Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high
1016 dimensional data in C++ and R. *Journal of Statistical Software*, 77(1), 1-17.
1017 <https://doi.org/10.18637/jss.v077.i01>
- 1018 Xu, X., He, P., Pampolino, M. F., Chuan, L., Johnston, A. M., Qiu, S., Zhao, S., & Zhou, W. (2013).
1019 Nutrient requirements for maize in China based on QUEFTS analysis. *Field Crops Research*,
1020 150, 115-125. <https://doi.org/10.1016/j.fcr.2013.06.006>

- 1021 Yawson, D. O., Kwakye, P. K., Armah, F. A., & Frimpong, K. A. (2011). The dynamics of
1022 potassium (K) in representative soil series of Ghana. *ARPN Journal of Agricultural and*
1023 *Biological Science*, 6(1), 48-55.
1024 [https://www.researchgate.net/publication/260793689_The_Dynamics_of_Potassium_K](https://www.researchgate.net/publication/260793689_The_Dynamics_of_Potassium_K_in_Representative_Soil_Series_of_Ghana)
1025 [_in_Representative_Soil_Series_of_Ghana](https://www.researchgate.net/publication/260793689_The_Dynamics_of_Potassium_K_in_Representative_Soil_Series_of_Ghana)
1026 Yeboah, E., Asamoah, G., Kofi, B., & Abunyewa, A. A. (2016). Effect of Biochar Type and Rate of
1027 Application on Maize Yield Indices and Water Use Efficiency on an Ultisol in Ghana.
1028 *Energy Procedia*, 93, 14-18. <https://doi.org/10.1016/j.egypro.2016.07.143>
1029 Zingore, S., Delve, R. J., Nyamangara, J., & Giller, K. E. (2007). Multiple benefits of manure: The
1030 key to maintenance of soil fertility and restoration of depleted sandy soils on African
1031 smallholder farms. *Nutrient Cycling in Agroecosystems*, 80(3), 267-282.
1032 <https://doi.org/10.1007/s10705-007-9142-2>
1033 Zingore, S., Mutegi, J., Agesa, B., Tamene, L., & Kihara, J. (2015). Soil degradation in sub-Saharan
1034 Africa and crop production options for soil rehabilitation. *Better Crop.*, 99(1), 24-26.
1035 <https://cgspace.cgiar.org/handle/10568/68702>

1036

1037 **Supplementary data**

1038 Appendix A (*Supplementary Table*)

1039 Appendix B (*Supplementary Figure*)

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

3 Anselme K. K. Kouame^a, Prem S. Bindraban^{a, *}, Isaac N. Kissiedu^a, Williams K. Atakora^a, Khalil El Mejahed^b

4 ^aInternational Fertilizer Development Center (IFDC), Muscle Shoals, AL 35662, USA.

5 ^bMohammed VI Polytechnic University (UM6P), AgroBioSciences, Lot 660, Hay Moulay Rachid, 43150 Ben Guerir, Morocco.

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

7 1 INTRODUCTION

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

17 However, farmers face changing and increasingly unpredictable weather conditions,
18 drastically reducing soil fertility, and typically use local or inbred crop varieties. Bationo et al.
19 (2018) reported that soil nutrient depletion rates of about 35 kg N, 4 kg P, and 20 kg K per hectare
20 are worrisome and prevalent in all agro-ecological zones (AEZs) in Ghana, with nitrogen (N)
21 and phosphorus (P) being the most deficient nutrients (Zingore et al., 2015). As a result, yields
22 obtained by smallholder farmers are far below the potentially attainable yields, hampering
23 agricultural production and jeopardizing economic development and food security (Adzawla,
24 et al., 2021).

25 One solution is to increase fertilizer application by farmers. However, it is increasingly
26 understood that crop yield in many areas of Africa, including Ghana, is depressed by a variety
27 of soil degradation problems and many other factors, such as crop variety, soil organic matter,
28 and soil depth (Sadras et al., 2001; Kpotor et al., 2014; Tetteh et al., 2016; Guilpart et al., 2017;
29 Leenaars et al., 2018). Furthermore, high variability in climatic conditions (rainfall and
30 temperature) causes uncertainties in agricultural productivity, with profound impacts on the
31 ecology, economy, and social welfare of rural farmers (Onduru et al., 2007; Kyei-Mensah et al.,
32 2019).

33 Despite current low crop productivity, Ghana could intensify production and significantly
34 close the current yield gaps of major cereals (Bationo et al., 2018; van Loon et al., 2019), since it
35 has been estimated that only about 20%, on average of the potential maize yield is being
36 achieved across the country (GYGA, 2021). For example, addressing nutrient deficiencies by

37 applying fertilizer alone would help to reduce the maize yield gap to 50% of the attainable yield
38 (Mueller et al., 2012). However, Adzawla, et al. (2021) and SRID/MoFA (2021) have reported
39 that the maize yield is around 2 t ha⁻¹ even with the application of nitrogen (N), phosphorus (P),
40 and potassium (K) compound fertilizers. Subsequently, Bua et al. (2020) found that maize yield
41 is highly variable, with yields ranging from a mere 500 kg ha⁻¹ to more than 8 t ha⁻¹. This large
42 yield variability depresses farmers' incentive and ability to purchase fertilizers in subsequent
43 seasons (Njoroge, 2019).

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

49 Models such as the QUantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) have
50 been advocated by several studies for estimating crop yield (Tabi et al., 2007; Tittonell et al.,
51 2008; Wijayanto et al., 2012; Xu et al., 2013; Ren et al., 2015), and specifically recommended for
52 use in Ghana as well (Wijayanto et al., 2012; Antwi et al., 2017). QUEFTS was developed for
53 maize and considers soil chemical properties (pH, organic carbon, extractable and total
54 phosphorus, exchangeable potassium, and organic nitrogen) and fertilizer application as the
55 input variables (Janssen et al., 1990; Sattaria et al., 2014). The model assumes that all other
56 production factors are optimal and does not consider the maize variety, soil physical properties,
57 or climatic variables. But in real life, maize grain yield could be expressed by equation (1)
58 according to Giller et al. (2013), as the result of "Variety", biophysical "Environment"
59 (temperature and precipitation), "Soil" (acidity, texture, root depth, limiting nutrients), and
60 management, including mineral fertilizers interactions:

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

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

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

70 2 MATERIALS AND METHODS

71 2.1 Study area

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

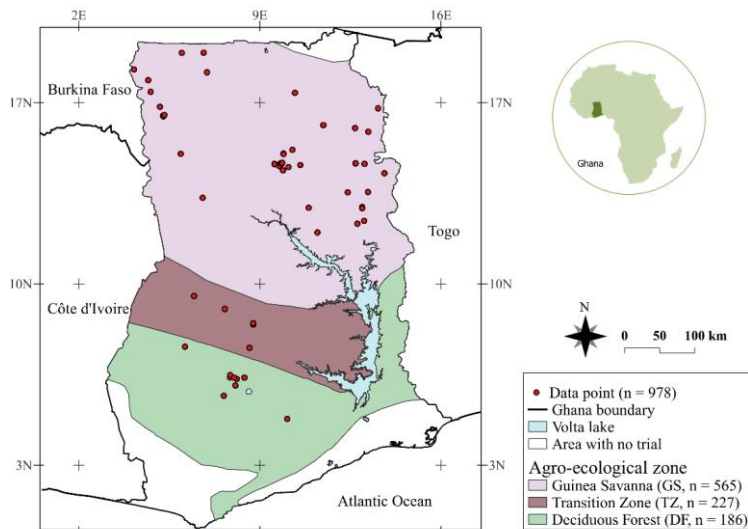


Figure 1: Geographic distribution of the experimental locations on a map of Ghana that shows the country's agro-ecological zones. Many experimental locations overlap.

[Agroecological map source: \(Antwi et al., 2014\)](#)

79 (Nkrumah et al., 2014). The average temperatures in these AEZs are between 24°C and 34°C
80 (Ghansah et al., 2018; Darko et al., 2019).

81 **2.2 Yield data**

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

87 **2.3 Fertilizer data**

88 Fertilizer treatments were heterogeneous and included organic fertilizers only, organic
89 fertilizers in combination with inorganic fertilizers, and inorganic fertilizers only. Organic
90 fertilizers, including cow dung; poultry, goat, and sheep droppings; compost; town waste;
91 biochar and palm bunch ash, were all converted into N, P₂O₅, and K₂O rates (Fening et al., 2009;
92 Adjei-Nsiah, 2012; Kanton et al., 2016; Badu et al., 2019). The lowest and highest rates of nitrogen
93 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
94 kg ha⁻¹. The lowest and highest rates of P₂O₅ (PF) and K₂O (KF) were 3 kg ha⁻¹ and 90 kg ha⁻¹,
95 with common rates of 20 kg ha⁻¹, 40 kg ha⁻¹, and 60 kg ha⁻¹. To assess the fertilizer contribution
96 to yield, data points were categorized into 2 treatments: “with_fertilizer” and
97 “without_fertilizer.”

98 **2.4 Soil and environmental data**

99 Table 1Table 1 presents the continuous variables used in the models. Soil chemical and physical
100 properties data were obtained from the African SoilGrids (ISRIC), at 250 m of resolution for the
101 0 - 30 cm topsoil. Rainfall and temperature datasets are from the WorldClim (Fick et al., 2017)
102 database, at a spatial resolution of 1 km² and land elevation data with a resolution of 30 meters
103 from Shuttle Radar Topography Mission (SRTM) Digital Terrain Elevation.

104 The geographic coordinates of each data point were overlaid on the maps to extract soil and
105 environmental data. Mean, minimum, and maximum temperatures and total rainfall for each
106 data point were obtained from aggregated monthly measurements according to the

107 recommended rainy season for planting maize in Ghana, on the basis that maize was harvested
 108 no later than 120 days after the date of sowing (Adu et al., 2014).

109 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	

Formatted Table

Formatted Table

110 ⁽¹⁾ Precipitation is the sum for the estimated maize growing season, ⁽²⁾ Minimum, maximum, and mean temperatures are the mean
 111 for the estimated maize growing season.

112 2.5 QUEFTS model

113 QUEFTS was implemented in the R software, based on Sattaria et al. (2014). QUEFTS' calibration
 114 and validation parameters from Wijayanto et al. (2012) and Antwi et al. (2017) were used to
 115 replace the Janssen et al. (1990)' default parameterization. The maximum yield was set to 10 t
 116 ha⁻¹. Soil-available P in P-Mehlich3 and the exchangeable K in K-NH₄Ac extracted from the
 117 ISRIC maps were converted into P-Olsen and K-Mehlich3 extractable based on Sawyer et al.
 118 (1999)' pedo-transfer functions. The performance of the QUEFTS model was assessed using the
 119 coefficient of determination (R²) (Krause et al., 2005), through a linear regression between
 120 QUEFTS-estimated and observed yield, and the significance of correlation (r) was determined
 121 based on a p-value < 0.05.

122 2.6 Linear Mixed Effects Model

123 Prior to statistical modelling, data points were subjected to the Shapiro-Wilk normality test and
 124 were not found to be normally distributed (Supplementary Table S2S 2). Therefore, maize yield
 125 was transformed and normalized to meet the assumptions of homoscedasticity and

126 homogeneity of analysis of variance (ANOVA) and LMM errors variance (Supplementary
127 Figure S_1). To evaluate the effectiveness of the normalization technique applied, the
128 “bestNormalize()” function of the bestNormalize R package (Peterson, 2021) was used. Thus,
129 the out-of-sample method via 10-fold cross-validation with 3 repeats was performed to estimate
130 the Pearson P-statistic (normality statistic). The transformation technique was selected
131 according to the calculated value of the normality statistic.

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

135 Variance inflation factors (VIFs) were calculated to measure the inflation of the predictor
136 coefficients due to collinearities between the independent numerical variables in ~~Table 1~~
137 ~~1~~. The “vif()” function of the car R package (Fox et al., 2019) was used to calculate the VIF values.
138 Silt and sand had $VIF_s > 10$, so multicollinearities were very likely. They were considered as
139 potential predictors to be eliminated. However, the VIF for sand became less than 3 when silt
140 was removed from the explanatory variables. Soil organic carbon (SOC) and total nitrogen
141 (TOTN) also had $VIF_s > 5$. As with sand, the VIF for SOC became less than 5 when TOTN was
142 removed. Therefore, in the rest of the ANOVA and LMM analysis, silt and total nitrogen were
143 removed as predictors. On the other hand, the maximum temperature (MMaT) and minimum
144 temperature (MMiT) also had a $VIF > 5$. Instead of deleting one of them, they were combined,
145 thus creating a new explanatory variable (i.e., mean temperature - MMeT).

146 The R function “TukeyHSD()” was used to perform the Tukey-Kramer test, and the
147 multcompView R package (Graves et al., 2019) was used to compact the letter display to indicate
148 significant differences between AEZ, Treatment, and Treatment*AEZ subgroups following
149 equation 2, where Y_{ijk} is the k th observed value of yield formed by the i th level of treatment
150 effect, the j th level of AEZ effect, the ij th level of interaction between treatment and AEZ, and
151 the error (ϵ_{ijk}).

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

152 The means of the maize yield were significantly different at $p < 0.05$. Tukey’s Honest
153 Significant Difference (TukeyHSD) multiple comparison analysis method tests were used

154 because there were unequal data point sizes among AEZ and Treatments subgroups (Marusteri
155 et al., 2010; McHugh, 2011).

156 One-way ANOVA equations 3, 4 and 5 were used to measure the granularity of the random
157 effect (i.e., Intraclass Correlation Coefficient ~~---~~ ICC) by Treatment, AEZ, and Year on maize
158 yields, respectively. The ICC was calculated as the ratio of the variance between the random
159 effect and the model's total variance (Nakagawa et al., 2010). The one-way ANOVA equations
160 with random effects were constructed as follows:

$$Y_{ij_Treat} = Treatment_i + \epsilon_{ij} \quad (3)$$

$$Y_{ij_AEZ} = AEZ_i + \epsilon_{ij} \quad (4)$$

$$Y_{ij_Year} = Year_i + \epsilon_{ij} \quad (5)$$

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

163 Since the objective was to determine if the variability in maize yield was related and could
164 be explained by the variables listed in [Table 1](#) ~~Table 1~~ and maize variety type, a LMM was
165 chosen. According to Dargie et al. (2022), LMM accounts for sample size imbalance and
166 confounding effects of uncontrolled variables, as in our case. So, maize yield was modelled in
167 three majors components: a linear function with a fixed effect trend " $f()$," with i th level of the
168 ~~covariate explanatory variable~~ (environmental factors ~~---~~ ENV, soil physical properties ~~---~~ SPP,
169 soil chemical properties ~~---~~ SCP, maize variety type ~~---~~ VAR, fertilizer rate ~~---~~ FER), where the
170 intercept was allowed to vary by AEZ_j (random effect of the j th AEZ) and $Year_k$ (random effect
171 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
172 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}
173 or ϵ_{ij}) denoting the small-scale fluctuations around " $f()$ ". The LMMs 11_{step} and 17_{step_AEZ} were
174 optimized by backward selection of variables using the lmerTest R package (Kuznetsova et al.
175 2017) "step()" function. At the level of each AEZ, LMM 17_{step_AEZ} was used to reveal the AEZ's
176 variables that were linked to maize yield variability. As a result, $Y_{ij_AEZ_step}$ represented the
177 maize yield modelled in a specific and known AEZ (i.e., GS, TZ, or DF). However, stepwise
178 LMM 17_{step_AEZ} ~~using the "step()" function~~ in DF, showed that the year's ICC was 0%. Therefore,
179 LMM 17_{step_AEZ} was converted into a Multiple Linear Regression (MLR_{step_DF}) formed with the
180 trend function " $f()$ " and the error term (ϵ_{ij}) ~~in DF~~. To avoid confusing models, GS, TZ, or DF

181 was assigned as a subscript to those LMMs or MLRs that referred to yield modelling in a
 182 particular AEZ, and all LMM where the AEZs were not assigned as a subscript referred to the
 183 yield modelling across the entire set of data points. The LMMs were constructed as follows:

$$(LMM\ 6_{ENV})\quad Y_{ijk_ENV} = f(ENV_{Environmental\ factors})_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (6)$$

$$(LMM\ 7_{SPP})\quad Y_{ijk_SPP} = f(SPP_{Soil\ physical\ properties})_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (7)$$

$$(LMM\ 8_{SCP})\quad Y_{ijk_SCP} = f(SCP_{Soil\ chemical\ properties})_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (8)$$

$$(LMM\ 9_{VAR})\quad Y_{ijk_VAR} = f(VAR_{Maize\ variety\ type})_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (9)$$

$$(LMM\ 10_{FER})\quad Y_{ijk_FER} = f(FER_{Fertilizer\ rate})_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (10)$$

$$(LMM\ 11_{step})\quad Y_{ijk_step} = f(ENV_{Environment}, SPP_{Soil}, SCP, VAR_{Variety}, FER_{Fertilizer})_i + AEZ_j + Year_k + \epsilon_{ijk} \quad (11)$$

$$(LMM\ 12_{ENV_AEZ})\quad Y_{ij_ENV_AEZ} = f(ENV_{Environmental\ factors})_i + Year_j + \epsilon_{ij} \quad (12)$$

$$(LMM\ 13_{SPP_AEZ})\quad Y_{ij_SPP_AEZ} = f(SPP_{Soil\ physical\ properties})_i + Year_j + \epsilon_{ij} \quad (13)$$

$$(LMM\ 14_{SCP_AEZ})\quad Y_{ij_SCP_AEZ} = f(SCP_{Soil\ chemical\ properties})_i + Year_j + \epsilon_{ij} \quad (14)$$

$$(LMM\ 15_{VAR_AEZ})\quad Y_{ij_VAR_AEZ} = f(VAR_{Maize\ variety\ type})_i + Year_j + \epsilon_{ij} \quad (15)$$

$$(LMM\ 16_{FER_AEZ})\quad Y_{ij_FER_AEZ} = f(FER_{Fertilizer\ rate})_i + Year_j + \epsilon_{ij} \quad (16)$$

$$(LMM\ 17_{step_AEZ})\quad Y_{ij_AEZ_step} = f(ENV_{Environment}, SPP_{Soil}, SCP, -VAR_{Variety}, FER_{Fertilizer})_i + Year_j + \epsilon_{ij} \quad (17)$$

Formatted: Left

184 Parameter estimation for the variance components in the LMMs was done with the REstricted
 185 Maximum Likelihood (REML) approach (Searle et al., 1992) using the “lmer()” function in the
 186 lme4 R package (Bates et al., 2015). To assess the significance of LMM and MLR fixed effects and
 187 also to accommodate imbalances in data point sizes among some predictors, an ANOVA table
 188 with F-tests and p-values, using Kenward-Roger’s method for denominator degrees-of-freedom
 189 and F-statistic, was used (Spilke et al., 2005). R² statistics for mixed-effects model from
 190 Nakagawa et al. (2017)’ were used to assess the goodness of fit of the LMMs and MLR_{step_DF}
 191 fixed effect trend “f()”. The predictors groups, i.e., ~~environmental factors (ENV), soil physical~~
 192 ~~properties (SPP), soil chemical properties (SCP), maize variety (VAR), and fertilizer rate (FER),~~
 193 involved in the fixed-effect statistics modellings, except for LMMs 11_{step} and 17_{step_AEZ} and
 194 MLR_{step_DF}, were compared using the marginal R² (R²_m) to explain the maize yield variability
 195 related to the trend function (f) and also to rank the predictor groups according to their
 196 explanatory power.

197 2.7 Random Forest model

198 The RF regression model was trained to (i) identify, prioritize, and rank the variables in Table 1
 199 Table 1 that were most important in explaining the variability of maize yields across AEZs, and
 200 (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

201 The measure of variable importance worked by calculating the increase in RF's prediction error
202 after permuting the variables in equation 18. RF is a non-parametric modelling machine learning
203 technique that has been gaining popularity in agricultural data analysis; it is resistant to outliers
204 and can better handle both straightforward linear and complex nonlinear associations compared
205 to LMM/MLR (Han et al., 2021).

$$Y_{i_rf} = f(\text{ENVEnvironment, SPPSoil, SCP, -VAR Variety, FERFertilizer})_i + \epsilon_i \quad (18)$$

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

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

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

228 3 RESULTS

229 3.1 Cropping system characteristics

230 In soils across AEZs, sand predominated over silt and clay, which were present in roughly equal
231 amounts (Supplementary Table [S3](#)). DF soils contained by far the greatest levels of soil
232 organic carbon (SOC), total phosphorus (TOTP), and cation exchange capacity (CEC) than TZ
233 and GS soils. Across the AEZs, potassium (K) and phosphorus (P) were generally present in
234 comparable amounts but at levels relatively below optimal values for P (Kugbe et al., 2019;
235 Daniel et al., 2021) and at a good level for K (Antwi et al., 2016). In addition, soil pH was around
236 6 and fluctuated only very slightly from one AEZ to another (Bationo et al., 2018). Generally, the
237 data points with the deepest root zones (RootDEP) on average were observed in soils of the TZ,
238 and the data points with the shallowest soils were in the GS (Supplementary Table S 3), as also
239 reported by (Bationo et al., 2018).

240 The coolest average temperatures \approx 26°C were observed in DF and TZ; however,
241 GS registered the highest average amount of precipitation during the maize growing season
242 (Table 1 Supplementary Table S 3). In general, the highest land elevations (ELV) were in DF at
243 an average of 270 m, followed by TZ at an average ELV elevation of 247 m.

244 Across AEZs, the average rate of nitrogen fertilizer (NF), phosphorus fertilizer (PF), and
245 potassium fertilizer (KF) was 67 kg ha⁻¹, 30 kg ha⁻¹, and 29 kg ha⁻¹, respectively. The highest
246 average NF rate was in DF at 80 kg ha⁻¹ (Supplementary Table S 3 Table S4). On the other hand,
247 the average PF (34 kg ha⁻¹) and KF (33 kg ha⁻¹) rates were higher in GS.

248 Open-pollinated varieties (OPVs) accounted for 69% of the maize varieties across all data
249 points, with Obatanpa being the most widely OPV variety (Supplementary Table S 4 Table S5).
250 GH 110, Etubi, Mamaba, and Pannar53 were the only 4 hybrid varieties, representing 20.5% of
251 all data points. Other variety types, categorized as “Others,” included “QPM,” “Entry,” and
252 “Local variety” and accounted for 10.5% of all data points.

253 3.2 Maize yield characteristics

254 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,
255 similar to those reported by (SRID/MoFA, 2021). The main effect of treatments was
256 significant and large ($p < 0.05$) (Supplementary Table S 5 Table S6). Tukey’s HSD test found that

Field Code Changed

257 the mean yield was significantly different between fertilized and control data points ($p < 0.05$)
258 (Supplementary [Table S 5 Table S7](#), Figure 2_A).

259 The main effect of AEZ was statistically significant and medium ($p < 0.05$), and the interaction
260 between treatment and AEZ was also statistically significant but weak ($p < 0.05$) ([Supplementary](#)

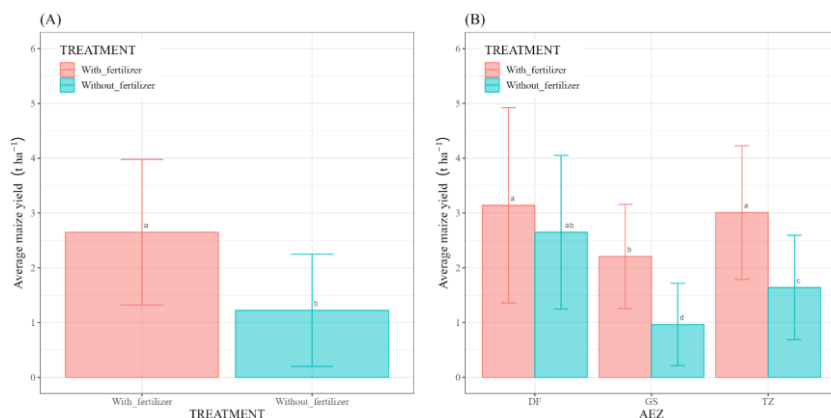


Figure 2: Average maize yield ($t\ ha^{-1}$) of treatments with fertilizer and without fertilizer (A) and the interaction between treatments and agro-ecological zone (AEZ) (B). DF (Deciduous Forest), GS (Guinea Savanna), and TZ (Transition Zone). Error bars are standard deviation bars. Different letters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

261 [Table S 5](#)). Mean yield was significantly ($p < 0.05$) different between GS and TZ data points and
262 between GS and DF data points, but not different between DF and TZ data points. In DF, mean
263 yield was not significantly ($p < 0.05$) different between fertilized and control data points;
264 however, mean yield was significantly ($p < 0.05$) different between fertilized and control data
265 points in TZ and GS (Supplementary [Table S 5 Table S8](#), Figure 2_B).

266 In the one-way ANOVA with random effects, yield varied with treatment, AEZ, and year.
267 The ICC indicated that the random effects (i.e., treatments, AEZs, and years) accounted for 46%,
268 20%, and 33% of the total variation in yield, respectively (Supplementary [Table S 6, S 7 and S 8](#)
269 [Table S9, S10, and S11](#)).

270 3.3 QUEFTS model yield estimated

271 When considering all the data points, the ~~estimated~~ average ~~of estimated yields using from~~ the
272 QUEFTS model ~~and observed yields averaged was~~ 3.3 $t\ ha^{-1}$ ~~and 2.2 $t\ ha^{-1}$, respectively~~. A linear

273 regression between observed and QUEFTS-estimated yields, based on the entire data points
 274 involved in the simulation (size effects of pH, SOC, P, TOTP, TOTN, K, NF, NP, and KF were
 275 confounded), revealed that soil chemical properties and fertilizer rate explained only 19% ($r =$
 276 0.44 , $p < 0.05$) of total yield variability (Figure 3_A).

277 **Figure 3** shows the linear regression between observed and QUEFTS-estimated yields
 278 in GS, TZ, and DF (size effects of pH, SOC, P, TOTP, TOTN, K, NF, NP, and KF were confounded
 279 in each AEZ). Indeed, there were weak correlations between observed and QUEFTS-estimated

Formatted: Indent: First line: 0"

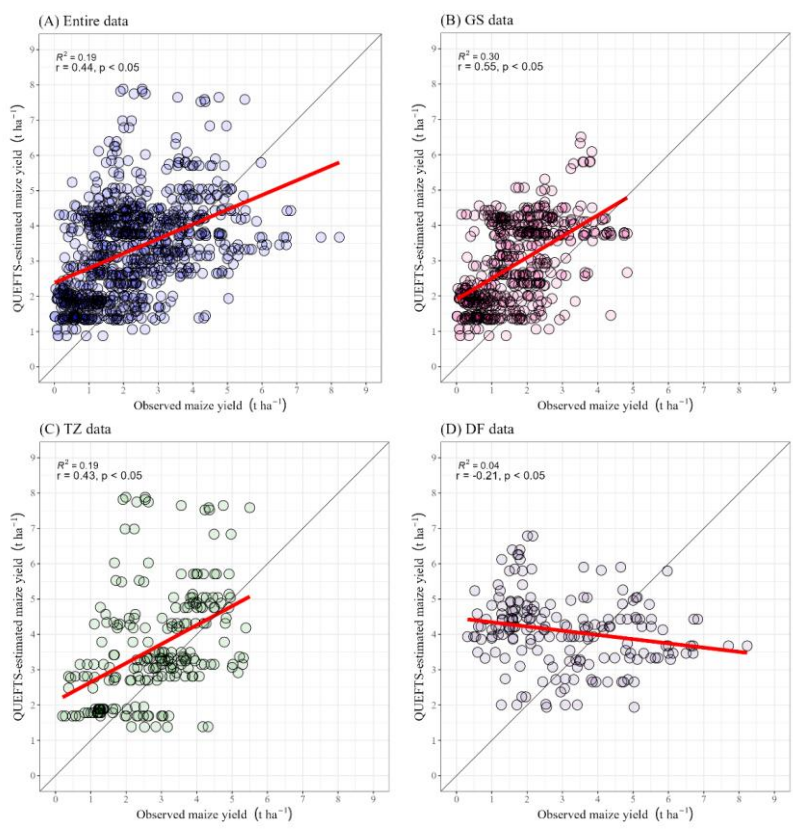


Figure 3: Relationship between observed and QUEFTS-estimated maize yield ($t\ ha^{-1}$) across all data points (A) and per AEZ (B, C, D). The R^2 indicates the portion of yield variability explained by soil chemical properties and fertilizers (confounded effect). The bold red line represents the linear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

280 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,
 281 observed and QUEFTS-estimated yields were negatively correlated ($r = -0.21$, $p < 0.05$), but were
 282 positively correlated in GS ($r = 0.55$, $p < 0.05$) and TZ ($r = 0.43$, $p < 0.05$)

283 Furthermore, when yields from the fertilized (size effects of pH, SOC, P, TOTP, TOTN, K,
 284 NF, NP, and KF were confounded) and non-fertilized (size effects of pH, SOC, P, TOTP, TOTN,
 285 K were confounded) trials were simulated separately, linear regression between observed and
 286 QUEFTS-estimated yields showed that QUEFTS explained only 2% ($r = 0.15$, $p < 0.05$) of the
 287 variability in yield in the fertilized data points and 8% ($r = 0.28$, $p < 0.05$) in the non-fertilized
 288 data points (Supplementary Figure [S_2_A](#), B).

289 At the annual level, in 2001, 2002, 2007, 2008, 2010, 2011, and 2012, linear regressions between
 290 observed and QUEFTS-estimated yields showed a good and statistically significant positive
 291 correlation (Supplementary Figure [S_2](#)). In 2008, the effects size of soil chemical properties and
 292 fertilizer rate, which were confounded, explained 75% of the variability in maize yield.
 293 However, the QUEFTS model overestimated maize yield in most simulation scenarios.

294 3.4 Modelling maize yield variability across all data points

295 Variability in maize yield was greatly associated with soil chemical properties ([SCP](#)) across the
 296 3 AEZs. In the LMM, yield varied significantly ($p < 0.05$) with SOC, P, and K ([Table 2](#)~~Table 2~~,
 297 Supplementary [Table S 9](#)~~Table S12~~). A correlation matrix and a plot of yield on the CEC did not
 298 show a clear trend (Supplementary Figure [S_3_a_d](#)). In addition, pH was not a significant
 299 predictor of yield in LMM 11_{step} but was significantly ($p < 0.05$) related to yield in LMM 8_{SCP}, and
 300 a plot of yield on pH also revealed a clear and negative trend (Supplementary Figure [S_3_a_c](#)).
 301 Compared to P, K, pH, and CEC, the variation in yield due to SOC was greater ([F value = 0.35](#))
 302 ([Table 2](#)).

303 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)	ICC		
							R^2_m	AEZ	Year
LMM 11 _{step} ^(f)	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) 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 2, Supplementary Table S9). 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 S3a).

In the LMM analysis, MP was the only significant ($p < 0.05$) environmental predictor (ENV) of yield (Supplementary Table S9), but the correlation matrix showed that MMeT and ELV were likewise significantly ($p < 0.05$) related to the variation in maize yield (Supplementary Figure S3a). A plot of maize yield on MP showed a significant ($p < 0.05$) and clear negative trend (Supplementary Figure S3b).

Maize variety type (VAR) was significantly ($p < 0.05$) related to yield (Table 2, Supplementary Table S9). 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 S10). Tukey's HSD test found that there was no significant difference in mean yield between OPV and hybrid ($p = 0.543$), 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 S9). 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 S9). The ICC values in Table 2 and Supplementary Table S9 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.

332 **3.5 Maize yield determinants for each agro-ecological zone**

333 In GS, LMM demonstrated that maize yield significantly ($p < 0.05$) varied with ELV, CLAY,
 334 SOC, VARIETY, NF, PF, SAND, and CEC (Table 3, Supplementary Table S11).
 335 However, the contributions of ELV and CEC to the variations in maize yield in GS did not show
 336 a clear and significant trend (Supplementary Figure S4 a, b, e). The PF rate accounted for 48%
 337 of the R^2_m (0.39); therefore, GS was found to have a larger significant yield fluctuation due to PF
 338 rate than the other AEZs. Random effects (i.e., year) accounted for 2-32% of the total variation
 339 in maize yield in GS. The highest explanation power of yield variability ($R^2_m = 0.35$) in GS was
 340 shown by fertilizer rate FER, followed by maize variety VAR ($R^2_m = 0.4$), soil physical SPP and
 341 chemical properties SCP ($R^2_m = 0.03$), and environmental factors ENV ($R^2_m = 0.02$).

342 In TZ, LMM analysis showed that maize yield variability was significantly ($p < 0.05$) related
 343 to MMeT, ELV, CLAY, VARIETY, SOC, NF, and PF (Table 4, Supplementary Table S
 344 12). The combined contribution of ELV, CLAY, VARIETY, NF, and PF explained 54%
 345 of the variation in maize yield in TZ, whereas the year's ICC was 37% (Table 4). Fertilizer
 346 rates had the best explanatory power of maize yield variability ($R^2_m = 0.26$) in TZ, followed by
 347 the environmental factors ENV ($R^2_m = 0.21$), then maize variety VAR ($R^2_m = 0.20$), soil chemical
 348 properties SCP ($R^2_m = 0.19$), and finally soil physical properties SPP ($R^2_m = 0.19$).

349 Table 3: Significance of the effects of the explanatory variables in the linear mixed effects modelling of the set of yield data points
 350 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		

351 (†) Maize yield modelled with equation (17-47) in Guinea Savanna using stepwise linear mixed effects modelling.

352 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
LMM 17 _{step_TZ} (^t)	ELV	8.2	1	6.1	23.1	< 0.05	0.55	0.37
	CLAY	19.5	1	25.9	54.8	< 0.05		
	VARIETY	37.6	2	143.3	52.7	< 0.05		
	NF	44.9	1	215.4	125.9	< 0.05		
	PF	4.2	1	215.9	11.7	< 0.05		

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

354 In the MLR analysis, ELV, MP, SOC, CLAY, SAND, RootDEP, TOTP, P, K, CEC, VARIETY, ~~and~~
 355 ~~NF and PF~~ were significant (p < 0.05) predictors of maize yield in DF, explaining 79% of maize
 356 yield variability (Table 5). ~~The PF rate was not significantly related to maize yield variability in~~
 357 ~~Table 5 but weakly related to yield in the supplementary Table S22.~~ The LMM-analysis also
 358 revealed that maize yield significantly (p < 0.05) varied with K but was not significantly related
 359 to the KF rate (supplementary Table S13 Table S23). The ICC indicated that the random effects
 360 (i.e., year) accounted for 4-64% of the total variation in maize yield in DF. NF and PF rates and
 361 maize variety did not exhibit the same level of explanatory power of yield variability as in GS
 362 and TZ. Indeed, the strongest explanatory power for the variation in yield in DF was found in
 363 the ~~soil chemical properties~~ SCP (R²_m = 0.48), followed by ~~environmental factors~~ ENV (R²_m = 0.43),
 364 ~~maize variety~~ VAR, and ~~soil physical properties~~ SCP (R²_m = 0.24).

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

Model	Variable	Sum Sq	DF	F value	Pr(>F)	R ²
MLR _{step_DF} (^t)	ELV	1.6	1	7.4	< 0.05	0.79
	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	
	TOTP	2.9	1	13.8	< 0.05	
	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	

367 (^t) Maize yield modelled in Deciduous Forest using stepwise multiple linear regression model.

368 LMM analysis showed that the contribution of maize variety type in the variations in yield
369 in the 3 AEZs was also significant ($p < 0.05$) (~~Table 3, 4, 5~~~~Table 3, Table 4, and Table 5,~~
370 ~~Supplementary Table S 14, S 15, S 16~~~~Table S24, S25, S26~~). A one-way ANOVA revealed that there
371 was a significant ($p < 0.05$) difference in mean yield between at least two variety types in GS,
372 TZ, and DF (Supplementary ~~Table S 14, S 15, S 16~~~~Table S27, S28, S29~~). Tukey's HSD test found
373 that there was no significant difference in mean yield between OPV and hybrid in GS ($p = 0.253$)
374 and TZ ($p = 0.694$), but there was significant difference in mean yield in DF ($p < 0.05$). There was
375 a significant ($p < 0.05$) difference in mean yield between OPV and "Others" variety types in GS,
376 TZ, and DF, and also between hybrid and "Others" variety types in ~~the 3 AEZs~~~~GS, TZ, and DF~~.

377 **3.6 Modelling maize yield using Random Forest**

378 Considering 75% of all data points, the results of the RF model showed that the NF rate
379 explained the largest portion (26.7%) of the variability in maize yield. The RF model revealed
380 that 4 variables (NF, MMeT, RootDEP, and MP) accounted for nearly 61% of the variability in
381 maize yield (Supplementary Figure ~~S1AS 7 A~~). In addition, a strong importance of RootDEP in
382 the entire maize yield variability was spotlighted, whereas LMM yield modelling did not.

383 The R^2 ranged between 0.71 (training data) and 0.75 (testing data), the RMSE ranged between
 384 750 kg ha⁻¹ (training data) and 700 kg ha⁻¹ (testing data), and the NSE ranged between 0.58
 385 (training data) and 0.64 (testing data), indicating that the RF model performed well in yield
 386 prediction (Figure 4 Figure 4A, Supplementary Figure S 72_B). Based on the importance of the
 387 variables in explaining yield using the training data points, maize yield prediction was made in
 388 each AEZ (Figure 4 Figure 4 B, C, D). In comparison to TZ and DF, the RF model's performance
 389 in GS was somewhat subpar. Indeed, in GS, RF-predicted maize yield showed the lowest values
 390 of NSE = 0.55 and R^2 = 0.74.

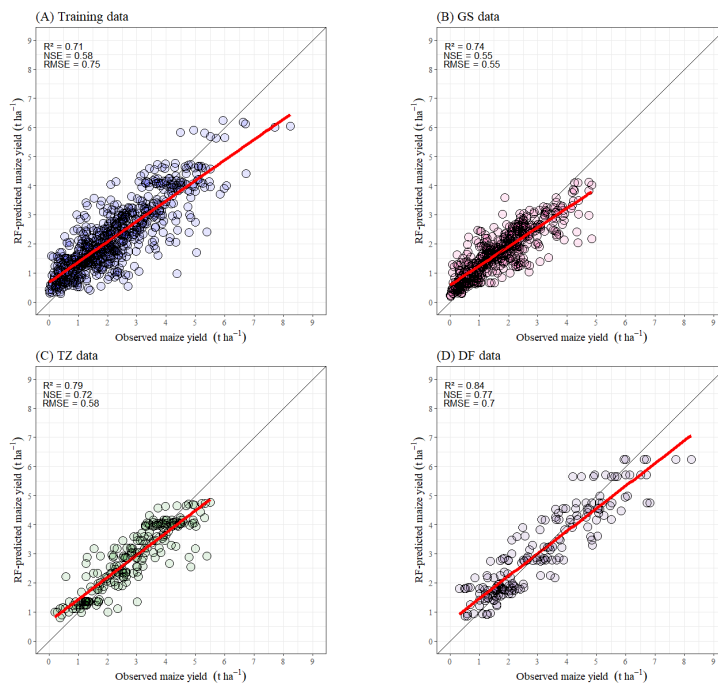


Figure 4: Relationship between observed maize yield and that predicted by the Random Forest (RF) model using 3 repeats of a 10-fold cross-validation approach. (A) The linear regression on ~~the~~-75% of the full entire data setpoint (training data point), (B) only the data setpoint from GS, (C) only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

391 **4 DISCUSSION**

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

398 ~~Data points in DF had high maize yields of 7-8 t ha⁻¹, which could be explained by the low MMiT~~
399 ~~(19°C) and MMaT (27°C) at the trial locations, sufficient MP during the crop growth period (520~~
400 ~~mm), high ELV (390 m), and the deep RootDEP of 150 cm for storing the rainwater. However,~~
401 ~~several studies on plant breeding have shown that the agronomic potential of certain maize~~
402 ~~varieties, such as Obatanpa (OPV), is between 4 and 5 t ha⁻¹ (Sallah et al., 2007; Adu et al., 2014;~~
403 ~~USAID/IFDC, 2015). The potential yields attributed to these varieties are averages from several~~
404 ~~research station trials (Sallah et al., 2007; Adu et al., 2014). Therefore, yields above these averages~~
405 ~~could be obtained in research trials under optimal agronomic practices and climatic conditions,~~
406 ~~which justifies why, in this study, the high yields of those maize trials were not considered~~
407 ~~outliers. In addition, Adu Gyamfi et al. (2019) reported average yields of 6.5 t ha⁻¹ with~~
408 ~~Obatanpa, well above its potential yield (according to research institutions), in GS which could~~
409 ~~be considered less fertile than TZ and DF given the soil and climatic conditions.~~

410 **4.1 Soil chemical properties and fertilizer rates impactingcontrolling yield**
411 **variability**

412 ~~In general, the QUEFTS model weakly explained the spatial and temporal observed maize yield~~
413 ~~variability. The QUEFTS model was calibrated and validated for the northern regions (i.e., the~~
414 ~~GS) and explained more variability there than in the other 2 AEZs. Yet, ~~t~~he low spatial and
415 temporal explanatory power of ~~the~~ QUEFTS model suggests ~~that other that~~ factors ~~other~~ than
416 ~~the~~ soil chemical properties and fertilizer rates only, ~~to not considered in QUEFTS~~ contributed
417 to the variability in maize yield as also reported by Debtanu et al. (2006) and (Onduru et al.,
418 (2007). Thus, the QUEFTS is unable to improve the understanding of native soil nutrient supply,
419 and maize yield for the wide range of soil, climate, and management conditions in the 3 AEZs,
420 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

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

Formatted: Superscript

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

438 The model overestimated grain yields at almost all data point locations. This overestimation
439 revealed bias in the model calibration. The accumulation and dilution coefficients and N, P, and
440 K recovery efficiency (RE) reported by Antwi et al. (2017) and Wijayanto et al. (2012) need to be
441 updated and adjusted considering the soil and climate specificity of the AEZs (Ezui et al., 2017).
442 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
443 kg K kg⁻¹. Antwi et al. (2017) and Wijayanto et al. (2012) also obtained similar RE values;
444 however, these RE values did not align with the RE in this study.

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

451 (p < 0.05) contributor to the variation in maize yield across the 3 AEZs in Ghana. Extractable P,
452 and K demonstrated a positive trend in the main effect among these soil chemical variables
453 (Figure S 33 aa), also identified by ~~Indeed, according to Braimoh et al. (2006),~~ maize yields will
454 be less than 520 kg ha⁻¹ under repeated cropping on soils with poor chemical properties. Yeboah
455 et al. (2016) and also showed that improving soil chemical properties, such as TOTN, ~~P, and K,~~
456 CEC, and pH, had a pronounced effect on maize grain yield ~~(Akolgo et al., (2020)).~~ Therefore,
457 knowledge about the condition of the soil's chemical properties and its evolution over time and
458 space is necessary to promote land management practices to maintain and improve maize
459 productivity and sustainable use of natural resources (Bindraban et al., 2000). However, a plot
460 of yield on CEC and soil pH revealed an inverse trend (Supplementary Figure S 4e3 c, d).

461 ~~Although Ghanaian soils are generally not very acidic (Supplementary Table S 343) and fairly~~
462 ~~optimal, however with the slightly negative relation with maize yield likely resultsing from the~~
463 ~~narrow pH range and associations with other variables such as , the ideal pH range for maize~~
464 ~~production is between 6 and 7 (Buah et al., 2017), and a reduction in pH will generally result in~~
465 ~~a negative trend for maize yield. However, because there was a positive and substantial~~
466 ~~association between pH and MMeT and temperature (Supplementary Figure S5 3 a, c, 5 a), it is~~
467 ~~likely that this negative trend between pH and yield was caused by the influence of another~~
468 ~~variable, such as temperature, as the pH did not vary too much and the standard deviation was~~
469 ~~extremely small.~~

470 ~~With tropical soils like in the DF having typical CEC values of around 10 cmol₍₊₎ kg⁻¹ (Osei,~~
471 ~~1995), the high values (83 cmol₍₊₎ kg⁻¹) of the CEC values for 30 data points situated in the DF~~
472 ~~suggested an negative though not significant (p = 0.09) relation with yield when the entire data~~
473 ~~point is considered (Supplementary Figure S 3 d). were much higher than average~~
474 ~~(Supplementary Figure S 26 e d), which accounted for the non significant (p = 0.09) negative~~
475 ~~trend between CEC and yield when the entire data point is considered (Supplementary Figure~~
476 ~~S 3 d). According to Osei (1995), tropical soils like those in the DF generally have low CEC~~
477 ~~(typically 10 cmol₍₊₎ kg⁻¹), and as a result, these extreme values (83 cmol₍₊₎ kg⁻¹) influenced the~~
478 ~~trend of linear relationship between CEC and yield (Supplementary Figure S 3 e). After~~
479 ~~excluding these data points turned the relation excessive CEC values from these 30 data points~~
480 ~~from the entire set of data points, the linear trend became significantly ((p < 0.05)) and positive~~

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

483 The LMM analyses revealed that compared to NF and PF rates, the KF rate was not a strong
484 predictor of maize yield in any of the AEZs. LMM and RF modelling of yields with all data
485 points also confirmed this result, with and showed that the contribution of KF in explaining 2.5-
486 3.5% of maize yield variability was only 2.5-3.5%. These findings suggest that K is not a limiting
487 nutrient for Ghana's maize production. Therefore, the significant ($p < 0.05$) association of K
488 and maize yield in Table 2 was probably due to the data points of the DF (Supplementary
489 Table S 13 Table S32, Supplementary Figure S 6 a). According to Yawson et al. (2011), Ghana's
490 forest soils will require frequent and split KF applications since they have a lower capacity to
491 maintain a long-term supply of K; however, the savanna soils will require less frequent but
492 higher K fertilization to satisfy the exchangeable pool and immediate plant nutrition
493 requirement. Different KF rates were not significantly related to maize yield, probably due to
494 Liebig's Law of the Minimum (Essel et al., 2020). Nevertheless, Application of KF will remain
495 important to prevent soil nutrient mining.

496 We observed significant but downward linear trends in the main effect for SOC in the 3-2
497 AEZs (Supplementary Figure S4a, S 5 a, S 6 a). This is explained by the fact that, in comparison
498 to soils where poor yields were reported, the bulk of high-yielding data points in GS, TZ and
499 DF were found in soils with a comparatively lower SOC. This negative trend between SOC and
500 yield in some AEZs data points, despite their inherently high soil SOC, indicates that suggests
501 yield variability to be is driven by other factors as well other than SOC. In research led by Kihara
502 et al. (2016), Shehu et al. (2019) and Sileshi et al. (2022), maize fertilization trials got poor yields
503 on soils with a relatively high larger SOC level obtained poor yields, while whereas the trials that
504 obtained the highest yields were found carried out on soils with a low SOC content. Across For
505 the entire set of data points, with a wider SOC range, LMM yield variability modelling
506 revealed a significant and positive trend of maize yield on SOC (Table 2, Supplementary Figure
507 S 3 f a), revealed a significant ($p < 0.05$) positive trend of maize yield on SOC. Logah et al. (2011)
508 and many other scholars found the same result (Lal, 2006; Solomon et al., 2016; Owoade et al.,
509 2020; Owoade et al., 2021). Therefore, SOC is used as an indicator of soil fertility (Tiftonell et al.,
510 2008), with and an application of NF in such soil will promote good plant nitrogen nutrition
511 and thus contribute to a significant and positive influence of NF on maize yield.

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

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

524 **4.2 Role of sSoil pPhysical characteristics role ion yield variability**

525 ~~Among the variables in the soil physical characteristics group, only RootDEP demonstrated a~~
526 ~~significant (p < 0.05) positive trend with yield variation when the entire set of maize yield data~~
527 ~~points was modelled using in LMM. It was indee even identified by the RF model as the third~~
528 ~~most important variable, right after the NF rate and MMeT (Supplementary Figure S 8A7 A).~~
529 ~~But a more specific analysis showed that this influence of RootDEP on maize yields of all the~~
530 ~~data points originated mostly from the DF data points (Supplementary Table S34S 13, Figure~~
531 ~~S9aS 6 a) with deeper soils. These results show how critical soil rooting depth is in maize~~
532 ~~production in Ghana. A study by Guilpart et al. (2017), revealed this importance of RootDEP on~~
533 ~~who used crop simulation to assess the sensitivity of rainfed maize yields as function of to the~~
534 ~~water-holding capacity of the root zone, showed that a modest maize surplus could be produced~~
535 ~~in SSA, including Ghana, provided rooting depths are comparable to those of other important~~
536 ~~granaries, such as the United States Maize Belt and the South American Pampas (Leenaars et al.,~~
537 ~~2018).~~

538 Sand and clay were found to be significantly (p < 0.05) associated with maize yield variability
539 in both LMM and MLR modelling. ~~However, Ghanaian soil being much sandier (Supplementary~~
540 ~~Table S35S 3), especially in GS where there is high potential for maize production (Boullouz et~~
541 ~~al., 2022), it has low fertility and less SOC content (Kanton et al., 2016; Bationo et al., 2018).~~

542 ~~Mobilian et al. (2021) reported that soil texture influences other soil properties, such as bulk~~
543 ~~density, water holding capacity, permeability, and porosity. For example, soils that are~~
544 ~~predominantly composed of sand particles have high permeability and low water holding~~
545 ~~capacity compared to soils with higher clay content. While a negative relation of sand content is~~
546 ~~expected with yield due to various associated soil properties like low water-holding capacity.~~
547 ~~This implies that the positive and significant trend between sand and maize yield~~
548 ~~(Supplementary Figure S.3 a, S.4 a, c, and S.6 a, g) on sandy soils could be related to the influence~~
549 ~~of organic fertilizers (poultry and cow manure) in this study (Dapaah et al., 2008; Quansah, 2010;~~
550 ~~Adjei-Nsiah, 2012; Kanton et al., 2016; Badu et al., 2019). According to Obi et al., (1995), Zingore~~
551 ~~et al. (2007), Uzoma et al. (2011) and; Frimpong et al., 2021; Obi et al., (1995); Zingore et al. (2017),~~
552 ~~the application of organic fertilizers such as cow and poultry manure biochar on sandy soil is~~
553 ~~not only beneficial for crop growth, but also significantly improves the physicochemical~~
554 ~~properties of the coarse sandy soils. Indeed, the application of organic fertilizer tends to~~
555 ~~improve the soil SOC content, which is favorably affecting e for good maize growth (Frimpong~~
556 ~~et al., 2021). The same result was found in Nigeria by Obi et al. (1995), who showed that~~
557 ~~application of poultry manure significantly improved maize yield in severely degraded sandy~~
558 ~~soil. Zingore et al. (2007) also reported that manure is a key nutrient resource in the tropics,~~
559 ~~especially on poorly buffered sandy soils, because of its multiple benefits to soil fertility.~~
560 ~~Moreover, the GS has a higher potential for maize production (Boullouz et al., 2022) that can be~~
561 ~~better exploited with appropriate fertilization and water management practices; organic manure~~
562 ~~being one of these favorable practices.~~

563 ~~According to our analyses, t~~he contribution of clay content to the variations in maize yield
564 was significant (~~Table 2~~Table 2). Clay plays an important role in the supply, retention, and
565 fixation of many macronutrients and micronutrients in the soil that improve maize crop
566 nutrition (O'Halloran et al., 1985; Batjes, 2011; Florence et al., 2017), ~~and improvesameliorates~~
567 ~~soil physical characteristics such as enhancing water holding capacity. However, i~~n GS,
568 ~~however,~~ a sizable ~~downward upward~~ trend in the maize yield with clay was observed
569 (Supplementary Figure S.4.d). ~~While Although~~ the correlation coefficient ~~was~~seems significant
570 ($p < 0.05$), the distribution of data points in the Supplementary Figure S.4.d does ~~clearly~~
571 ~~underline not seem to make~~ this negative trend ~~very clear,~~ likely for similar reasons as indicated
572 ~~for sand. In a study,~~Sileshi et al. (2022) ~~reports~~showed the ~~samsimilare~~ results, and ~~Similarly,~~

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

576 4.3 Contribution of environmental factors to yield variation

577 Climate has always been central to agricultural production in Ghana (Peprah, 2012; Nyuor et al.,
578 2016; Adjei et al., 2018). The analysis of LMM (Table 2, Supplementary Table S36S 9, S37S
579 11, S38S 12, S39S 13, and Figure S40 3 a, b, e, S41 5 b, S 6 a) and RF yield modelling
580 (Supplementary Figure S42 7 BA) showed that environmental variables (MMeT, MP, and ELV)
581 were related to maize yield variability, as also reported by (Onduru et al. (2007), Mugwe et al.
582 (2009), Fosu-Mensah et al., (2019;-), Kyei-Mensah et al., (2019;-), and Cudjoe et al., (2021).
583 However, the contribution of MP of to the entire data set point reveals a negative trend with
584 point yield variation showed a negative pattern (Supplementary Figure S43 3 a, b). Indeed, in
585 the DF locations where very high yields of (7-8 t ha⁻¹) were reported obtained, the amount of
586 rainfall was sufficient although lower compared to other areas with lower yields, was sufficient
587 Their larger water holder capacity of the deeper soils (RootDEP of 150 mm) made sufficient
588 water available for maize growth, as also reported by (Durodola et al., (2020) and, Bagula et al.,
589 (2022), to fill the soil reservoir where deep RootDEP (150 mm) and high ELV (390 m) were
590 observed, so that water would not be a limiting factor for maize growth. In addition, a
591 negative trend between MP and RootDEP support this logic and could also help explain why
592 high MP in some locations (with low RootDEP) did not tend to result in high maize yields
593 (Supplementary Figure S44 6 a, hi). Furthermore, optimal rainfall in low yielding areas confirms
594 our view that variables other than rainfall, which had a significant relationship with yield, also
595 play an important role in maize variability.

596 Baffour-Ata et al. (2021) reported that temperature was significantly and positively related to
597 maize yield in some regions of Ghana. This result was also observed in our analysis in TZ
598 (Supplementary Figure S45 5 a, b). However, according to Bationo et al. (2018), high air
599 temperatures, high light levels, and heat-trapping sandy soils combine to make the local
600 environment too hot for good plant growth. Therefore, the positive relation found increase in
601 maize yield in TZ may have resulted from coinciding with increasing temperatures may be due
602 to other non climatic factors, such as the use of improved high-potential, drought-tolerant
603 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

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

606 ELV, in addition to MP and MMeT, was also a significant factor in explaining maize yield
607 variability. In fact, the correlation matrices (Supplementary Figure ~~S16.3 a~~, S 4 a and S 517a)
608 revealed that ELV and some soil variables, including RootDEP, CLAY, and SOC content, were
609 positively correlated. Thus, for example, the significant ($p < 0.05$) trend between ELV and
610 RootDEP demonstrates that a high yield is more likely because the higher the ELV, the larger
611 the water pool available to the plant. A study by [Jiang et al. \(2004\)](#) ~~also confirmed that ELV is a~~
612 very useful factor in understanding variation in maize production and that many soil properties
613 are significantly dependent upon ELV ([Cooper, 1979](#); [Ovalles et al., 1986](#); [Kravchenko et al.,](#)
614 [2007](#)).

615 4.4 Yield variability related to maize variety

616 Overall, the results showed that maize variety type was significantly ($p < 0.05$) related to maize
617 yield variability, ~~as found by [Kpotor et al., \(2014\)](#) also.~~ [Bawa et al. \(2021\)](#) evaluated the
618 influence of different NF rates on maize yield and found ~~differential that varieties performed~~
619 ~~differently- responses depending on the rate NF rate, as did [Adu et al., \(2014\)](#) in Ghana, and~~
620 ~~[Abera et al. \(2017\)](#) in Ethiopia. The use of different types of maize varieties with different genetic~~
621 ~~potential and drought resistance, could induce highly variable yields depending on AEZ and~~
622 ~~fertilizer application ([Adu et al., 2014](#)). Similar outcomes were also documented in Ethiopia by~~
623 ~~[Abera et al. \(2017\)](#). Indeed, the use of different types of maize varieties coupled with different~~
624 ~~fertilizer rates (also highly variable) has undoubtedly contributed to maize yields variability~~
625 ~~([van Loon et al., 2019](#); [Bawa et al., 2021](#)). For example, [Afreh et al. \(2022\)](#) reported that,~~
626 depending on the interaction between maize variety and NF rate, farmers would need to apply
627 their NF during planting for Omankwa (OPV ~~variety~~) and 14-28 days after planting for hybrid
628 varieties to achieve 4.7 t ha⁻¹ and 6.5 t ha⁻¹, respectively. ~~However,~~ the coupled contribution of
629 environmental factors and fertilizer rates could explain the aberrant responses of varieties
630 between the AEZ's, ~~lack of significant difference between the mean yield of hybrids and OPV in~~
631 ~~GS and TZ, the significant difference between the mean yield of hybrids and OPV in DF, and~~
632 ~~the significant difference between the mean yield of hybrids and OPV with "Others" in GS, TZ,~~
633 ~~and DF. In fact, 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

634 substantial interactions between variety, NF rate, and trial site, which were very important in
635 determining maize yield.

636 4.5 Model performance

637 Crop yield is intricately influenced by a variety of genetic, environmental, and management
638 factors as well as their interactions. For those involved in agriculture, the ability to precisely
639 predict crop yield in a variety of geographic settings with changing environmental
640 circumstances is becoming more and more crucial (Wang, 2021). ~~However, the identifying of~~
641 ~~variables influencing maize yield (in Ghana) is highly complex because yields are influenced by~~
642 ~~interdependent and frequently drastically different climates, soil, and management variables~~
643 ~~(see also van Loon et al. (2019)).~~ Model cross-validation showed that RF performed well in maize
644 yield prediction across AEZs (Figure 4). The small value of RMSE observed in GS in
645 yield prediction, compared to TZ and DF, could be explained by the low mean yield observed
646 (μ) and the standard deviation (σ) less dispersed in GS ($\mu = 1.68 \text{ t ha}^{-1}$, $\sigma = 1$) than in TZ ($\mu =$
647 2.82 t ha^{-1} , $\sigma = 1.3$) and DF ($\mu = 3.05 \text{ t ha}^{-1}$, $\sigma = 2$).

648 QUEFTS is a model based on empirical processes and internal ~~curve~~ interaction (Janssen et
649 al., 1990), whereas statistical and machine learning models (i.e., LMM, MLR, and RF) directly
650 explain and predict yield without many internal processes. This could also justify why
651 LMM/MLR and RF performed better than QUEFTS at both ~~identifying the driving variables of~~
652 ~~explaining~~—maize yield variability and estimating yield (Bonilla-Cedrez et al., 2021).
653 Additionally, it was demonstrated that RF outperformed in terms of maize yield explanation
654 (R^2) since it is not subject to the same amount of normality assumptions that LMM and MLR are.
655 It was thus possible to clearly illustrate the importance of RootDEP in explaining yield
656 variability. The use of linear models, such as LMM and MLR, alongside RF facilitates the
657 interpretation of how variables are related to maize yield variability in cases where RF has only
658 highlighted and ranked the importance of factors related to maize yield variability and not how
659 these factors vary with maize yield across AEZs. ~~This is why a~~ Our multi-model approach ~~was~~
660 used in this study ~~reveals this added value to unravel driving factors of maize yield.~~
661 ~~particularly given that van Loon et al. (2019) claimed that identifying the variables influencing~~
662 ~~maize yield in Ghana is much more difficult because yields are influenced by interdependent~~
663 ~~and frequently drastically different climates, soil, and management variables.~~

Formatted: Heading 2

Formatted: Indent: First line: 0.2"

Formatted: Superscript

664 5 CONCLUSION

665 Increasing maize yields to meet food demand while increasing farmers' profitability remains a
666 major challenge for Ghanaian farming systems. The overall objective of this study was to
667 characterize rainfed maize yield variation, understand the sources of variability, and predict
668 maize yield using robust statistical [and modelling](#) tools. The high variability in yield [to fertilizer](#)
669 [application](#), both within and across AEZs and years, reflects a high degree of heterogeneity in
670 ~~the fertilizer rate~~, soil characteristics, environmental factors, maize varieties, and growing
671 conditions at various spatial and temporal scales. This study provided new information on
672 maize yield variability in Ghana. When the entire set of data points was involved in the yield
673 modelling, LMM/MLR and RF models showed that the NF rates were the most important
674 factors explaining the maize yield variability in Ghana. RF revealed also that the second most
675 important factor was MMeT, followed by RootDEP and MP. Since yield variability was
676 significantly related to AEZ, in DF, soil chemical properties and environmental factors ~~instead~~
677 ~~guided the explained portion~~ of variability. In TZ and GS, ~~it was rather~~ NF and ~~NPF drove~~
678 ~~that led the~~ yield explanation. This ~~suggest that may imply~~ the inherently high soil fertility in
679 DF overrules the importance of fertilizers, while fertilizers drive yield increase in the less-fertile
680 TZ and GS. In all 3 AEZs, the ~~high used of variability different types of maize varieties of the~~
681 ~~variety (OPV, hybrid, or local)~~ also played an important role in the overall yield variability
682 observed. [Importantly, our multi-model approach that combines advanced statistical methods,](#)
683 [crop-soil modelling, and machine learning, reveals its ability to identify drivers for yield despite](#)
684 [the huge complexity of the production system.](#)

685 **Author contribution**

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

690 ***Acknowledgement***

691 This study received funding from the Université Mohammed VI Polytechnique and the OCP
692 group, through the Fertilizer Research and Implementation Program, led by the International
693 Fertilizer Development Center (IFDC). The funder was not involved in the study design,

694 collection, analysis, interpretation of data, the writing of this article, or the decision to submit it
695 for publication.

696 **Declaration of Competing Interest**

697 The authors declare that they have no conflict of interest.

698 **Data availability**

699 The authors do not have permission to share data.

700 **Supplementary data**

701 Appendix A (*Supplementary Table*)

702 Appendix B (*Supplementary Figure*)

703 **References**

- 704 Abera, T., Debele, T., & Wegary, D. (2017). Effects of Varieties and Nitrogen Fertilizer on Yield
705 and Yield Components of Maize on Farmers Field in Mid Altitude Areas of Western
706 Ethiopia. *International Journal of Agronomy*, 2017, 1-13.
707 <https://doi.org/10.1155/2017/4253917>
- 708 Adjei-Nsiah, S. (2012). Response of Maize (*Zea mays* L.) to Different Rates of Palm Bunch Ash
709 Application in the Semi-deciduous Forest Agro-ecological Zone of Ghana. *Applied and*
710 *Environmental Soil Science*, 2012, 870948. <https://doi.org/10.1155/2012/870948>
- 711 Adjei, V., & Kyerematen, R. (2018). Impacts of Changing Climate on Maize Production in the
712 Transitional Zone of Ghana. *American Journal of Climate Change*, 07(03), 463-476.
713 <https://doi.org/10.4236/ajcc.2018.73028>
- 714 ~~Adu Gyamfi, R., Agyin Birikorang, S., Tindjina, I., Ahmed, S. M., Twumasi, A. D., Avornyo,~~
715 ~~V. K., & Singh, U. (2019). One Time Fertilizer Briquettes Application for Maize~~
716 ~~Production in Savanna Agroecologies of Ghana. *Agronomy Journal*, 111(6), 3339-3350.~~
717 ~~<https://doi.org/10.2134/agronj2019.04.0292>~~
- 718 Adu, G. B., Abdulai, M. S., Alidu, H., Nustugah, S. K., Buah, S. S., Kombiok, J. M., Obeng-Antwi,
719 K., Abudulai, M., & Etwire, P. M. (2014). Recommended Production Practices for Maize
720 in Ghana. <https://doi.org/10.13140/2.1.4376.3527>
- 721 Adzawla, W., Atakora, W. K., Gouzaye, A., & Bindraban, P. S. (2021). *Crop Yield and Fertilizer*
722 *Use Among Farmers in Guinea Savannah and Transitional Zones of Ghana.*
723 [https://ifdc.org/wp-content/uploads/2021/07/IFDC-FERARI-Research-Report-No-5-](https://ifdc.org/wp-content/uploads/2021/07/IFDC-FERARI-Research-Report-No-5-Final.pdf)
724 [Final.pdf](https://ifdc.org/wp-content/uploads/2021/07/IFDC-FERARI-Research-Report-No-5-Final.pdf)
- 725 Adzawla, W., Atakora, W. K., Kissiedu, I. N., Martey, E., Etwire, P. M., Gouzaye, A., &
726 Bindraban, P. S. (2021). Characterization of farmers and the effect of fertilization on
727 maize yields in the Guinea Savannah, Sudan Savannah, and Transitional agroecological
728 zones of Ghana. *EFB Bioeconomy Journal*, 1.
729 <https://doi.org/10.1016/j.bioeco.2021.100019>
- 730 Afreh, D. N., Afari, M. A. B., Adjei, R. R., Sarfo Boateng, A., Santo, K. G., Abdulai, M., & Popovic,
731 V. (2022). Response of Two Maize (*Zea mays* L.) Varieties to Times of NPK (15-15-15)

732 Fertilizer Application. *International Journal of Agronomy*, 2022, 1-7.
733 <https://doi.org/10.1155/2022/7186913>

734 Akolgo, G. A., Kemausuor, F., Awafo, E. A., Amankwah, E., Atta-Darkwa, T., Essandoh, E. O.,
735 Bart-Plange, A., & Maia, C. M. B. d. F. (2020). Biochar as a Soil Amendment Tool: Effects
736 on Soil Properties and Yield of Maize and Cabbage in Brong-Ahafo Region, Ghana. *Open*
737 *Journal of Soil Science*, 10(03), 91-108. <https://doi.org/10.4236/ojss.2020.103005>

738 Antwi, A., Duker, A., Fosu, M., & Abaidoo, R. C. (2017). Simulation of major soil nutrients
739 requirement for maize production using the QUEFTS model in the Northern region of
740 Ghana. *Direct Research Journal of Agriculture and Food Science*, 5(3), 133-140, Article
741 DRJA43071532. <http://directresearchpublisher.org/aboutjournal/drjafs>

742 Antwi, E. K., Boakye-Danquah, J., Asabere, S. B., Yiran, G. A. B., Loh, S. K., Awere, K. G.,
743 Abagale, F. K., Asubonteng, K. O., Attua, E. M., & Owusu, A. B. (2014). Land use and
744 landscape structural changes in the ecoregions of Ghana. *Journal of Disaster Research*, 9(4),
745 452-467.

746 Antwi, M., Duker, A. A., Fosu, M., Abaidoo, R. C., & Pirasteh, S. (2016). Geospatial approach to
747 study the spatial distribution of major soil nutrients in the Northern region of Ghana.
748 *Cogent Geoscience*, 2(1), 1201906. <https://doi.org/10.1080/23312041.2016.1201906>

749 Atiah, W. A., Amekudzi, L. K., Akum, R. A., Quansah, E., Antwi- Agyei, P., & Danuor, S. K.
750 (2021). Climate variability and impacts on maize (*Zea mays*) yield in Ghana, West Africa.
751 *Quarterly Journal of the Royal Meteorological Society*. <https://doi.org/10.1002/qj.4199>

752 Badu, E., Kaba, J. S., Abunyewa, A. A., Dawoe, E. K., Agbenyega, O., & Barnes, R. V. (2019).
753 Biochar and inorganic nitrogen fertilizer effects on maize (*Zea mays* L.) nitrogen use and
754 yield in moist semi-deciduous forest zone of Ghana. *Journal of Plant Nutrition*, 42(19),
755 2407-2422. <https://doi.org/10.1080/01904167.2019.1659347>

756 Baffour-Ata, F., Antwi-Agyei, P., Nkiaka, E., Dougill, A. J., Anning, A. K., & Kwakye, S. O.
757 (2021). Effect of climate variability on yields of selected staple food crops in northern
758 Ghana. *Journal of Agriculture and Food Research*, 6.
759 <https://doi.org/10.1016/j.jafr.2021.100205>

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

765 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using
766 lme4. *Journal of Statistical Software*, 67(1), 1 - 48. <https://doi.org/10.18637/jss.v067.i01>

767 Bationo, A., Fening, J. O., & Kwaw, A. (2018). Assessment of Soil Fertility Status and Integrated
768 Soil Fertility Management in Ghana. In A. Bationo, et al. (Eds.), *Improving the Profitability,*
769 *Sustainability and Efficiency of Nutrients Through Site Specific Fertilizer Recommendations in*
770 *West Africa Agro-Ecosystems* (Vol. 1, pp. 93-138). Springer International Publishing AG
771 2018. https://doi.org/10.1007/978-3-319-58789-9_7

772 Batjes, N. H. (2011). *Global Distribution of Soil Phosphorus Retention Potential* (ISRIC Report
773 2011/06, 42 pp, Issue. <https://library.wur.nl/WebQuery/wurpubs/fulltext/185289>

774 Bawa, A., & Tang, Y. (2021). Yield and Growth Response of Maize (*Zea mays* L.) to Varietal and
775 Nitrogen Application in the Guinea Savanna Agro-Ecology of Ghana. *Advances in*
776 *Agriculture*, 2021, 1-8. <https://doi.org/10.1155/2021/1765251>

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

Field Code Changed

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

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

784 Boullouz, M., Bindraban, P. S., Kissiedu, I. N., Kouame, A. K. K., Devkota, K. P., & Atakora, W.
785 K. (2022). An integrative approach based on crop modeling and geospatial and statistical
786 analysis to quantify and explain the maize (*Zea mays*) yield gap in Ghana. *Front. Soil Sci.*
787 2:1037222. <https://doi.org/10.3389/fsoil.2022.1037222>

788 Braimoh, A. K., & Vlek, P. L. G. (2006). Soil quality and other factors influencing maize yield in
789 northern Ghana. *Soil use and management*, 22(2), 165-171. <https://doi.org/10.1111/j.1475-2743.2006.00032.x>

790 Breure, M. S., Kempen, B., & Hoffland, E. (2022). Spatial predictions of maize yields using
791 QUEFTS - A comparison of methods. *Geoderma*, 425.
792 <https://doi.org/10.1016/j.geoderma.2022.116018>

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

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

800 Cicin-Sain, I. (2018). *Population growth and food: A systems perspective*. Retrieved 01 July 2021 from
801 <https://sustainablefoodtrust.org/articles/population-growth-and-food-a-systems-perspective/>

802 Cooper, P. J. M. (1979). The association between altitude, environmental variables, maize growth
803 and yields in Kenya. *The Journal of Agricultural Science*, 93(3), 635-649. <https://doi.org/10.1017/S0021859600039058>.

804 Cudjoe, G. P., Antwi-Agyei, P., & Gyampoh, B. A. (2021). The Effect of Climate Variability on
805 Maize Production in the Ejura-Sekyedumase Municipality, Ghana. *Climate*, 9(10).
806 <https://doi.org/10.3390/cli9100145>

807 Daniel, K. A. N., Alice, A. A., & Thomas, A.-G. (2021). Response of maize (*Zea mays* L.) to foliar
808 and soil applied fertilizers in the Semi-deciduous forest zone of Ghana. *African Journal of*
809 *Agricultural Research*, 17(8), 1114-1122. <https://doi.org/10.5897/AJAR2021.15561>

810 Dapaah, H. K., Ennin, S. A., & Asafu-Agyei, J. N. (2008). Combining inorganic fertilizer with
811 poultry manure for sustainable production of quality protein maize in Ghana. *Ghana*
812 *Journal of Agricultural Science*, 41(1). <https://doi.org/10.4314/gjas.v41i1.46144>

813 Dargie, S., Girma, T., Chibsa, T., Kassa, S., Boke, S., Abera, A., Haileselassie, B., Addisie, S.,
814 Amsalu, S., Haileselassie, M., Soboka, S., Abera, W., & Weldesemayat, S. G. (2022).
815 Balanced fertilization increases wheat yield response on different soils and
816 agroecological zones in Ethiopia. *Experimental Agriculture*, 58.
817 <https://doi.org/10.1017/s0014479722000151>

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

Formatted: Dutch (Netherlands)

Field Code Changed

- 825 Debtanu, M., Das, D. K., & Pathak, H. (2006). Simulation of fertilizer requirement for irrigated
826 wheat in eastern India using the QUEFTS model. *Archives of Agronomy and Soil Science*,
827 52(4), 403-418. <https://doi.org/10.1100/tsw.2006.43>
- 828 Diao, X., Hazell, P., Kolavalli, S., & Resnick, D. (2019). Ghana's economic and agricultural
829 transformation: Past performance and future prospects. 277.
830 <https://doi.org/10.1093/oso/9780198845348.001.0001>
- 831 Durodola, O. S., & Mourad, K. A. (2020). Modelling Maize Yield and Water Requirements under
832 Different Climate Change Scenarios. *Climate*, 8(11). <https://doi.org/10.3390/cli8110127>
- 833 Essel, B., Abaidoo, R. C., Opoku, A., & Ewusi-Mensah, N. (2020). Economically Optimal Rate for
834 Nutrient Application to Maize in the Semi-deciduous Forest Zone of Ghana. *J Soil Sci
835 Plant Nutr*, 20(4), 1703-1713. <https://doi.org/10.1007/s42729-020-00240-y>
- 836 Ezui, K. S., Franke, A. C., Ahiabor, B. D. K., Tetteh, F. M., Sogbedji, J., Janssen, B. H., Mando, A.,
837 & Giller, K. E. (2017). Understanding cassava yield response to soil and fertilizer nutrient
838 supply in West Africa. *Plant and Soil*, 420(1-2), 331-347. <https://doi.org/10.1007/s11104-017-3387-6>
- 840 FAO/OECD. (2018). *Food Security and Nutrition: Challenges for Agriculture and the Hidden Potential
841 of Soil*. FAO. <http://www.fao.org/3/CA0917EN/ca0917en.pdf>
- 842 Farr, T. G., & Kobrick, M. (2000). Shuttle Radar Topography Mission produces a wealth of data.
843 *Eos Trans. AGU*, 81:583-583.
- 844 Fening, J. O., Yeboah, E., Gyapong, T. A., & Gaizie, E. (2009). On farm evaluation of the
845 contribution of three green manures to maize yield in the semi-deciduous forest zone of
846 Ghana. *African Journal of Environmental Science Technology*, 3(9).
847 <https://www.ajol.info/index.php/ajest/article/view/46071>
- 848 Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1- km spatial resolution climate surfaces
849 for global land areas. *International Journal of Climatology*, 37(12), 4302-4315.
850 <https://doi.org/10.1002/joc.5086>
- 851 Florence, A., Ransom, M., & Mengel, D. (2017). Potassium Fixation by Oxidized and Reduced
852 Forms of Phyllosilicates. *Soil Science Society of America Journal*, 81(5), 1247-1255.
853 <https://doi.org/10.2136/sssaj2016.12.0420>
- 854 Fosu-Mensah, B. Y., Manchadi, A., & Vlek, P. L. G. (2019). Impacts of climate change and climate
855 variability on maize yield under rainfed conditions in the sub-humid zone of Ghana: A
856 scenario analysis using APSIM. *West African Journal of Applied Ecology*, 27(1), 108-126.
857 <https://www.ajol.info/index.php/wajae/article/view/189216>
- 858 Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression*, Third edition. Sage,
859 Thousand Oaks CA. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- 860 Frimpong, K. A., Phares, C. A., Boateng, I., Abban-Baidoo, E., & Apuri, L. (2021). One-time
861 application of biochar influenced crop yield across three cropping cycles on tropical
862 sandy loam soil in Ghana. *Heliyon*, 7(2), e06267.
863 <https://doi.org/10.1016/j.heliyon.2021.e06267>
- 864 Ghansah, B., Forkuo, E. K., Osei, E. F., Appoh, R. K., Asare, M. Y., & Kluste, N. A. B. (2018).
865 Mapping the spatial distribution of small reservoirs in the White Volta Sub-basin of
866 Ghana. *Remote Sensing Applications: Society and Environment*, 9, 107-115.
867 <https://doi.org/10.1016/j.rsase.2017.12.003>
- 868 Giller, K. E., Franke, A. C., Abaidoo, R., Bajjukya, F., Bala, A., Boahen, S., Dashiell, K.,
869 Kantengwa, S., Sanginga, J.-M., Sanginga, N., Simmons, A., & Turner, A. (2013).
870 N2Africa: putting nitrogen fixation to work for smallholder farmers in Africa. In *Agro-*

Field Code Changed

871 ecological intensification of agricultural systems in the African highlands (pp. 176-194).
872 Routledge. <https://doi.org/10.4324/9780203114742>
873 Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J.,
874 Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food security: the challenge of feeding
875 9 billion people. *Sci. (80-.)*, 327(5967), 812-818. <https://doi.org/10.1126/science.1185383>
876 Graves, S., Piepho, H.-P., & Selzer, L. (2019). *Visualizations of Paired Comparisons: Package*
877 *'multcompView'*. <https://cran.r-project.org/web/packages/multcompView/index.html>
878 Guilpart, N., Grassini, P., van Wart, J., Yang, H., van Ittersum, M. K., van Bussel, L. G. J., Wolf,
879 J., Claessens, L., Leenaars, J. G. B., & Cassman, K. G. (2017). Rooting for food security in
880 Sub-Saharan Africa. *Environmental Research Letters*, 12(11), 1-8.
881 <https://doi.org/10.1088/1748-9326/aa9003>
882 GYGA. (2021). *Agriculture in Ghana*. University of Nebraska-Lincoln and Wageningen
883 University & Research. Retrieved 24th March 2021 from
884 <https://www.yieldgap.org/ghana>
885 Han, S., & Kim, H. (2021). Optimal Feature Set Size in Random Forest Regression. *Applied*
886 *Sciences*, 11(8). <https://doi.org/10.3390/app11083428>
887 Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., Sila, A.,
888 MacMillan, R. A., Mendes de Jesus, J., Tamene, L., & Tondoh, J. E. (2015). Mapping Soil
889 Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current
890 Predictions. *PLoS One*, 10(6), 1-26. <https://doi.org/10.1371/journal.pone.0125814>
891 IFDC. (2012). *Ghana Fertilizer Assessment*. International Fertilizer Development Center Muscle
892 Shoals Alabama 35662 USA.
893 [http://ghana.countrystat.org/fileadmin/user_upload/countrystat_fenix/congo/docs](http://ghana.countrystat.org/fileadmin/user_upload/countrystat_fenix/congo/docs/Ghana%20Fertilizer%20Need%20Assessment%20for%20USAID%20FtF%20AFAPIFD)
894 [/Ghana%20Fertilizer%20Need%20Assessment%20for%20USAID%20FtF%20AFAPIFD](http://ghana.countrystat.org/fileadmin/user_upload/countrystat_fenix/congo/docs/Ghana%20Fertilizer%20Need%20Assessment%20for%20USAID%20FtF%20AFAPIFD)
895 [C%202012.pdf](http://ghana.countrystat.org/fileadmin/user_upload/countrystat_fenix/congo/docs/Ghana%20Fertilizer%20Need%20Assessment%20for%20USAID%20FtF%20AFAPIFD)
896 ISRIC. *Africa SoilGrids nutrients*. ISRIC. Retrieved 12 Apr. 2021 from
897 <https://data.isric.org/geonetwork/srv/eng/catalog.search#/search>
898 Janssen, B. H., Guiking, F. C. T., van der Eijk, D., Smaling, E. M. A., Wolf, J., & van Reuler, H.
899 (1990). A system for quantitative evaluation of the fertility of tropical soils (QUEFTS).
900 *Geoderma*, 46, 299-318. [https://doi.org/10.1016/0016-7061\(90\)90021-Z](https://doi.org/10.1016/0016-7061(90)90021-Z)
901 Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim,
902 K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random Forests for Global and
903 Regional Crop Yield Predictions. *PLoS One*, 11(6), e0156571.
904 <https://doi.org/10.1371/journal.pone.0156571>
905 Jiang, P., & Thelen, K. D. (2004). Effect of soil and topographic properties on crop yield in a
906 North- Central corn-soybean cropping system. 96(1), 252-258.
907 <https://doi.org/10.2134/agronj2004.0252>
908 Kanton, R. A. L., Prasad, P. V. V., Mohammed, A. M., Bidzakin, J. K., Ansoba, E. Y., Asungre, P.
909 A., Lamini, S., Mahama, G., Kusi, F., & Sugri, I. (2016). Organic and Inorganic Fertilizer
910 Effects on the Growth and Yield of Maize in a Dry Agro-Ecology in Northern Ghana.
911 *Journal of Crop Improvement*, 30(1), 1-16. <https://doi.org/10.1080/15427528.2015.1085939>
912 Kihara, J., Nziguheba, G., Zingore, S., Coulibaly, A., Esilaba, A., Kabambe, V., Njoroge, S., Palm,
913 C., & Huisling, J. (2016). Understanding variability in crop response to fertilizer and
914 amendments in sub-Saharan Africa. *Agric. Ecosyst. Environ.*, 229, 1-12.
915 <https://doi.org/10.1016/j.agee.2016.05.012>
916 Kpotor, P., Akromah, R., Ewool, M. B., Kena, A. W., Owusu-Adjei, E., & Tuffour, H. O. (2014).
917 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)

- 918 Genotypes under Different Levels of Nitrogen Fertilization across Two Agro-Ecological
 919 Zones in Ghana. *International Journal of Scientific Research in Agricultural Sciences*, 1(7),
 920 128-141. <https://doi.org/10.12983/ijstras-2014-p0128-0141>
- 921 Kranjac-Berisavljevic, G., Bayorbor, T. B., Abdulai, A. S., Obeng, F., Blench, R. M., Turton, C. N.,
 922 Boyd, C., & Drake, E. (1999). Rethinking natural resource degradation in semi-arid Sub-
 923 Saharan Africa: the case of semi-arid Ghana. *University for Development Studies, Tamale*
 924 *ODI, London*.
- 925 Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for
 926 hydrological model assessment. *Adv. Geosci.*(5), 89-97. [https://doi.org/10.5194/adgeo-](https://doi.org/10.5194/adgeo-5-89-2005)
 927 [5-89-2005](https://doi.org/10.5194/adgeo-5-89-2005)
- 928 Kravchenko, A. N., & Robertson, G. P. (2007). Can Topographical and Yield Data Substantially
 929 Improve Total Soil Carbon Mapping by Regression Kriging ? *Agronomy Journal*(99), 12-
 930 17. <https://doi.org/10.2134/agronj2005.0251>
- 931 Kugbe, J. X., Kombat, R., Atakora, W., & Tejada Moral, M. (2019). Secondary and micronutrient
 932 inclusion in fertilizer formulation impact on maize growth and yield across northern
 933 Ghana. *Cogent Food & Agriculture*, 5(1). <https://doi.org/10.1080/23311932.2019.1700030>
- 934 [Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. \(2017\). lmerTest Package: Tests in
 935 Linear Mixed Effects Models. *Journal of Statistical Software*, 82\(13\), 1-23.
 936 <https://doi.org/10.18637/jss.v082.i13>](https://doi.org/10.18637/jss.v082.i13)
- 937 Kyei-Mensah, C., Kyerematen, R., & Adu-Acheampong, S. (2019). Impact of Rainfall Variability
 938 on Crop Production within the Worobong Ecological Area of Fanteakwa District, Ghana.
 939 *Advances in Agriculture*, 2019, 1-7. <https://doi.org/10.1155/2019/7930127>
- 940 Lal, R. (2006). Enhancing crop yields in the developing countries through restoration of the soil
 941 organic carbon pool in agricultural lands. *Land degradation development*, 17(2), 197-209.
 942 <https://doi.org/10.1002/ldr.696>
- 943 Lamos-Díaz, H., Puentes-Garzón, D. E., & Zarate-Caicedo, D. A. (2020). Comparison Between
 944 Machine Learning Models for Yield Forecast in Cocoa Crops in Santander, Colombia.
 945 *Revista Facultad de Ingeniería*, 29(54), e10853-e10853.
 946 <https://doi.org/10.19053/01211129.v29.n54.2020.10853>
- 947 Leenaars, J. G. B., Claessens, L., Heuvelink, G. B. M., Hengl, T., Ruiperez Gonzalez, M., van
 948 Bussel, L. G. J., Guilpart, N., Yang, H., & Cassman, K. G. (2018). Mapping rootable depth
 949 and root zone plant-available water holding capacity of the soil of sub-Saharan Africa.
 950 *Geoderma*, 324, 18-36. <https://doi.org/10.1016/j.geoderma.2018.02.046>
- 951 Logah, V., Ewusi-Mensah, N., & Tetteh, F. K. M. (2011). Soil Organic Carbon and Crop Yield
 952 under Different Soil Amendments and Cropping Systems in the Semi-deciduous Forest
 953 Zone of Ghana. *Journal of Plant Sciences*, 6(4), 165-173.
 954 <https://doi.org/10.3923/jps.2011.165.173>
- 955 MacCarthy, D. S., Vlek, P. L. G., & Fosu-Mensah, B. Y. (2012). The Response of Maize to N
 956 Fertilization in a Sub-Humid Region of Ghana: Understanding the Process Using a Crop
 957 Simulation Model. In J. Kihara, et al. (Eds.), *Improving Soil Fertility Recommendations in*
 958 *Africa using the Decision Support System for Agrotechnology Transfer (DSSAT)*. Springer
 959 Science + Business Media. https://doi.org/10.1007/978-94-007-2960-5_5
- 960 Marusteri, M., & Bacarea, V. (2010). Comparing groups for statistical differences: how to choose
 961 the right statistical test? *Biochimica medica*, 20(1), 15-32. <https://hrcak.srce.hr/file/73801>
- 962 McHugh, M. L. (2011). Multiple comparison analysis testing in ANOVA. *Biochimica medica*, 21(3),
 963 203-209. [https://www.biochimica-](https://www.biochimica-medica.com/en/journal/21/3/10.11613/BM.2011.029)
 964 [medica.com/en/journal/21/3/10.11613/BM.2011.029](https://www.biochimica-medica.com/en/journal/21/3/10.11613/BM.2011.029)

Field Code Changed

Field Code Changed

- 965 Mobilian, C., & Craft, C. B. (2021). Wetland Soils: Physical and Chemical Properties and
 966 Biogeochemical Processes. *Reference Module in Earth Systems and Environmental Sciences*,
 967 Elsevier. <https://doi.org/10.1016/B978-0-12-819166-8.00049-9>
- 968 Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., & Foley, J. A. (2012).
 969 Closing yield gaps through nutrient and water management. *Nature*, 490(7419), 254-257.
 970 <https://doi.org/10.1038/nature11420>
- 971 Mugwe, J., Mugendi, D., Kungu, J., & Muna, M. M. (2009). Maize Yields Response to Application
 972 of Organic and Inorganic Input under on-Station and on-Farm Experiments in Central
 973 Kenya. *Experimental Agriculture*, 45(1), 47-59.
 974 <https://doi.org/10.1017/S0014479708007084>
- 975 Mulder, I. (2000). *Soil Fertility: QUEFTS and Farmer's Perceptions* (Vol. Working Paper No 30).
 976 IIED.
 977 https://www.researchgate.net/publication/228856865_Soil_fertility_QUEFTS_and_farmers%27_perceptions
- 979 Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and
 980 intra-class correlation coefficient from generalized linear mixed-effects models revisited
 981 and expanded. *Journal of the Royal Society Interface*, 14(134), 20170213.
 982 <https://doi.org/10.1098/rsif.2017.0213>
- 983 Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non- Gaussian data: a
 984 practical guide for biologists. *Biological Reviews*, 85(4), 935-956.
 985 <https://doi.org/10.1111/j.1469-185x.2010.00141.x>
- 986 Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A
 987 discussion of principles. *Journal of hydrology*, 10(3), 282-290.
 988 https://en.wikipedia.org/wiki/Nash-Sutcliffe_model_efficiency_coefficient
- 989 Nevavuori, P., Narra, N., Linna, P., & Lipping, T. (2020). Crop Yield Prediction Using
 990 Multitemporal UAV Data and Spatio-Temporal Deep Learning Models. *Remote Sensing*,
 991 12(23), 4000. <https://www.mdpi.com/2072-4292/12/23/4000>
- 992 Njoroge, K. S. (2019). *Explaining variability in maize yield responses to nutrient applications in
 993 smallholder farms of western Kenya* (Publication Number 978-94-6395-164-7) Wageningen
 994 University & Research]. The Netherlands. <https://doi.org/10.18174/503185>
- 995 Njoroge, R., Otinga, A. N., Okalebo, J. R., Pepela, M., & Merckx, R. (2018). Maize (*Zea mays* L.)
 996 Response to Secondary and Micronutrients for Profitable N, P and K Fertilizer Use in
 997 Poorly Responsive Soils. *Agronomy*, 8(4). <https://doi.org/10.3390/agronomy8040049>
- 998 Nkrumah, F., Klutse, N. A. B., Adukpo, D. C., Owusu, K., & Quagraine, K. A. (2014). Rainfall
 999 variability over Ghana: model versus rain gauge observation. *International Journal of
 1000 Geosciences*, 5(7). <https://doi.org/10.4236/ijg.2014.57060>
- 1001 Nyuor, B. A., Donkor, E., Aidoo, R., Buah, S. S., Naab, J., Nutsugah, S., Bayala, J., & Zougmore,
 1002 R. (2016). Economic Impacts of Climate Change on Cereal Production: Implications for
 1003 Sustainable Agriculture in Northern Ghana. *Sustainability*, 8(8).
 1004 <https://doi.org/10.3390/su8080724>
- 1005 O'Halloran, I. P., Kachanoski, R. G., & Stewart, J. W. B. (1985). Spatial variability of soil
 1006 phosphorus as influenced by soil texture and management. *Canadian journal of soil science*,
 1007 65(3), 475-487. <https://doi.org/10.4141/cjss85-051>
- 1008 Obi, M. E., & Ebo, P. O. (1995). The effects of organic and inorganic amendments on soil physical
 1009 properties and maize production in a severely degraded sandy soil in southern Nigeria.
 1010 *Bioresource Technology*, 51(2-3), 117-123. [https://doi.org/10.1016/0960-8524\(94\)00103-8](https://doi.org/10.1016/0960-8524(94)00103-8)

- 1011 Onduru, D. D., & Du Preez, C. C. (2007). Spatial and temporal aspects of agricultural
 1012 sustainability in the semi-arid tropics: a case study in Mbeere district, Eastern Kenya.
 1013 *Tropical Science*, 47(3), 134-148. <https://doi.org/10.1002/ts.207>
- 1014 Osei, B. A. (1995). Effects of different lime application rates and time on some chemical
 1015 properties of an acid soil in Ghana. *Soil use and management*, 11, 25-29.
 1016 <https://doi.org/10.1111/j.1475-2743.1995.tb00491.x>
- 1017 Ovalles, F. A., & Collins, M. E. (1986). Soil- landscape relationships and soil variability in north
 1018 central Florida. *Soil Science Society of America Journal*, 50(2), 401-408.
 1019 <https://doi.org/10.2136/sssaj1986.03615995005000020029x>
- 1020 Owoade, F. M., Adiku, S. G. K., Atkinson, C. J., & MacCarthy, D. S. (2021). Differential Impact
 1021 of Land Use Types on Soil Productivity Components in Two Agro-ecological Zones of
 1022 Southern Ghana. In *African Handbook of Climate Change Adaptation* (pp. 1721-1733).
 1023 https://doi.org/10.1007/978-3-030-45106-6_144
- 1024 Owoade, F. M., Adiku, S. G. K., Atkinson, C. J., MacCarthy, D. S., Kumahor, S. K., & Kolawole,
 1025 G. O. (2020). Location and Land use effects on Soil Carbon Accretion and Productivity in
 1026 the Coastal Savanna Agro-ecological Zone of Ghana. *West African Journal of Applied*
 1027 *Ecology*, 28(2), 1-13.
 1028 [https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahU](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKewijgM_d2cD5AhVNLewKHfCfCYgQFnoECAIQAQ&url=https%3A%2F%2Fwww.a)
 1029 www.a
 1030 [jol.info%2Findex.php%2Fwajae%2Farticle%2Fview%2F202673%2F191158&usg=AOvVa](https://www.a)
 1031 [w1WdXQq6X2x3K-dU3roFeYD](https://www.a)
- 1032 Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylaniadis, C., & Athanasiadis, I. N.
 1033 (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, 187.
 1034 <https://doi.org/10.1016/j.agry.2020.103016>
- 1035 Peprah, K. (2012). Rainfall and temperature correlation with crop yield: The case of Asunafo
 1036 forest, Ghana. www.ijsr.net
- 1037 Peterson, R. A. (2021). Finding Optimal Normalizing Transformations via bestNormalize. *The R*
 1038 *Journal*, 13(1), 310-329. <https://doi.org/10.32614/RJ-2021-041>
- 1039 Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for
 1040 random forest. *WIREs Data Mining and Knowledge Discovery*, 9(3).
 1041 <https://doi.org/10.1002/widm.1301>
- 1042 Quansah, G. W. (2010). *Effect of organic and inorganic fertilizers and their combinations on the growth*
 1043 *and yield of maize in the semi-deciduous forest zone of Ghana* [Msc, Kwame Nkrouma
 1044 University of Science and Technology]. Kumsi, Ghana.
 1045 <http://csirspace.csirgh.com/handle/123456789/2379>
- 1046 Rahman, A. N., Larbi, A., Kotu, B., Asante, M. O., Akakpo, D. B., Mellon- Bedi, S., & Hoeschle-
 1047 Zeledon, I. (2021). Maize-legume strip cropping effect on productivity, income, and
 1048 income risk of farmers in northern Ghana. *Agronomy Journal*, 1-12.
 1049 <https://doi.org/10.1002/agj2.20536>
- 1050 Ren, T., Zou, J., Wang, Y., Li, X. K., Cong, R. H., & Lu, J. W. (2015). Estimating nutrient
 1051 requirements for winter oilseed rape based on QUEFTS analysis. *Journal of Agricultural*
 1052 *Science*, 154(3), 425-437. <https://doi.org/10.1017/S0021859615000301>
- 1053 RStudio Team. (2022). RStudio: Integrated Development for R. . *RStudio* (2021.09.2), PBC, Boston,
 1054 MA URL <http://www.rstudio.com/>.
- 1055 Sadras, V. O., & Calvino, P. A. (2001). Quantification of grain yield response to soil depth in
 1056 soybean, maize, sunflower, and wheat. *Agronomy Journal*, 93(3), 577-583.
 1057 <https://doi.org/10.2134/agronj2001.933577x>

Field Code Changed

Formatted: Dutch (Netherlands)

Field Code Changed

Field Code Changed

- 1058 [Sallah, P. Y. K., Twumasi Afriyie, S., Ahenkora, K., Asiedu, E. A., Obeng Antwi, E., Osei-](#)
1059 [Yeboah, S., Frimpong Manso, P. P., Ankomah, A., & Dzah, B. D. \(2007\). Agronomic](#)
1060 [potentials of quality protein maize hybrids developed in Ghana. *Journal of Agricultural*](#)
1061 [Science and Applications, 40\(1\), 81-89. <https://doi.org/10.4314/gjas.v40i1.2157>](#)
- 1062 [Sattaria, S. Z., van Ittersuma, M. K., Bouwmanb, A. F., Smitd, A. L., & Janssen, B. H. \(2014\).](#)
1063 [Crop yield response to soil fertility and N, P, K inputs in different environments: Testing](#)
1064 [and improving the QUEFTS model. *Field Crops Research, 157*, 35-46.](#)
1065 [<https://doi.org/10.1016/j.fcr.2013.12.005>](#)
- 1066 Sawyer, J. E., & Mallarino, A. P. (1999). Differentiating and understanding the Mehlich 3, Bray,
1067 and Olsen soil phosphorus tests. *Agronomy Conference Proceedings and Presentations, 12*.
1068 [<https://dr.lib.iastate.edu/handle/20.500.12876/4349>](#)
- 1069 Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. John Wiley & Sons.
1070 [<https://doi.org/10.1002/9780470316856>](#)
- 1071 Shehu, B. M., Lawan, B. A., Jibrin, J. M., Kamara, A. Y., Mohammed, I. B., Rurinda, J., Zingore,
1072 S., Craufurd, P., Vanlauwe, B., Adam, A. M., & Merckx, R. (2019). Balanced nutrient
1073 requirements for maize in the Northern Nigerian Savanna: Parameterization and
1074 validation of QUEFTS model. *Field Crops Res, 241*, 107585.
1075 [<https://doi.org/10.1016/j.fcr.2019.107585>](#)
- 1076 Sileshi, G. W., Kihara, J., Tamene, L., Vanlauwe, B., Phiri, E., & Jama, B. (2022). Unravelling
1077 causes of poor crop response to applied N and P fertilizers on African soils. *Experimental*
1078 *Agriculture, 1-17*. [<https://doi.org/10.1017/S0014479721000247>](#)
- 1079 Solomon, D., Lehmann, J., Fraser, J. A., Leach, M., Amanor, K., Frausin, V., Kristiansen, S. M.,
1080 Millimouno, D., & Fairhead, J. (2016). Indigenous African soil enrichment as a climate-
1081 smart sustainable agriculture alternative. *Frontiers in Ecology the Environment, 14(2)*, 71-
1082 76. [<https://doi.org/10.1002/fee.1226>](#)
- 1083 Spilke, J., Piepho, H. P., & Hu, X. (2005). Analysis of unbalanced data by mixed linear models
1084 using the MIXED procedure of the SAS system. *Journal of Agronomy and Crop Science,*
1085 *191(1)*, 47-54. [<https://doi.org/10.1111/j.1439-037X.2004.00120.x>](#)
- 1086 SRID/MoFA. (2021). *Agriculture in Ghana: Facts and Figures (2020)*.
1087 [\[https://srid.mofa.gov.gh/sites/default/files/Agriculture%20In%20Ghana%20Facts%20%26%20Figures_%202020%20FINAL.pdf\]\(https://srid.mofa.gov.gh/sites/default/files/Agriculture%20In%20Ghana%20Facts%20%26%20Figures_%202020%20FINAL.pdf\)](#)
- 1088
- 1089 Tabi, F. O., Diels, J., Ogunkunle, A. O., Iwuofor, E. N. O., Vanlauwe, B., & Sanginga, N. (2007).
1090 Potential nutrient supply, nutrient utilization efficiencies, fertilizer recovery rates and
1091 maize yield in northern Nigeria. *Nutrient Cycling in Agroecosystems, 80(2)*, 161-172.
1092 [<https://doi.org/10.1007/s10705-007-9129-z>](#)
- 1093 Tetteh, F., Larbi, A., Nketia, K. A., Senayah, J. K., Hoeschle-Zeledon, I., & Abdulrahman, N.
1094 (2016). *Suitability of soils for cereal cropping in Northern Ghana*. I. I. o. T. Agriculture.
1095 [\[https://www.researchgate.net/publication/335581532_Suitability_of_soils_for_cereal_cropping_in_northern_Ghana/link/5d6e6bbf299bf16522f2b9b6/download\]\(https://www.researchgate.net/publication/335581532_Suitability_of_soils_for_cereal_cropping_in_northern_Ghana/link/5d6e6bbf299bf16522f2b9b6/download\)](#)
- 1096
- 1097 Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable
1098 intensification of agriculture. *Proc Natl Acad Sci U S A, 108(50)*, 20260-20264.
1099 [<https://doi.org/10.1073/pnas.1116437108>](#)
- 1100 Timsina, J., Dutta, S., Devkota, K. P., Chakraborty, S., Neupane, R. K., Bishta, S., Amgain, L. P.,
1101 Singh, V. K., Islam, S., & Majumdar, K. (2021). Improved nutrient management in cereals
1102 using Nutrient Expert and machine learning tools: Productivity, profitability and
1103 nutrient use efficiency. *Agricultural Systems, 192*.
1104 [<https://doi.org/10.1016/j.agsy.2021.103181>](#)

Formatted: Dutch (Netherlands)

Formatted: English (United States)

Formatted: English (United States)

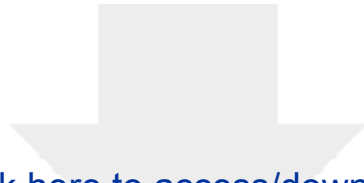
Formatted: English (United States)

- 1105 Tittonell, P., Corbeels, M., van Wijk, M. T., Vanlauwe, B., & Giller, K. E. (2008). Combining
 1106 Organic and Mineral Fertilizers for Integrated Soil Fertility Management in Smallholder
 1107 Farming Systems of Kenya: Explorations Using the Crop-Soil Model FIELD. *Agronomy*
 1108 *Journal*, 100(5), 1511-1526. <https://doi.org/10.2134/agronj2007.0355>
- 1109 USAID. (2022). *Agriculture and food security*. USAID. Retrieved 18/02/2022 from
 1110 <https://www.usaid.gov/ghana/agriculture-and-food-security>
- 1111 ~~USAID/IFDC. (2015). Seed Guide Recommended Commercial Maize, Rice and Soybean~~
 1112 ~~Varieties Available for Northern Ghana. In U. S. A. f. I. Development (Ed.), (pp. 28): Feed~~
 1113 ~~the Future Ghana Agriculture Technology Transfer Project.~~
- 1114 Uzoma, K. C., Inoue, M., Andry, H., Fujimaki, H., Zahoor, A., & Nishihara, E. (2011). Effect of
 1115 cow manure biochar on maize productivity under sandy soil condition. *Soil use and*
 1116 *management*, 27(2), 205-212. <https://doi.org/10.1111/j.1475-2743.2011.00340.x>
- 1117 van Loon, M. P., Adjei-Nsiah, S., Descheemaeker, K., Akotsen-Mensah, C., van Dijk, M., Morley,
 1118 T., van Ittersum, M. K., & Reidsma, P. (2019). Can yield variability be explained?
 1119 Integrated assessment of maize yield gaps across smallholders in Ghana. *Field Crops*
 1120 *Research*, 236, 132-144. <https://doi.org/10.1016/j.fcr.2019.03.022>
- 1121 Wallach, D., Makowski, D., Jones, J. W., & Brun, F. (2018). *Working with dynamic crop models*.
 1122 Academic Press-Elsevier. <https://doi.org/10.1016/C2016-0-01552-8>
- 1123 Wang, L. (2021). Data Driven Explanation of Temporal and Spatial Variability of Maize Yield in
 1124 the United States. *Front Plant Sci*, 12, 701192. <https://doi.org/10.3389/fpls.2021.701192>
- 1125 ~~Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.~~
 1126 ~~Retrieved from <https://ggplot2.tidyverse.org>.~~
- 1127 Wijayanto, Y., & Prastyanto, E. (2012). A study of using QUEFTS model for establishing site
 1128 specific fertilizer recommendation in maize on the basis of farmer fields. *AGRIVITA,*
 1129 *Journal of Agricultural Science*, 33(3), 273-278.
 1130 [https://media.neliti.com/media/publications/53171-EN-a-study-of-using-quefts-](https://media.neliti.com/media/publications/53171-EN-a-study-of-using-quefts-model-for-establ.pdf)
 1131 [model-for-establ.pdf](https://media.neliti.com/media/publications/53171-EN-a-study-of-using-quefts-model-for-establ.pdf)
- 1132 Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high
 1133 dimensional data in C++ and R. *Journal of Statistical Software*, 77(1), 1-17.
 1134 <https://doi.org/10.18637/jss.v077.i01>
- 1135 Xu, X., He, P., Pampolino, M. F., Chuan, L., Johnston, A. M., Qiu, S., Zhao, S., & Zhou, W. (2013).
 1136 Nutrient requirements for maize in China based on QUEFTS analysis. *Field Crops*
 1137 *Research*, 150, 115-125. <https://doi.org/10.1016/j.fcr.2013.06.006>
- 1138 Yawson, D. O., Kwakye, P. K., Armah, F. A., & Frimpong, K. A. (2011). The dynamics of
 1139 potassium (K) in representative soil series of Ghana. *ARP Journal of Agricultural and*
 1140 *Biological Science*, 6(1), 48-55.
 1141 [https://www.researchgate.net/publication/260793689_The_Dynamics_of_Potassium_](https://www.researchgate.net/publication/260793689_The_Dynamics_of_Potassium_K_in_Representative_Soil_Series_of_Ghana)
 1142 [K_in_Representative_Soil_Series_of_Ghana](https://www.researchgate.net/publication/260793689_The_Dynamics_of_Potassium_K_in_Representative_Soil_Series_of_Ghana)
- 1143 Yeboah, E., Asamoah, G., Kofi, B., & Abunyewa, A. A. (2016). Effect of Biochar Type and Rate of
 1144 Application on Maize Yield Indices and Water Use Efficiency on an Ultisol in Ghana.
 1145 *Energy Procedia*, 93, 14-18. <https://doi.org/10.1016/j.egypro.2016.07.143>
- 1146 Zingore, S., Delve, R. J., Nyamangara, J., & Giller, K. E. (2007). Multiple benefits of manure: The
 1147 key to maintenance of soil fertility and restoration of depleted sandy soils on African
 1148 smallholder farms. *Nutrient Cycling in Agroecosystems*, 80(3), 267-282.
 1149 <https://doi.org/10.1007/s10705-007-9142-2>

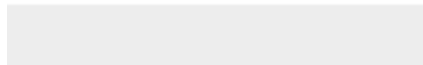
Field Code Changed

Formatted: Space After: 8 pt

1150 Zingore, S., Mutegi, J., Agesa, B., Tamene, L., & Kihara, J. (2015). Soil degradation in sub-Saharan
1151 Africa and crop production options for soil rehabilitation. *Better Crop.*, 99(1), 24-26.
1152 <https://cgspace.cgiar.org/handle/10568/68702>
1153

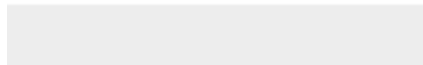


Click here to access/download
Supplementary Material
Appendix A (Supplementary Table).docx





Click here to access/download
Supplementary Material
Appendix B (Supplementary Figure).docx





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,

A handwritten signature in black ink, appearing to read "Prem Bindraban".

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