FERARI

FERTILIZER RESEARCH & RESPONSIBLE IMPLEMENTATION

Evaluation and Geospatial Analysis of Variability in Maize Yield Response to Fertilizer (NPK) Using Modeling in Ghana

IFDC FERARI Research Report No. 7

K.K.A. Kouame, P.S. Bindraban, I.N. Kissiedu, K. El Mejahed, and W.K. Atakora

IFDC Developing Agriculture from the Ground Up

MOHAMMED VI POLYTECHNIC UNIVERSITY

ESAFE ÉCOLE DES SCIENCES DE L'AGRICULTURE, DE LA FERTILISATION ET DE L'ENVIRONNEMENT

Evaluation and Geospatial Analysis of Variability in Maize Yield Response to Fertilizer (NPK) Using Modeling in Ghana

K.K.A. Kouame[1,2], P.S. Bindraban[2*], I.N. Kissiedu[2], K. El Mejahed[1], and W.K. Atakora[2]

[1] *Mohammed VI Polytechnic University (UM6P), Ben Guerir, Morocco*
[2] *International Fertilizer Development Center (IFDC), USA*
**Correspondence: pbindraban@ifdc.org*

This report is based on the thesis by Kouame K.K.A., submitted to the Faculty of the School of Agriculture, Fertilization and Environmental Sciences (ESAFE) for obtaining the diploma of master's degree in "Fertilizer Science and Technology". This thesis was conducted under the Fertilizer Research and Responsible Implementation (FERARI) program.

Citation:

Kouame, K.K.A., P.S. Bindraban, I.N. Kissiedu, K. El Mejahed, and W.K. Atakora. 2021. *Evaluation and Geospatial Analysis of Variability in Maize Yield Response to Fertilizer (NPK) Using Modeling in Ghana*. IFDC FERARI Research Report No. 7.

ACKNOWLEDGMENTS

Acknowledgments by Kouakou K.A. Kouame from his thesis.

We would like to express our gratitude to the people without whom we could not have completed this internship. We thank them for the warm welcome they gave us, the support they gave us, and their availability. This study was made possible with support from the Fertilizer Research and Responsible Implementation (FERARI) program. I am highly indebted for the technical and financial support offered to me under this project. Sincere thanks go to:

- Prof. Abdallah Oukarroum, Director of School of Agricultural, Fertilizer and Environmental Sciences (ESAFE), and all his staff.
- Prof. Aziz Yasri, former Director of ESAFE and current General Secretary of INRA Morocco.
- Dr. Prem S. Bindraban, FERARI Program Director, our IFDC supervisor, for his valuable technical and design advice, I enjoyed our working sessions, sir.
- Mr. Isaac N. Kissiedu, Database Administrator and GIS Specialist, IFDC, for his precious assistance in geostatistical spatial analysis and mapping.
- Prof. Khalil El Mejahed, our UM6P co-supervisor, for his guidance and advice.
- Dr. Williams K. Atakora and Ms. Victorine Goly for the favorable environment created to accomplish our work.
- My father, the late Kouame Yao, my mother, Kouadio Akissi, my brothers and sisters, and my fiancée for their efforts and their commitment at my side during these years of study.

All those from far and near have been instrumental in the development of this document.

Thank you all.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

Adj. R^2	Adjusted Coefficient of Determination
AEZs	Agroecological Zones
AIC	Akaike Information Criterion
AYO	Average Yield Observed
AYP	Average Yield Predicted
C	Carbon
Ca	Calcium
Corg	Organic Carbon
CV	Coefficient of Variation
CS	Coastal Savannah
DSSAT	Decision Support System for Agrotechnology Transfer
DYPM	Digital Yield Prediction Map
Elev	Elevation
FST	Forest-Savannah Transitional
GDP	Gross Domestic Product
GS	Guinea Savannah
K	Potassium
KExch	Exchangeable Potassium
M	Million
MoFA	Ministry of Food and Agriculture
Mg	Magnesium
ML	Machine Learning
MLR	Multiple Linear Regression
N	Nitrogen
Norg	Organic Nitrogen
NUE	Nutrient Use Efficiency
OLSR	Ordinary Least Squares Regression
P	Phosphorus
POlsen	Phosphorus Olsen
Pt	Total Phosphorus
QUEFTS	Quantitative Evaluation of Fertility of Tropical Soils
R^2	Coefficient of Determination
RF	Random Forest
RFsp	Random Forest for spatial predictions framework
RMSE	Root Means Square Error
SDF	Semi-Deciduous Forest
SSA	Sub-Saharan Africa
SS	Sudan Savannah

SUMMARY

Maize is the main cereal crop in Ghana. However, yields are very low (around 1-1.5 mt/ha), and despite the increase in fertilizer application per hectare (21-22 kg/ha), there are large differences in yields in on-farm and on-station trials.

Maize production is hampered by several biotic and abiotic factors that negatively impact its yield response to fertilizer application. Therefore, we sought to understand why, despite fertilizer application, maize yields do not increase consistently over space and time and what major factors explain this variability.

To answer this question, we chose a yield-modeling approach based on yield data from on-farm and on-station trials. Quantitative Evaluation of Fertility of Tropical Soils (QUEFTS) and Multiple Linear Regression-Akaike Information Criterion (MLR-AIC) models were used to evaluate observed yield variability, while random forest for spatial predictions framework modeling was used for geospatial analysis and mapping of yield predicted.

The QUEFTS model cannot significantly explain yield variability at the station and farm level (R^2=12% and R^2=24.6%, respectively). MLR showed that soil physical properties explained more of the yield variability (R^2=24%) at the station level than environmental parameters (R^2=8%), with chemical soil properties explaining the highest fraction (R^2=41%). At the farm level, environmental covariates (R^2=26%) explained more variability in yield response than physical (R^2=21%) and soil chemical (R^2=16%) variables. Detailed regression analysis revealed that high temperature and high rainfall combined with shallow rooting depth (<50 cm) were determinants that reduced the effectiveness of fertilizer application.

Understanding the yield variability observed in Ghana for better fertilizer recommendations must be done comprehensively because yield variability is the result of the interaction and combination of several covariates. Other covariates, such as management, pest and diseases, and solar radiation, must be considered in further modeling analysis.

CHAPTER 1: INTRODUCTION

1.1 Background

The increase in population growth and consumption means that the global demand for food will continue to increase for at least 50 years (Cicin-Sain, 2018; EU, 2019), and the effects of climate change are not helping matters. Population growth and shifting diet imply that 60-70% more food must be produced by 2050. Agricultural production in sub-Saharan Africa (SSA) must double or even triple to meet that estimated food demand (Godfray et al., 2010; Rahman et al., 2021).

Scientists have begun to speak about soil fertility as a natural resource at risk of depletion. According to Sanchez et al. (1997), an average of 660 kg N/ha, 75 kg P/ha, and 450 kg K/ha have been used by crops and not replaced during the last 30 years from about 200 million (M) ha of cultivated land in 37 African countries, including Ghana. Bationo et al. (2018) reported that soil nutrient depletion rates are projected as 35 kg N, 4 kg P, and 20 kg K per hectare, and the extent of nutrient depletion is widespread in all of Ghana's agroecological zones (AEZs), with N and P being the most deficient nutrients (Zingore et al., 2015). As a result, the yields obtained by farmers are far below the attainable yield, crop production is limited, and food security is in jeopardy.

It is increasingly understood that crop response to fertilizer in many areas of Africa, including Ghana, is depressed by a variety of factors, including soil degradation problems and many others. Furthermore, variability in climatic conditions (rainfall and temperature) is considered another one of the major challenges to agricultural production other than soil fertility issues (Kyei-Mensah et al., 2019). For instance, rainfall variability has been reported to affect production crops, increase crop disease incidents, and cause drastic reductions in soil fertility (Thornton et al., 2009; Kashaigili et al., 2014; Leng and Huang, 2017). High variability of climatic parameters (rainfall, temperature) causes uncertainties in agricultural productivity, with profound impacts on the ecology, economy, and people's welfare (Onduru and Du Preez, 2007). According to Tetteh et al. (2014), the impacts of climate change in Ghana are expected to worsen soon, especially if nothing is done to mitigate its effects.

Low yield is partly caused by soil variability and varying topographic features of the field (Jiang and Thelen, 2004; Kravchenko and Robertson, 2007). Field topography can have a direct influence on crop growth and yield by redirecting and changing soil water availability and an indirect effect through its influence on the distribution of certain soil chemical and physical properties, such as organic matter content, base saturation, soil temperature, and particle size distribution (Stone et al., 1985; Kravchenko and Robertson, 2007). In addition to the growth-defining factors, yield is also influenced by root zone depth (Sadras and Calvino, 2001; Guilpart et al., 2017; Leenaars et al., 2018a). According to Tetteh et al. (2016), soils in Ghana are suitable for cereal crop production but most sites are marginal, with shallow soil depth as the major limitation. For instance, 52% of soils in the three northern regions of Ghana have a soil depth <50 cm. In addition, a study on quantification of grain yield response to soil depth in soybean, maize, sunflower, and wheat revealed that the harvest index was most affected by shallow soil in the maize plot (Sadras and Calvino, 2001).

Despite current low crop productivity, Ghana has a large potential to intensify production and significantly close current yield gaps of major cereals (Bationo et al., 2018; van Loon et al., 2019), since it was estimated that, on average, 20% of maize yield potential is achieved across Ghana (GYGA, 2021). For example, only addressing nutrient deficiencies by applying fertilizer would help to reduce yield gaps to 50% of attainable yields (Mueller et al., 2012).

1.2 Problem Statement

Despite calls to ramp up efforts to design and implement fertilizer programs that recognize the spatial variability of soil fertility and climatic conditions, Ghana's current fertilizer recommendations tend to be very general, assuming uniform soil fertility and geographic conditions throughout the country (Chapoto and Tetteh, 2014). According to Abunyewa and Mercer-Quarshie (2004) and Bationo et al. (2018), maize grain yield rarely exceeds 1 mt/ha in farmers' fields. Despite N, P, and K compound fertilizer applied, they observed low yield in maize farms. The increase in fertilizer use, however, has not lead to substantial increases in crop productivity. Subsequently, Bua et al. (2020) did a study on 1,684 yields and fertilizer data points from legacy and peer-reviewed publications. Yield responses of maize to fertilizer (organic, inorganic) application in Ghana were explored, and findings showed that grain yield responses to fertilization are highly variable across AEZs. Some locations showed a significant grain yield response (8 mt/ha) to fertilizers, while other areas had a small grain yield response (0.5 mt/ha). This variability in maize response to different fertilizer rate applications depresses farmers' incentives and ability to purchase fertilizers in subsequent seasons (Njoroge, 2019). Hence, the question arises of why the response of maize yield fluctuates so greatly even though the rate of NPK applied per hectare in Ghana has grown rapidly, from 8 kg/ha in 2016 to 21 kg/ha in 2020 (MoFA, 2020; AfricaFertilizer.org, 2021)?

It is acknowledged that solutions to such challenges lie in sustainable agricultural production systems (Onduru and Du Preez, 2007) via improved nutrient use efficiency (NUE) with integrated soil fertility management (Sanginga and Woomer, 2009; Vanlauwe et al., 2015; Mugwe et al., 2019), crop management, and other inputs. However, to produce sustainably, it is crucial to assess factors that affect maize production and their impact on yield response to native soil fertility and applied fertilizers. Several of these soil factors are beyond farmers' control. In addition, the climatic variability of rainfall, evaporation, solar radiation, temperature, relative humidity, and wind is also beyond farmers' command. Each of these parameters has a spatial and temporal impact that influences maize yield response.

An essential way to capture spatio-temporal variability to identify site-specific factors that determine yield and the assessment of crop requirements to reduce the yield gap can be done by model-based approaches (Silva and Giller, 2021). Crop production models can be characterized as empirical and mechanistic (process-oriented) models (Sattaria et al., 2014). Decision support approaches, such as the Quantitative Evaluation of Fertility of Tropical Soils (QUEFTS) model, have been advocated by several studies to estimate field-specific N, P, and K recommendations (Tabi et al., 2007; Tittonell et al., 2008a; Wijayanto and Prastyanto, 2012; Xu et al., 2013; Ren et al., 2015). QUEFTS could be used to explain yield variability concerning soil chemical properties (pH, organic carbon, available and total phosphorus, and organic nitrogen) (Onduru and Du Preez, 2007; Njoroge, 2019). QUEFTS assumes all other production factors are optimal and does not consider soil physical properties, climatic variables, or crop-specific characteristics. A single modeling approach, such as QUEFTS, cannot be expected to disentangle the highly variable and complex smallholder farming systems, which calls for caution in drawing oversimplified conclusions. According to Dutta et al. (2020) and Timsina et al. (2021), for investigating multiple interactions among the outcomes, the explanatory variables often demand adaptive and non-parametric multivariate analyses only, due to their ability to negotiate non-linear relationships. Thus, here, multiple linear regression (MLR) and machine learning (ML) have been added to further analyze yield variability. In this study, maize yield response variability will be assessed using data from on-farm and on-station trials through these approaches.

1.3 Activity Statement

The study of variability in yield response of maize to fertilizer began by pre-processing and diagnosing yield data from articles and covariates to obtain ordered and scientifically sound data. Then, the ordered yield data and covariates were used to estimate yield using different analytical approaches, including QUEFTS, MLR, and ML). The observed and predicted yields were confronted using Ordinary Least Squares Regression (OLSR) to analyze the error between average model prediction and ground-truth to determine how much models help us to explain maize yield response variance in Ghana. Finally, the observed and predicted yields were used to digitize the maps using advanced geostatic techniques.

1.4 Hypotheses

- H_0-The QUEFTS model can significantly explain the yield response of maize to fertilizer.
- H_0-MLR statistics can identify factors that significantly affect the yield response of maize to fertilization.
- H_0-Digital yield prediction maps (DYPMs) can be used to arrive at site-specific fertilizer recommendations.

1.5 Forecasting Statement

The objective of this study is to understand and explain the spatial variability of maize yield response components using the selected approach. Specifically, it will:

- Assess maize yield response to chemical soil properties and NPK fertilizer.
- Assess maize yield response to physical soil properties, landscape elements, and climatic variables.
- Produce yield maps to arrive at site-specific fertilizer recommendations.

1.6 Justification Statement

Mineral (NPK) and organic fertilizers have played and will continue to play a crucial role in increasing maize yield in Ghana. However, the application of organic or mineral fertilizers or both in combination in farmers' plots and the associated low yield responses and high variability in time and place deserve in-depth analyses. To be useful, fertilizer must be applied in a site-specific manner and must be aligned with edaphic and climatic conditions. To do this, it is imperative to capture, understand, and explain the spatial, temporal, and distributional variation in yield responses of maize to fertilizer. This assessment will help in making farm decisions and fertilizer recommendations that are appropriate and specific to the crop site in the agroclimatic zone.

CHAPTER 2: LITERATURE REVIEW

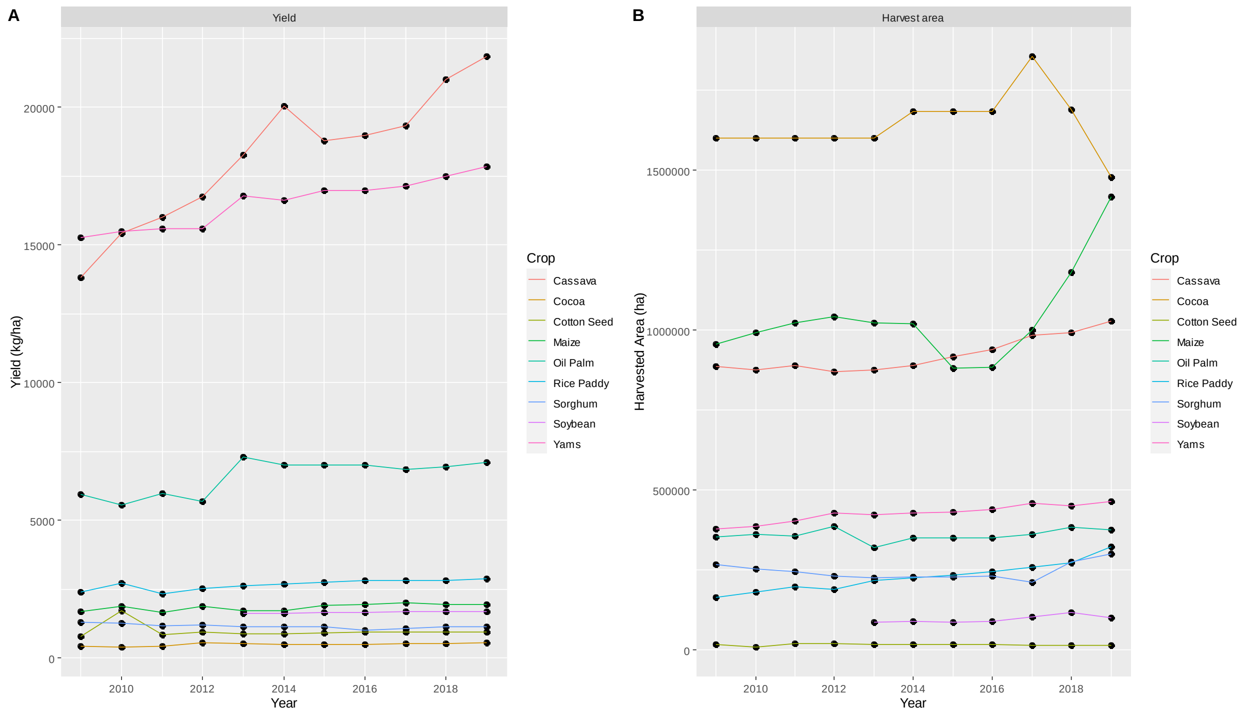
2.1 Agriculture in Ghana

Agriculture plays a notable role in the economies of most developing countries in SSA (FAO/OECD, 2018). In Ghana, it has been the principal sector for the development and growth of the economy for several decades (Diao et al., 2019). About 40% of Ghana's labor force is engaged in agriculture, which contributes 14.3% to the country's gross domestic product (GDP) (Addo and Amponsah, 2018; FAO, 2019), even though this contribution to real GDP growth has been declining for the past five years (World Bank, 2018). According to Shereen et al. (2019) and Ghana Statistical Services (GSS, 2019), cocoa export alone generates about U.S. $2.71 billion in foreign exchange. It makes up about 20-25% of total export receipts, provides about two-thirds of cocoa farmers' incomes, supports the livelihoods of approximately 4 million farming households, and reduces poverty, especially in southern Ghana, due to the overall agricultural growth (Darfour and Rosentrater, 2016a).

Agriculture in Ghana is predominantly on a smallholder basis (less than 2 ha) using rudimentary technology to produce about 80% of the country's total agricultural output. Agriculture occupies 56% of the country's total land area of 23,884,245 ha (SRID/MoFA, 2011). About 45.4% of the households in Ghana are agricultural households, 73.3% are rural households, and 26.7% are urban and large plantations with rubber, oil palm, coconut, and, to a lesser extent, rice, maize, and pineapples (Mohammed et al., 2013; Darfour and Rosentrater, 2016a; Bua et al., 2020).

Generally, 51% of Ghana's cereal needs are locally produced and less than 30% of agro-based industry raw materials are locally produced (Darfour and Rosentrater, 2016a); Ghana's major export crops are cocoa, oil palm, and cotton. Several challenges undermine Ghana's agriculture sector, according to Banson et al. (2014) and Darfour and Rosentrater (2016a). Diversity in agroecologies, a lack of human resources and managerial skills, policy constraints, poor management of natural resources, and inadequate technological development are all challenges that the Ghanaian authorities must address for sustainable agricultural development and food security.

Ghana's agriculture sector is characterized by low yields for both staple and cash crops (World Bank, 2017). The average cocoa yield in Ghana, estimated at 400-450 kg/ha (Figure 2-1A), is among the lowest in the world. In the Forest zone, cocoa, oil palm, coffee, and rubber are of particular importance. The food crops in this zone are mainly inter-cropped mixtures of maize, plantain, cocoyam, and cassava. Rice is important throughout the country, but cassava and cocoa cover the largest cultivated area (Figure 2-1B).



Source: FAOSTAT/FAO (2021).

Figure 2-1. Annual trend of (A) yield and (B) harvested area of major crops

Ghana's agriculture has recorded a generally increasing growth; the agricultural GDP increased from 2.3% in 2015 to 4.8% in 2018 (MoFA, 2020). Food price inflation decreased from 9.7 in 2016 to 7.2 in 2019, with a positive impact on overall inflation. The global food security index, which considers affordability, availability, and quality of food across 113 countries, placed Ghana in the 59th position in 2019, up from the 79th position in 2018.

Maize importation has decreased in these three last years, from 982,044 M mt in 2016, down to 830,127 M mt in 2018 (MoFA, 2020). The decreasing imports demonstrate the ambition of the Government of Ghana to become self-sufficient in production of maize and other cereals, including rice, sorghum, millet, and soybean. Overall, Ghana depends on the import of wheat, rice, and maize by 100%, 61%, and 3%, respectively.

In Ghana, 50% of the population depends on rainfed crops (SRID/MoFA, 2017). According to Nkrumah et al. (2014), the northern part of the region receives 150-250 mm of rainfall per month in the peak months of the wet season (July to September) and the southern part has two wet seasons: the major season from March to July and a minor season from September to November. The majority of West African rainfall comes from the West African Monsoon, which is controlled by the movement of the Intertropical Convergence Zone, an area where the southeast and northeast trade winds meet and a belt of convective clouds is formed due to this convergence and the high amount of energy from the sun. The movement of the Intertropical Convergence Zone results in high variability in rainfall (Israelsson et al., 2020). Mitigation strategies are challenging to implement in Ghana due to this complex spatial climate variability with the coexistence of different rainfall regimes, from a bimodal wet coastal forest to a dry savanna region in the north (Nkrumah et al., 2014). The variability negatively affects agricultural production, which threatens the existence of its smallholders.

2.2 Generality on the Maize Crop in Ghana

2.2.1 Ecology of Maize

Maize, *Zea mays ssp. mays* (2n = 20), is a monoecious plant cultivated as an annual plant but behaves under certain conditions like a biennial plant. It reproduces by fertilization cross (allogamous), unlike most other cereals. Carbon fixation in maize is affected by the C4 photosynthetic pathway. Maize differs from other cereals in that it forms its grains not in terminal ears or panicles but axillary cobs. Mature cobs can be harvested 3-5 months after sowing. The main maize-producing areas in Ghana are Eastern, Ashanti, and Brong-Ahafo regions, which account for over 80% of the country's total maize production (Darfour and Rosentrater, 2016b). The three northern regions (Northern, Upper East, and Upper West) supply the rest. Maize is produced in annual single-crop systems in the higher rainfall area in the Southern Forest zone and annual double-crop systems in the Forest-Savannah Transitional zone (GYGA, 2021). Typical double-crop systems in this zone include maize-maize, maize-cowpea, and groundnut-maize (GYGA, 2021). In the three northern regions, sorghum and millet are often intercropped with cowpea and/or maize, and in the Southern Forest zone, maize is often intercropped with one or more other crops such as cassava, cocoyam, and plantain. Maize adapts well to different soil types with a pH range of 5.0-7.0. High yields are obtained from maize planted on deep, fine-structured, well-aerated, well-drained loamy soils that are rich in organic matter.

2.2.2 Maize Production in Ghana

Zea mays L. ranks first as Ghana's most important cereal produced and consumed (MiDA, 2010; MoFA, 2012, 2020; FAOStat, 2021), occupying an area of about 1.5 M ha and constituting about 50-60% of Ghana's cereal production; it contributes 3.3% to total agricultural production value. Post-harvest losses between 5% and 70% have been reported, according to Darfour and Rosentrater (2016b). Maize farming systems differ across AEZs with their unique characteristics in terms of the available technology sets, environmental effects, and management practices.

The bimodal rainfall pattern in the Tropical Rainforest, Semi-Deciduous Forest, and Forest-Savannah Transitional zones results in major and minor cropping seasons. The Guinea Savannah and the Sudan Savannah zones are characterized by a single growing/raining season from July to September (Asante et al., 2019). Consequently, major maize-producing zones are the Semi-Deciduous Forest, Forest-Savannah Transitional, and part of the Guinea Savannah.

2.3 Nutrient Use in Maize Cropping

The fertilizer recommendation for maize was updated in 1974, and since then, only sporadic and inconclusive attempts have been made to update the recommendation (Tetteh et al., 2017). The fertilizer recommendation for maize was 64-38-38 kg/ha N-P_2O_5-K_2O according to Adu et al. (2014). Safo (1990) reports that 67-45-45 kg/ha of N-P_2O_5-K_2O was the Ministry of Food and Agriculture (MoFA)-recommended rate. These fertilizer recommendations were updated to 90-60-60 kg/ha N-P_2O_5-K_2O (Tetteh et al., 2008). The main subsidized fertilizers used in maize production are NPK (15:15:15) and urea (46:0:0). Average fertilizer use in 2020 was about 20 kg/ha (MoFA, 2020), which is slightly higher than the SSA average of about 10 kg/ha, and in Ghana, less than half of maize farmers apply fertilizer, although in the north, high adoption by up to 87% of farmers is reported (Chapoto and Ragasa, 2013).

2.4 Model Use in Maize Cropping

Maize production is at the center of a *perfect storm* that encompasses the grand challenges of ensuring cereal food security in Ghana in the face of climate change, soil degradation, and water scarcity while preventing further conversion of forests to agricultural land. The role of crop models and crop modeling in addressing these grand challenges is very important (Bindraban et al., 1999; Bindraban et al., 2000; Silva and Giller, 2021). To further support this claim, at iCropM2020 Hammer (2020)[1] and Giller (2020)[2] stated that "*If you don't understand it, you can't model it; if you don't model it, you can't understand it*" and "*We learn the most when the models don't work*" to emphasize the role of crop models in generating and testing research hypotheses and finding solutions to the challenges faced by farmers. Several crop models have been published. Initiated by modeling at Wageningen University by late Professor Cornelis Teunis de Wit (de Wit and Goudriaan, 1974) in the late 1960s with several models emerging, including BACROS and WOFOST (Van Diepen et al., 1989), many other institutions helped pursue these production concepts to further develop models, including the CERES-Maize model of the Decision Support System for Agrotechnology Transfer (DSSAT) suite (Jones, 1993; Jones et al., 1998), Agricultural Productions Systems simulator, or APSIM (McCown et al., 1996; Keating et al., 2003), AquaCROP (Hsiao et al., 2009; Raes et al., 2009; Steduto et al., 2009), and CropSyst (Stockle et al., 1994), among others. Of these, the DSSAT suite and APSIM are the most widely used in Ghana to estimate crop yields under varied soil, weather, and management conditions (Masika, 2016). The use of a crop model for yield estimation starts with calibration, followed by validation and finally application. For example, crop models have been used to assess the maize yield gap (Masika, 2016), maize variability and yield gap in Ghana AEZs (MacCarthy et al., 2017), and corn production and the role of fertilizer (Scheiterle and Birner, 2018). Crop models have also been found to be indispensable in evaluating and selecting the most promising options for fertilizer recommendations (Atakora et al., 2014; Antwi et al., 2017) for the best maize nutrition, thus reducing the yield gap. All of this research done with crop models does not concretely explain the yield response to fertilizer variability at the Ghanaian scale. And even when it has been done, it is only for a few regions and rarely mapped after the variability is explained. Thus, further analysis, taking into account a wide range of explanatory variables, is needed to explain and map the variability in observed maize yield responses to fertilizer across Ghana.

The tradition of using the models in Ghana is thus becoming more established. However, the QUEFTS model, originally designed and calibrated for maize in tropical regions, is used less because it is not widely cited in Ghana's agricultural literature. Articles on QUEFTS are mainly focused on its use for fertilizer recommendations, not as a model for assessing maize yield variability. The relationship between nutrient (NPK) uptake and yield, and the balance between these nutrients, is the basis of the QUEFTS model (Janssen and Guiking, 1990). The majority of studies in SSA in which QUEFTS was used show a fairly high coefficient of determination ($71\%<R^2<92\%$ [Antwi et al., 2017], $77\%<R^2<87\%$ [Ezui et al., 2017], $84\%<R^2<82\%$ [Jamada, 2019; Sattaria et al., 2014], $69\%<R^2<75\%$ [Shehu et al., 2019], $66\%<R^2<78\%$ [Smaling and Janssen, 1993], $83\%<R^2<84\%$ [Tabi et al., 2007; Tittonell et al., 2008b; Wijayanto and Prastyanto, 2012]) between predicted and observed yields.

Various studies around the world have shown that application of machine learning techniques, such as linear mixed-effects, classification and regression trees, and random forest (RF), can be useful in determining and prioritizing the relative importance of factors that contribute to yields and yield variability (Jeong et al., 2016; Lamos-Díaz et al., 2020; Nevavuori et al., 2020; Paudel et al., 2021; Timsina et al., 2021). However, according to Hengl et al. (2018), the modeling

[1] Hammer, G. 2020. "On the Nature of Crop Models (and Modelers) Needed to Advance Crop Adaptation and Improvement," Keynote to the Second International Crop Modelling Symposium, Montpellier, France.
[2] Giller, K.E. 2020. "Grand Challenges for the 21st Century: What Crop Models Can and Can't (Yet) Do," Keynote to the Second International Crop Modelling Symposium, Montpellier, France.

relationship between observed yield, as in our case study, with covariates and spatial autocorrelation jointly using machine learning techniques is relatively novel and not entirely worked out. In Ghana, few studies have used complex tools such as the above to quantify, assess, and explain persistent variability in grain crops, particularly the response of maize crop yield to fertilizer. Approaches such as machine learning model algorithms, however, require large datasets that may not be readily available, disallowing their use.

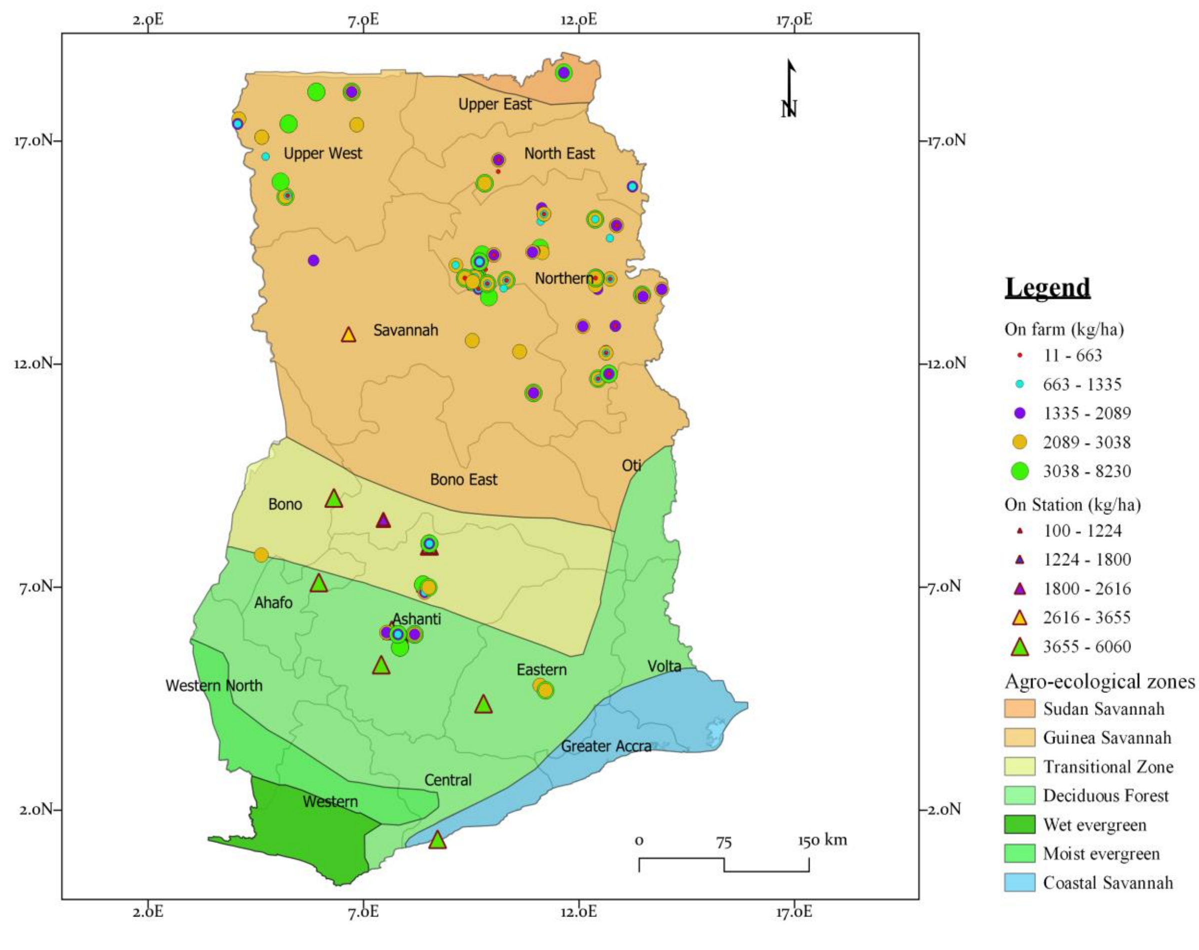
CHAPTER 3: DATA AND METHODOLOGY

3.1 Study Area

The study area covered almost the entire country of Ghana, which is located on the west coast of Africa. Ghana shares boundaries with Burkina Faso, Gulf of Guinea, Togo, and Côte d'Ivoire to the North, South, East, and West, respectively. The country covers a total land area of 238,539 km² and lies between longitude 2.0° E and 17.0 E° and latitude 1.0° N and 18.0° N (Figure 3-1). Administratively, the country is divided into 16 regions, and these regions comprise six agroecological zones (AEZs): Sudan Savannah (SS), Guinea Savannah (GS), Forest-Savannah Transitional (FST), Semi-Deciduous Forest (SDF), Rainforest (moist and wet evergreen), and Coastal Savannah (CS). Experiments were implemented in all AEZs where legacy data of maize yield were gathered.

3.2 Data Sources

Maize yields were compiled from peer-reviewed publications and various field and station trials between 1990 and 2017. The yield data concerned 1,818 sites from maize production under farmer trials and research trials assessing *maize nitrogen grain yield responses*, *maize nitrogen* * *phosphorus grain yield responses*, *NPK grain yield responses*, *fertilizer grain yield responses in Ghana*, *response of maize to fertilizer applications in Ghana*, *fertilizer yield*



Source: Author.

***Figure 3-1.** Map of Ghana showing AEZs and geographical distribution of experimental sites* *responses in Ghana*, and *fertilizer trials in Ghana*; for a more elaborate explanation of the trials, refer to Bua et al. (2020). Six hundred and twenty sites (farmer trials and research trials) with missing geographical coordinates and their trial years, unquantified NPK fertilizer amounts, and micronutrient fertilizer treatments (sulfur, zinc, boron, magnesium) were deleted. After cleaning the data, a total of 1,198 trial sites with their respective maize yields could be used for the analysis.

3.2.1 Covariates

In statistics, a covariate is a variable that is possibly predictive of the outcome under study. The direct effect of fertilizer on maize yields can be analyzed straightforwardly, but a significant part of the unexplained variation may be related to a secondary variable indirectly influencing yield response. In addition to fertilizer application, crop yields are also influenced by local climate, land characteristics, levels of inputs other than fertilizers, pests and diseases, and other management practices applied to the field.

3.2.2 Soil and Climate Data Collection

Soil properties are from ISRIC Africa SoilGrids (a collection of gridded soil property maps) at a resolution of 1 km² (Hengl et al., 2015; Hengl et al., 2017; Leenaars et al., 2018a). Rainfall and temperature datasets are from spatially interpolated monthly climate data for global land areas at a very high spatial resolution (approximately 1 km²) (Harris et al., 2014) and downscaled with WorldClim 2.1 (Fick and Hijmans, 2017). Temperature data were calculated using covariates including mean MODIS cloud cover (Cld), distance to oceanic coast (cdist), elevation (Elev), MODIS daytime land surface temperature (Tmax), nighttime land surface temperature (Tmin), and average nighttime and daytime land surface temperatures (Tmean) (Fick and Hijmans, 2017). Regarding elevation, topography data are from SRTM with a resolution of 1 arc-second (30 meters) .[3]

Table 3-1. Soil and climate data used as input to QUEFTS-R and spatial modeling

Covariate	Map	Source of Map
Soil H_2O pH	pH	pHH2O_M_agg30cm_AF_1km.tif[1a]
Soil organic carbon (g/kg)	Corg	OC_M_agg30cm_AF_1km.tif[b]
Soil total nitrogen (g/kg)	Norg	N_M_agg30cm_AF_1km.tif[(b]
Soil available phosphorus (mg/kg)	Pav	P_M_agg30cm_AF_1km.tif[b]
Soil total phosphorus (mg/kg)	Ptot	P.T_M_agg30cm_AF_1km.tif[b]
Soil exchangeable potassium (mmolc/kg)	KExch	K_M_agg30cm_AF_1km.tif[b)]
Soil root zone depth	Rzd	gyga_af_erzd__m_1km.tif[b]
Precipitation (mm)	prec	wc2.1_2.5m_prec.tif[c]
Maximum temperature (°)	Tmax	wc2.1_2.5m_tmax.tif[c]
Minimum temperature (°)	Tmin	wc2.1_2.5m_tmin.tif[c]
Digital elevation (m)	Elev	SRTM 1 Arc-Second Global.tif[d]

a. Hengl et al. (2015); b. Hengl et al. (2017); c. Fick and Hijmans (2017); and d. Download – CGIAR-CSI SRTM.

[3] Download – CGIAR-CSI SRTM.

The map of soil available P (in ppm) represents data according to POlsen (ppm) that were converted from the source map, which represents data according to PMehlich3 (100 ppm). The conversions are given by Eq. 3-1, based on rules derived from data presented by Sawyer and Mallarino (1999) and suggested by de Pater (2015), according to Leenaars et al. (2018b). The data presented by Sawyer and Mallarino (1999) are visualized in Figure 3-2.

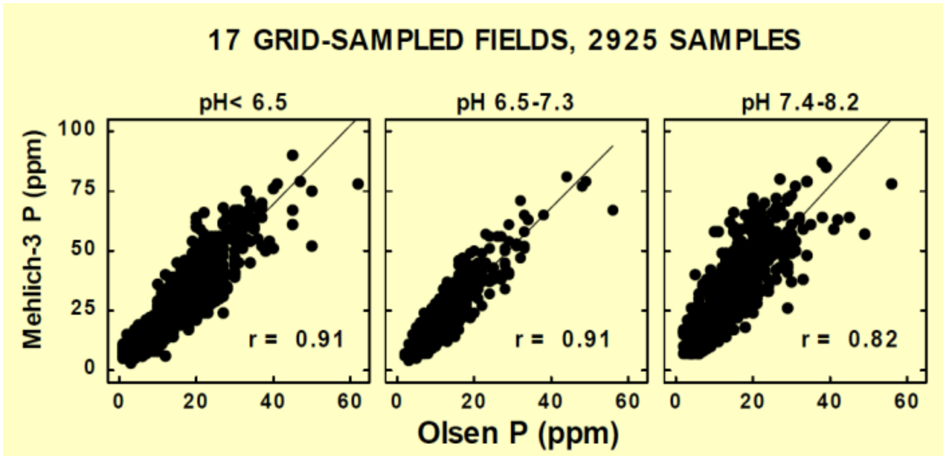


Figure 3-2. ***Relationships between soil P contents measured according to Olsen and according to Mehlich3, differentiated according to soil pH (Sawyer and Mallarino, 1999)***

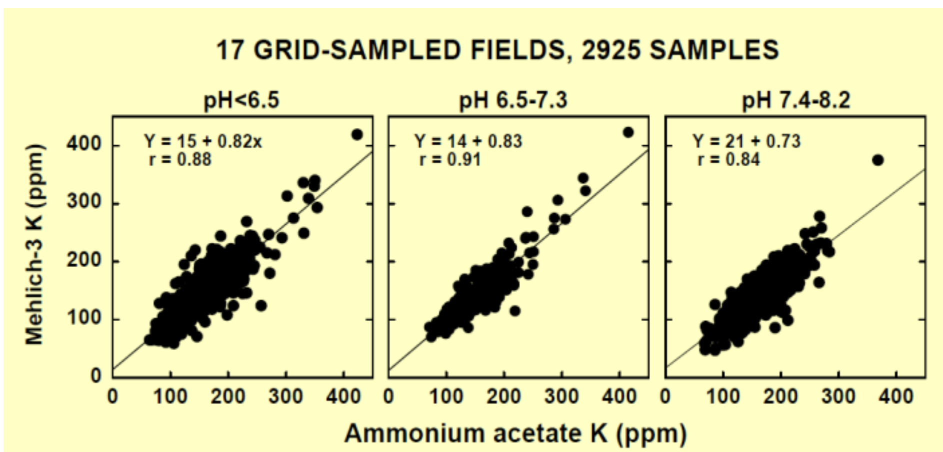


Figure 3-3. ***Relationships between soil K contents measured according to NH4Ac and according to Mehlich3, differentiated according to soil pH (Sawyer and Mallarino, 1999)***

$$
\begin{cases}
\text{if pH} < 6.5 \ \rightarrow \text{POlsen (ppm)} = 0.55 * \dfrac{\text{PMehlich3(ppm/100)}}{100} \\
\text{else if pH} < 7.3 \ \rightarrow \text{POlsen (ppm)} = 0.50 * \dfrac{\text{PMehlich3(ppm/100)}}{100} \\
\text{else POlsen (ppm)} = 0.45 * \dfrac{\text{PMehlich3(ppm/100)}}{100}
\end{cases}
\tag{3-1}
$$

$$
\begin{cases}
\text{if pH} < 7.3 \ \rightarrow \text{KExch(ppm)} = \dfrac{[\text{KMehlich3(ppm)} - 14.5]}{0.825}, \\
\text{and KExch(mmolc/kg)} = \dfrac{\left[\dfrac{[\text{KMehlich3(ppm)} - 14.5]}{0.825}\right]}{39} = -0.45 + 0.031 * \text{KMehlich3(ppm)} \\
\text{else KExch(ppm)} = \dfrac{[\text{KMehlich3(ppm)} - 14.5]}{0.825}, \\
\text{and KExch(mmolc/kg)} = \dfrac{\left[\dfrac{[\text{KMehlich3(ppm)} - 21]}{0.73}\right]}{39} = -0.738 + 0.035 * \text{KMehlich3(ppm)}
\end{cases}
\tag{3-2}
$$

.

The map of soil exchangeable K represents data ($Cmol_c$/kg) measured according to K-NH4Ac, which were converted from the source map that represented data for extractable K (in ppm) measured according to KMehlich3. The conversions given by Eq. 3-2 were also based on rules derived from data presented by (Sawyer and Mallarino, 1999) and suggested by de Pater (2015), according to Leenaars et al. (2018b). The data presented by Sawyer and Mallarino (1999) are visualized in Figure 3-3. The soils of sites were mainly sandy (Bua et al., 2020) and acidic, with a low level of exchangeable potassium (KExch). The mean pH was 6.06. The minimum and maximum POlsen values were 2.03 mg/kg and 9.02 mg/kg, respectively. The majority of soil chemical properties were below average for optimal maize production (Table 3-2).

Table 3-2. Summary of critical soil property levels for the 889 on-farm trials data

	Elev[a] (mm)	Rzd[b]	Corg[b] (g/kg)	Norg[b] (g/kg)	Pt[b] (mg/kg)	POlsen[b] (mg/kg)	KExch[b] (mmol/kg)	pH[b]
Min	115	10	3	0.20	112	2.03	1.88	5.4
Max	446	150	16	1.25	468	9.92	3.64	6.5
Mean	203.49	74.85	5.80	0.54	214.05	3.87	2.82	6.06
SD	64.18	28.29	2.72	0.23	81.97	1.51	0.34	0.17
CV	32%	38%	47%	43%	38%	39%	12%	3%

Table 3-3. Summary of critical soil property levels for the 318 on-station trials data

	Elev[a] (mm)	Rzd[b]	Corg[b] (g/kg)	Norg[b] (g/kg)	Pt[b] (mg/kg)	POlsen[b] (mg/kg)	KExch[b] (mmol/kg)	pH[b]
Min	13	0	4	0.311	113	2.00	1.99	5.4
Max	356	150	15	1.206	267	8.49	4.85	6.4
Mean	241.57	116.80	7.48	0.79	209.33	4.25	2.74	6.04
SD	49.50	42.32	4.05	0.24	34.25	0.98	0.54	0.17
CV	20%	36%	54%	30%	16%	23%	20%	3%

a. Download – CGIAR-CSI SRTM; b. Hengl et al. (2017), c. Hengl et al. (2015); Elev = Elevation, Rzd = Root zone depth, Corg = Organic Carbon, Norg= Organic nitrogen, Pt = Total phosphorus, POlsen = Phosphorus Olsen, KExch = Exchangeable Potassium; Min = Minimum; Max = Maximum, SD = Standard Deviation; CV = Coefficient of Variation.

Average monthly rainfall and minimum and maximum temperatures were derived from the WorldClim database and are summarized in Figure 3-4. Those data were classified based on the suggested planting date for maize in Ghana as a function of AEZ (Adu et al., 2014) and on the assumption that maize was harvested four months after planting. There are two seasons in terms of planting: minor and major. September is the month with the highest rainfall amount; March and December are the hottest and coolest months, with minimum and maximum average temperatures from 1992 to 2017 of 19.5°C and 37.4°C (Figure 3-4).

3.2.3 Diagnostic to Select Grain Yield Data

The dataset used in this paper contained experimental data with yields expressed in different units: kilograms per hectare (kg/ha), metric tons per hectare (mt/ha), and megagrams per hectare (Mg/ha). Therefore, to harmonize and facilitate the reading of the yield units, they have all been converted to kg/ha.

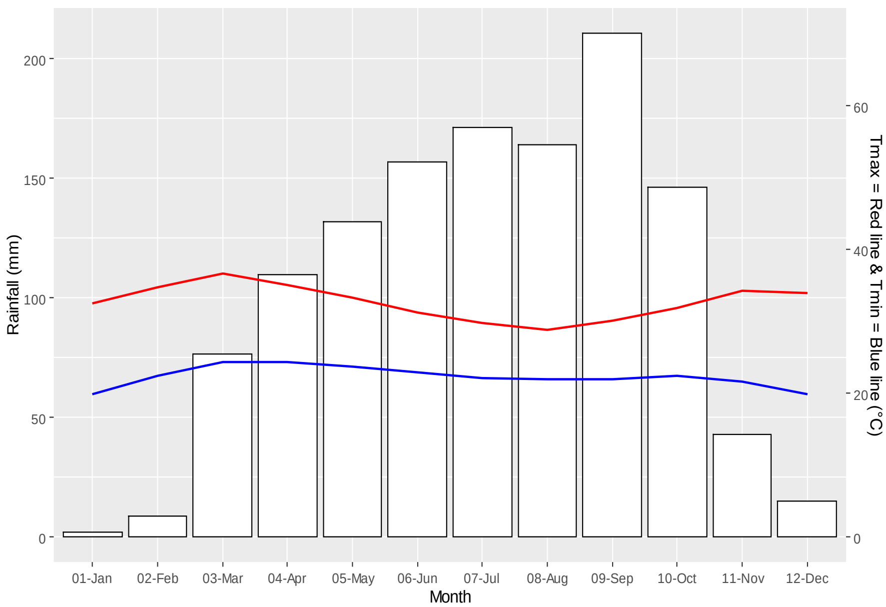


Figure 3-4. Mean monthly maximum and minimum temperatures (°C) and rainfall during maize planting and growth (1990-2017) (Fick and Hijmans, 2017)

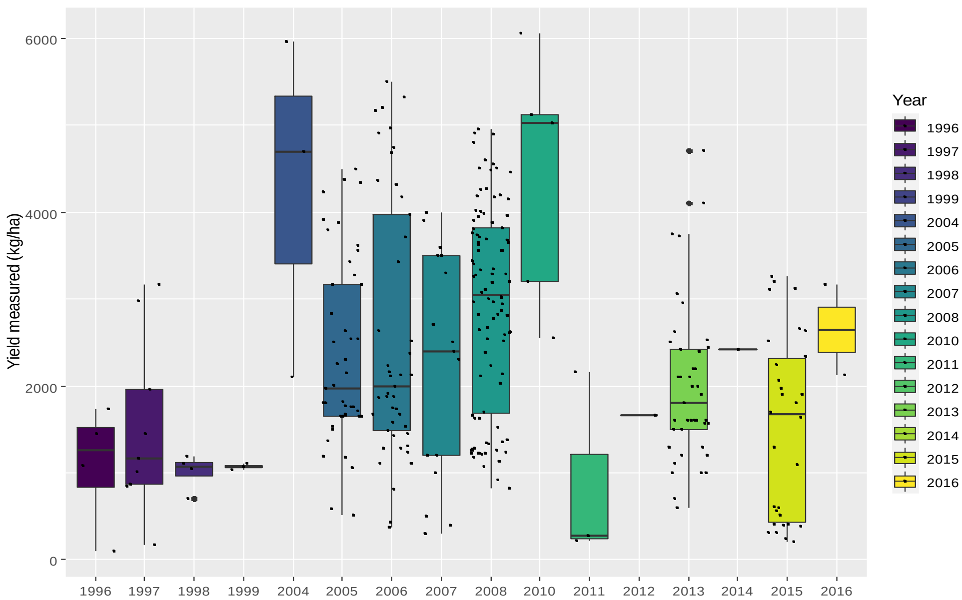


Figure 3-6. ***Maize yield spatio-temporal variability at station experimental level***

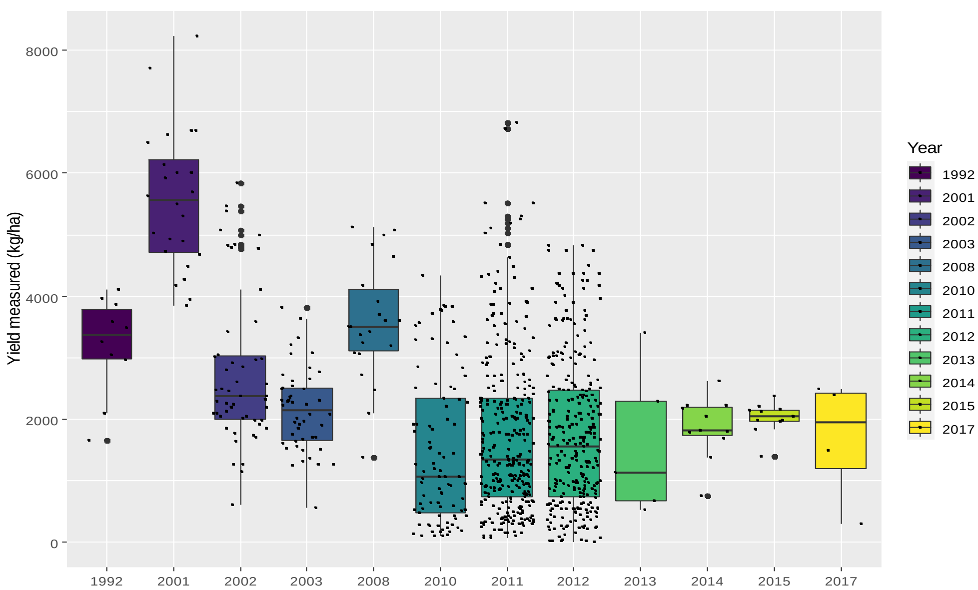


Figure 3-5. ***Maize yield spatio-temporal variability at farm experimental level***

In farm trials, the yield varied from 11 to 5,030 kg/ha in the control plot (*no fertilizer*) and from 141 to 8,230 kg/ha with fertilizer application. The average yield over the years was nearly 1,956 kg/ha. For the station trials, the yield varied from 100 to 5,030 kg/ha in the control plot (*no fertilizer*) and from 400 to 6,030 kg/ha with fertilizer. Figure 3-6 and Figure 3-5 show grain yield per year, boxplots of yields per year, and the dispersion of yields across boxplots. Eighty percent of the yield data from farm trials were obtained between 2010 and 2012. Some yield data are beyond 2-3 mt/ha in control and above 5-6 mt/ha, reaching 6.5, 7, and 8 mt/ha when fertilizers were applied (yield of 2001, Figure 3-6). Those yield data points were inconsistent and extremely higher than the range reported in the literature. According to USAID/IFDC (2015), potential yield with Ghana maize varieties *Obatanpa* and *Mamaba* is 4-5 mt/ha and 6-7 mt/ha, respectively. However, Adu-Gyamfi et al. (2019) conducted a study with Obatanpa and showed that maize yield can reach 6.5 mt/ha by applying fertilizer. In addition, the data point where yield reaches 8 mt/ha in 2001 is located in the Deciduous Forest/Ashanti region, where climatic conditions are adequate to get such yield: Tmin = 19°, Tmax = 28°, summation of rainfall during the four growing months of 540 mm, field elevation 400 m, and rootable zone depth 150 cm. Based on factual parameters and scientific background on maize fertilizer research, we did not exclude those exceptional (7, 8 mt/ha) yield data in the study.

3.3 Experimental Design

On the farms, trials concerned four AEZs, and on the stations, trials concerned five AEZs: Guinea Savannah (*farm n=698, station n=43*), Semi-Deciduous Forest (*farm n=124, station n=85*), Sudan Savannah (*farm n=4; station n=19*), Forest-Savannah Transitional (*farm n=63; station n=164*), and Coastal Savannah (*farm n=0; station n=7*). The treatments included a control (T_0 = *no fertilizer*), those in which PK, NK, and NP were applied and thus N, P, or K was omitted, respectively, and those in which NPK was applied. All experiments in which NPK was combined with micronutrients were excluded from the study. Organic fertilizers were quantified as NPK in the experiments in which they were applied (Appendix , B, and C).

Table 3-4. Fertilizer rate and combination on farm

NPK Treatment	Observations	Nutrient Rate (kg/ha)			Screened Treatment Combinations for the Correlation Test
		N	P	K	
T_0	373	0	0	0	Included
HHH	20	90-180	40-90	40-60	Included
HHL	1	120	60	0	Dropped
HLH	1	120	0	60	Dropped
HLL	38	90-135	0	0	Included
LHH	2	0	50	60	Dropped
LLL	15	7-23	3-15	5-15	Included
LLM	4	24-25	3-4	26	Dropped
LMM	1	23	23	23	Dropped
MHH	236	40-86	40-60	40-74	Included
MLH	6	33-75	09-12	42-78	Dropped
MLL	61	45-90	0-19	0-20	Included
MMH	4	69-70	33-34	86	Dropped
MML	5	58-86	25-38	13-19	Dropped
MMM	122	30-82	20-38	20-39	Included
All	889				

Source: Bua et al. (2020). T_0 = control, H = high level, M = medium level, and L = low level.

Various types of inorganic and organic fertilizers were used, including NPK 15-15-15; urea; ammonium sulfate; nitrogen, phosphorus, and sulfur; diammonium phosphate; NPK 20-10-10; NPK 23-10-5; NPK 20-20-20; muriate of potash; and potassium. Organic fertilizers included poultry manure, cow dung, household waste, market waste, fertisol, biochar, palm bunch ash, and plant residues (*C. ordorata, C. juncea, and maximum P*) as green manure (Appendices A, B, and C). Field fertilization treatments included farmer practice, optimal nutrient management, nutrient omission treatments based on optimal nutrient management, and various fertilizer treatment rates (Bua et al., 2020). For inorganic fertilizers, all rates were applied in kilograms of product per hectare. All P_2O_5 and K_2O values were converted back to kilograms of P or K by dividing P_2O_5 by 2.29 and K_2O by 1.21, based on molar weights, respectively.

Table 3-5. Fertilizer rate and combination on-station

NPK Treatment	Observations	Nutrient Rate (kg/ha)			Screened Treatment Combinations for the Correlation test
		N	P	K	
T_0	69	0	0	0	Included
HHH	32	90-180	45-90	45-91	Included
HHL	14	120-180	60-90	0	Included
HHM	2	180	90	45	Dropped
HLH	19	109-281	0-90	45-90	Included
HLL	16	90-127	0	18	Included
HLM	4	118-163	1	36	Dropped
HMH	3	180	45	90	Dropped
HML	12	120	30	0	Included
LHH	6	0	90	90	Dropped
LML	18	0	30-60	0	Included
LMM	12	0-24	20-45	20-46	Included
MHH	14	45-60	40-90	40-90	Included
MHL	22	40-80	45-60	0	Included
MLH	5	45	0	45	Dropped
MLL	33	37-82	0-18	0-19	Included
MLM	2	60-73	0	36	Dropped
MML	22	40-80	26-30	0	Included
MMM	4	30-60	20-40	20-41	Dropped
All	309				

Source: Bua et al. (2020). T_0 = control, H = high level, M = medium level, and L = low level.

3.4 QUEFTS, MLR, and ML Models

3.4.1 QUEFTS Model

The QUEFTS model was first proposed by Janssen et al. (1990) to estimate maize yield in tropical areas with and without fertilization. Over time, researchers modified it and adapted the model for several crops, such as cassava (Ezui et al., 2017), wheat (Sattaria et al., 2014), watermelon (Kang et al., 2020), and tea (Tang et al., 2020).

Step 1

In the first step, the QUEFTS model uses purely empirical linear and nonlinear multiple regression equations to estimate the potential soil supply of available N, P, and K, based on organic carbon (Corg), phosphorus Olsen (POlsen), exchangeable potassium (KExch), and pH, and optionally organic nitrogen (Norg) and total phosphorus (Pt), as independent variables (Sattaria et al., 2014). Additional to the nutrient supply from the soil, nutrient supply from fertilizer application is determined by adding a term that calculates the fertilizer recovery of applied fertilizers. Default values for maximum recovery fractions of N, P, and K in QUEFTS are 0.5, 0.1, and 0.5, respectively. If appropriate data are available, values can be calculated for each case, but in our study, these default values were used. A crucial requirement for the assessment of the maximum supply of an available nutrient, from the soil as well as from input, is that all other growth factors, including the availability of nutrients other than the one under study, are at the optimal level. For that assessment, the following relations are developed (Janssen et al., 1990):

$$S_N = \alpha_N f_N \mathrm{Corg} + I_N R_N \tag{3-3}$$

$$S_P = \alpha_P f_P \mathrm{Corg} + \beta_P \mathrm{POlsen} + I_P * R_P \tag{3-4}$$

$$S_K = \frac{\alpha_K f_K \mathrm{KExch}}{\gamma_K + \beta_K \mathrm{Corg}} + I_K R_K \tag{3-5}$$

where S_N, S_P, and S_K are supplies of crop-available N, P, and K, respectively; α, β, and γ are empirical parameters; I_N, I_P, and I_K refer to N, P, and K inputs to the system, f_i is a pH dependency coefficient (Eqs. 3-8, 3-9, and 3-10), and R_N, R_P, and R_K refer to the maximum recovery fraction of each fertilizer. When data on organic nitrogen are available, the soil supply of available nitrogen can be calculated by Eq. 3-6 (Sattaria et al., 2014). The value of α_{NN} is 10 times that of α_N in Eq. 3-3, assuming that the C:N ratio of the organic matter is 10.

$$S_N = \alpha_{NN} f_N Corg + I_N R_N \qquad (3\text{-}6)$$

$$S_P = q_P f_P Pt + \beta_P POlsen + I_P R_P \qquad (3\text{-}7)$$

Eq. 3-7 calculates S_P with Pt (Sattaria et al., 2014). The default values of α_P in Eq. 3-4 and q_P in Eq. 3-7 are 0.35 and 0.014, and hence, their ratio is 25. This suggests that the ratio of Pt/Corg is also 25 when Pt is expressed in mg/kg and Corg in g/kg. Such a value was found as an average for this ratio in areas where no fertilizer P had been applied, according to Janssen et al. (1990). Once farmers start to apply inorganic fertilizer P, the ratio increases, and then it is recommended to use only Eq. 3-7 (Sattaria et al., 2014). The coefficient f_i (i = N, P, and K) in Eqs. 3-3 through 3-7 is used to describe the pH dependency of soil organic matter mineralization, P solubility, and K exchangeability, as discussed in more detail by Diest (1980), Janssen and Guiking (1990), and Janssen et al. (1990).

$$f_N = 0.25 * (pH - 3) \qquad (3\text{-}8)$$

$$f_P = 1 - 0.5 * (pH - 6)^2 \qquad (3\text{-}9)$$

$$f_K = 0.625 * (3.4 - 0.4pH) \qquad (3\text{-}10)$$

Eqs. 3-8, 3-9, and 3-10 express the pH correction factors for Eqs. 3-3, 3-4, and 3-5, respectively.

Step 2

In this step, QUEFTS quantifies the relation between potential soil nutrient supply and actual N, P, and K uptake. The relationship between the potential supply of nutrients and actual absorption is based on the following considerations. First, the nutrients are paired for comparison. Therefore, the relationship between the actual uptake and the potential supply of N is calculated twice, resulting in two estimates of the actual uptake of each of the three nutrients. Sattaria et al. (2014) reported that $U_i(j)$ refers to the uptake of *i* with *j*. If i = N, j may be P or K. In other words, two values for N uptake can be calculated based on the potential supplies of P and K. In compliance with the Law of the Minimum, QUEFTS utilizes the lowest of the two N uptake estimates for further calculations. P and K uptake are calculated in the same way (Janssen et al., 1990). Eq. 3-11 models the process of nutrient uptake calculation, where $U_i(j)$ is actual nutrient (i) uptake as a function of nutrient (j), (ai) is physiological efficiency (PhE) or internal efficiency (IE) at the maximum accumulation of nutrient (i) (kg grain/kg nutrient [i]), and (di) is physiological efficiency (PhE) or internal efficiency (IE) at maximum dilution of nutrient (i) (kg grain/kg nutrient [i]) (Shehu et al., 2019); and r_i is the minimum nutrient uptake required to produce any grain. According to Janssen and Guiking (1990) and Janssen et al. (1990), r_i is about 5 kg for N, 0.4 kg for P, and 2 kg for K per hectare.

$$
U_i(j) = \begin{cases} S_i & \text{if} \quad S_i < r_i + (S_j - r_j)(a_j/d_j)\,; \\ r_i + (S_j - r_i)(d_j/d_i) & \text{if} \quad S_i > r_i + (S_j + r_j)[2(d_j/a_i) - (a_j/d_i)]\,; \\ S_i - \dfrac{0.25\,[S_i - r_i - (S_j - r_j)(a_j/d_j)]^2}{(S_j - r_j)(d_j/a_i - a_j/d_i)} & \text{Else. With } i, j = N, P, K, i \neq j \end{cases} \tag{3-11}
$$

Step 3

The third step determines the relationship between the actual absorption of N, P, and K of step 2 and the range of yield based on data collected from fertilized and unfertilized maize. This results in six yield estimates. For each *i* (N, P, or K), a yield estimate is calculated for maximum accumulation and dilution of that nutrient within the crop. Yield ranges are calculated between Y_i^a, yield at maximum accumulation (a), and Y_i^d, yield at maximum dilution (d), as functions of the actual uptake (U_i) and the minimum uptake required to produce any grain (r_i). The potential yield should thus be estimated before running the QUEFTS model. For this study, we assumed that the maximum attainable maize yield is 10 mt/ha.

Step 4

In the last step, yield ranges were combined for the ultimate yield estimate. One yield estimate is predicted by averaging the six yield (Y_i^a, Y_i^d) estimates calculated in step 3. The process of combining the production ranges calculated in step 3 consists of two parts. The yield ranges are first combined in pairs (N and P, N and K, and P and K), through Eq. 3-12, and then the average yield of nutrients is calculated using Eq. 3-13. This average is an estimate of the final (Y_E) prediction of the actual yield (Smaling and Janssen, 1993).

$$
Y_{ij} = Y_j^a \frac{2\left(\min(Y_j^d, Y_k^d, Y_{max}) - Y_j^a\right)\left(U_i - r_i - (Y_j^a/d_i)\right)}{\left(\min(Y_j^d, Y_k^d, Y_{max})/a_i\right) - Y_j^a/d_i} - \frac{\left(\min(Y_j^d, Y_k^d, Y_{max}) - Y_j^a\right)\left(U_i - r_i - (Y_j^a/d_i)\right)^2}{\left(\left(\min(Y_j^d, Y_k^d, Y_{max})/a_i\right) - Y_j^a/d_i\right)^2} \tag{3-12}
$$

With i, j, k = N, P, K, and i ≠ j ≠ k

$$
Y_E = \frac{Y_{NP} + Y_{NK} + Y_{PN} + Y_{PK} + Y_{KN} + Y_{KP}}{6} \tag{3-13}
$$

In the analysis of soil nutrient supply against soil parameters (step 1), no distinction was made for AEZs. The main reason for this was that, ideally, one model calibration should be made for the whole dataset (Ravensbergen et al., 2021). A practical reason was that few data points from certain AEZs remained after data preparation.

R (http://www.R-project.org) was used to simulate QUEFTS model scenarios employing scripts based on Sattaria et al. (2014).

3.4.2 Multiple Linear Regression Models

Multiple Linear Regression (MLR) was used to identify the combination of variables affecting maize yield at the farm and station levels. The linear multiple regression is of the form:

$$y = \beta_0 + \sum_{i=n}^{n} \beta_i x_i \tag{3-14}$$

where y is the maize yield, β is the estimated regression coefficient, and x_i (i=1,2,…, n) is the set of predictor variables. MLR models were constructed using the stepwise variable selection method based on Akaike Information Criterion (AIC) using R "MASS" (Venables and Ripley, 2002) and "CAR" (Fox and Weisberg, 2019) packages. AIC is an estimator of prediction errors and, thereby, the relative quality of statistical models for a given set of data. It evaluates how well a model fits the data from which it was generated (Bevans, 2020), compares different possible models, and determines which one is the best fit for the data (Mack, 2016). It is calculated from:

- The number of independent variables used to build the model.
- The maximum likelihood estimates of the model (how well the model reproduces the data).

The formula for AIC is:

$$AIC = 2 * K - 2\ln(L) \tag{3-15}$$

where K is the number of independent variables used and L is the log-likelihood estimate (the likelihood that the model could have produced the observed yield). The best-fit model according to AIC is the one that explains the greatest amount of variation using the fewest possible independent variables (Mack, 2016). We performed the likelihood ratio test using the "ANOVA" function. The coefficients of determination (R^2) of final models (containing all significant treatments and covariates) were calculated as the squared Pearson correlation between predicted and observed values. Predicted values were calculated using the estimated fixed effects coefficients for treatments and covariates.

3.4.3 Random Forest Prediction

Random forest (RF) methodology is used to address two main classes of problems: (1) constructing a prediction rule in a supervised learning problem and (2) assessing and ranking variables based on their ability to predict the response (Boulesteix et al., 2012). In this study, RF models were trained to predict maize yield maps using multiple biophysical variables as predictors (Table 3-1). The RF algorithm has become attractive in several applications because it can cope with high-dimensional data (the so-called "*n << p curse*") and can even be applied in difficult settings with highly correlated predictors (Boulesteix et al., 2012; Kassambara, 2018; Genuer and Poggi, 2020). In addition, it does not depend on a specific stochastic model and can also run and include non-linear association models between the covariates and the dependent variable (Jeong et al., 2016; Genuer and Poggi, 2020; Timsina et al., 2021). The RF algorithm combines numerous prediction trees in which each tree is built from a bootstrap[4] sample drawn from the calibration set. The random forest package in R Random Forest was used to simulate the RF model.

3.5 Geospatial and Statistical Analysis

The first step in the geostatistical analysis was the definition of a regularly spaced grid covering the study area, with a grid size equal to 0.25 km^2. Second, spatial models were defined to predict crop yield at visited and unvisited grid cells using the values of the covariates as:

[4] A Gentle Introduction to the Bootstrap Method (machinelearningmastery.com).

$$Y(s) = f\,[\mathrm{Rain, Tmax, Corg, Norg, POlsen, pH, Rzd, Sand, Clay, Silt}](s) + \varepsilon(s) \qquad (3\text{-}16)$$

where Y(s) represents the maize yield at location "s" that is modeled in two components: a trend function "f" and an error model ε(s), denoting the small-scale fluctuations around "f" with variance "Var (ε(s))." The function "f" determines the global influence of the external covariates (Table 3-1) except for fertilizers (FN, FP, FK), which are only used in statistical analysis through MLR and QUEFTS and not in the geospatial analysis modeled by ML because the applied fertilizers cannot be associated with any node of the created grids. Thus, "f" was modeled as a global predicted linear function using ML.

3.5.1 Geospatial Analysis

QUEFTS and MLR were applied to model the maize yield. The coefficients of these models were used to predict yield at unvisited locations using the grid cell values of the external covariates.

Moran's index (I_m) (Moran, 1948) was used to assess the significance of the pattern of yield data distribution (Antwi et al., 2016; Salima and Bellefon, 2018). The significance of Moran's index can also be expressed numerically as the probability of rejecting the null hypothesis (H_0). H_0 states that the observed yield value is random. The p-value is the limit value of rejection and depends on the number of permutations performed (*Monte Carlo* method). A low p-value indicates that the risk of rejecting the null hypothesis when it is true is low and therefore the observed value is significantly different from a random distribution.

The datasets were subjected to exploratory analysis to identify the outliers, and square root transformations were carried out to ensure normal distribution. After this, semivariogram analysis was done using R software. The normalized data were then analyzed in a geostatistical way by fitting different semivariogram models iteratively to measure the spatial variability (Vieira and Gonzalez, 2003; Liu et al., 2006; Oliver, 2010). The examination of the semivariogram can be calculated using the equation:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N} \left[Z_{(x_i)} - Z_{(x_i+h)} \right]^2 \qquad (3\text{-}17)$$

where N(h) is the number of pairs of points distant from each other separated by a vector h (distance). In addition, the semivariogram provided the necessary input parameters for spatial interpolation (Oliver, 2010). If the semivariogram increases with distance and stabilizes at the a priori variance value, it means that the regionalized variable under study is spatially correlated and all neighbors within the correlation range can be used to interpolate values that were not measured (Vieira and Gonzalez, 2003). Any geostatistical calculation will require semivariances for any distance within the measured domain; there is a need to fit a mathematical model that would describe the variability (Oliver, 2010; Méneroux, 2018-2019) through theoretical semivariograms. In this study, the experimental semivariograms were all fitted using the "*autofitVariogram*"[5] function in the R software.

3.5.2 Digital Yield Prediction Map

Random forest and similar machine learning techniques are already used to generate spatial predictions (Gräler et al., 2016; Jeong et al., 2016; Hengl et al., 2018; Dutta et al., 2020; Lovelace

[5] R: Automatically fitting a variogram (r-project.org).

et al., 2021; Paudel et al., 2021; Timsina et al., 2021), but the spatial location of points (geography) is often ignored in the modeling process. Spatial autocorrelation, especially if still existent in the cross-validation residuals, indicates that the predictions are possibly biased, and this is suboptimal (Hengl et al., 2018).

In this study, the digital yield prediction map (DYPM) was designed using the RF algorithm coupled with principal component analysis (Bertolini et al., 2015; Bertolini et al., 2018), called the random forest for spatial predictions framework (RFsp) (Hengl et al., 2018). RFsp is a method in which buffer distances from observation points are used as explanatory variables, thus incorporating geographical proximity effects into the prediction process. Dimensionality reduction by ordinations is a popular tool in vegetation science to extract key information, frequently corresponding to ecological gradients, from large covariate matrices (Lovelace et al., 2021). Buffer distances are first derived for each model (QUEFTS, MLR-AIC, RF) yield-predicted point, using the buffer function in the "GSIF" R package (Hengl, 2012), which produces a gridded map for each observation point. The spatial prediction model is defined in the formula:

$$\mathrm{Y_{Model}} \sim \mathrm{Layer.1} + \mathrm{Layer.2} + \mathrm{Layer.3} + \ldots + \mathrm{Layer.\ldots Layer.n} \quad (3\text{-}18)$$

which means that the model yield predicted is a function of "n" covariates for on-farm and on-station trials. In this method, covariates are the principal component (PC) from principal component analysis. Next, we overlay the yield-predicted point and covariates to create a regression matrix so that we can tune and fit a "ranger model" (Wright and Ziegler, 2017) and generate a predictions yield map base on QUEFTS, MLR-AIC, and RF maize yield estimated.

3.5.3 Statistical Analysis

Associations between data were evaluated by Pearson's correlation analysis, and linear regression equations were estimated by least squares regression. Differences between observed and predicted yield were evaluated by paired Student's t-test at the 0.05 significance level. Variability explanation and model accuracy were assessed using three statistical tests, i.e., root means square error (RMSE), coefficient of determination (R^2), and adjusted coefficient of determination (Adj. R^2):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\mathrm{Y_{P_i}} - \mathrm{Y_{O_i}})^2}{n}} \quad (3\text{-}19)$$

$$\mathrm{R}^2 = \left(\frac{\sum_{i=1}^{n}(\mathrm{Y_{P_i}} - \overline{\mathrm{Y}}_{\mathrm{O_i}})(\mathrm{Y_{P_i}} - \overline{\mathrm{Y}}_{\mathrm{O_i}})}{\sqrt{\sum_{i=1}^{n}(\mathrm{Y_{P_i}} - \overline{\mathrm{Y}}_{\mathrm{O_i}})^2}\sqrt{\sum_{i=1}^{n}(\mathrm{Y_{P_i}} - \overline{\mathrm{Y}}_{\mathrm{O_i}})^2}}\right)^2 \quad (3\text{-}20)$$

$$\mathrm{Adj.\,R}^2 = 1 - \left[\frac{(1 - \mathrm{R}^2)(n-1)}{n-k-1}\right] \quad (3\text{-}21)$$

where $\mathrm{Y_{P_i}}$ and $\mathrm{Y_{O_i}}$ represent the values of predicted and observed yield (kg/ha), respectively; n is the number of data point values; $\overline{\mathrm{Y_O}}$ represents the average value of observed yield (kg/ha); and k is the number of independent regressors. The equations of RMSE measure the average discrepancy between the predicted and observed yield data with the same unit (kg/ha). It is an error-index in which the lower the value indicates better model performance (Moriasi et al., 2007). The R^2 estimates the combined dispersion against the single dispersion of the observed and predicted yield series (Krause et al., 2005); it ranges between 0 and 1, where a value of 0 means no correlation at all and a value of 1 means the dispersion of prediction is equal to that of observation. Both R^2 and Adj. R^2 give an idea of how many yield data points fall within the line of the regression equation. However, there is one main difference between R^2 and the Adj. R^2: R^2 assumes that every single covariate explains the variation in the observed yield response to fertilizer, whereas the adjusted Adj. R^2 tells us the percentage of variation explained by only the covariates that affect the observed yield response to fertilizer. Adj. R^2 also is between 0 and 1.

Data were analyzed, modeled, simulated, and visualized via RStudio© (http://rstudio.org/) and GeoDa™ (https://geodacenter.github.io/) spatial analysis software. Only Figure 4-4 was done using Microsoft 365 Excel.[6] For geostatistical analyses, the R packages "gstat" (Pebesma, 2004; Gräler et al., 2016) and "geoR" (Ribeiro Jr and Diggle, 2006), representing the state-of-the-art according to Heuvelink and Rossiter (2021), were used in the RStudio© integrated development environment.

[6] https://office.microsoft.com/excel.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 QUEFTS-Simulated Maize Yield

Average yield observed (AYO) and average yield predicted (AYP) in farmers' trials (Figure 4-1A) were lower than those in the on-station trials (Figure 4-1B) at 1,957-2,389 kg/ha and 2,328-2,916 kg/ha, respectively. The differences observed in yields between farmers' trials and on-station research trials can be attributed to various biophysical constraints (Onduru and Du Preez, 2007; Mugwe et al., 2009). Mugwe et al. (2009) identified that possible explanations for this discrepancy in yields are better management and agronomic practices on-station than on farms. The strong effects of management often result in patterns of decreasing soil fertility, as reported.

The QUEFTS model failed to account for the actual levels of maize yields: the AYPs by QUEFTS were higher than the AYOs in the most treatment combinations (Table 4-1 and Table 4-2). The model-predicted yield varied with soil chemical properties and fertilizer input levels. There were low positive correlations between the observed maize yields and the QUEFTS-predicted yields on the farm ($r = 0.50$, $p < 0.001 < 0.05$) and on-station ($r = 0.35$, $p < 0.001 < 0.05$) (Figure 4-2). QUEFTS explained up to 25% and 12% of the yield response variability from fertilizer and soil chemical properties, respectively, in farm and station trials. There was also low correlation between the observed maize yields and the QUEFTS-predicted yields based on the NPK combination (Table 4-1 and Table 4-2) and the AEZ.

On-station, among the NPK combinations, LML (0:30-60:0 kg/ha) and MHH (45-60:40-90:40-90 kg/ha) showed the highest negative correlation between the AYO and AYP ($r = -0.80$, $p < 0.05$ and $r = -0.69$, $p < 0.05$, respectively), and the highest AYP (4,688 kg/ha) was from the HHH (90-180:45-90:45-91 kg/ha) treatment. The control treatment (T_0) showed low positive correlation ($r = 0.17$, $p > 0.05$) at 1,570 kg/ha AYP against 1,473 kg/ha AYO. However, the relation between observed and predicted yield of most NPK treatment combinations is not statistically significant, with a high a standard deviation (SD). Treatment MHH suggests that QUEFTS calculates lower yields when P and K increase because there is a negative coefficient of correlation.

Table 4-1. Summary of observed and predicted yield responses for on-station NPK combinations

NPK Treatment	Yield Observed		Yield Predicted		R^2	r	*p*
	Mean	SD	Mean	SD			
T_0	1473	1042	1570	569	0.03	0.17	
HHH	2812	1233	4688	556	0.11	-0.33	
HHL	3639	1027	4641	424	0.07	-0.27	
HLH	1927	597	3775	468	0.16	-0.4	
HLL	2311	1116	2737	470	0.19	-0.44	
HML	4042	472	3773	268	0.19	-0.43	*
LML	1261	367	1646	492	0.65	-0.8	*
LMM	2361	1258	2364	846	0.19	-0.43	
MHH	2427	1387	3441	771	0.48	-0.69	*
MHL	3266	1014	3209	488	0.00	0.03	
MLL	2171	851	2471	527	0.07	-0.26	
MML	3258	668	3093	484	0.14	0.38	
All	2389	1300	2916	1225	0.12	0.35	*

* denotes a statistically significant test.

At AEZ level, where on-station trial data had been collected, only in the Forest-Savannah Transitional (FST) and Sudan Savannah (SS) was the relation between AYO and AYPstatistically significant ($r = 0.49$, $p < 0.05$ and $r = 0.48$, $p < 0.05$, respectively), with positive correlation. The AYPs in those two AEZs were 1,464 kg/ha for SS and 3,000 kg/ha for FST, which are above the AYOs of 794 kg/ha and 2,900 kg/ha. These AYOs and AYPs tend to increase in function of forest gradient. The poor correlations suggest that factors other than soil fertility were limiting in the AEZs, causing actual yields to be lower than QUEFTS yields.

Table 4-2. Summary of observed and predicted yield responses based on-farm NPK combinations

NPK Treatment	Yield Observed		Yield Predicted		R^2	r	p
	Mean	SD	Mean	SD			
T_0	1181	979	1508	539	0.13	0.37	*
HHH	3397	702	4217	386	0.27	0.52	*
HLL	3208	1647	2887	899	0.16	0.40	*
LLL	2485	883	1812	451	0.08	0.28	>0.05
MHH	2212	1124	3164	290	0.01	-0.06	>0.05
MLL	3806	1912	2863	583	0.00	0.06	>0.05
MMM	1999	956	2371	469	0.09	0.30	*
All	1957	1389	2328	925	0.25	0.50	*

* denotes a statistically significant test.

On-farm, the NPK treatment combination HHH (90-180:40-90:40-60 kg/ha) showed the highest positive correlation between the observed yield and predicted yield ($r = 0.52$, $p < 0.05$). The observed control (T_0) had a positive correlation with the predicted control ($r = 0.37$, $p < 0.05$) at 1,181 kg/ha AYO vs. 1,508 kg/ha AYP. Variability studies at AEZ level reveal that, in FST, the correlation between AYO and AYP was positive ($r = 0.52$, $p < 0.05$), with 28.82% of yield variation in AYO explained by the regression model. In SS and SDF, the variability explanation by QUEFTS is very weak at 7.77% and 1.57%, respectively.

The coefficients of variation (CVs) for observed maize yields and QUEFTS-predicted yields based on NPK combination rate and AEZ were high. On-station and on-farm, CVs for observed yields were 54% and 71%, respectively. However, the CVs decreased with increasing amounts of NPK in the model, at 42% and 40% for predicted yields based on NPK treatment combination (high, medium, or low). The same is true when the yields of the trials were categorized according to AEZ. Indeed, the coefficient of variation increased along the gradient (SDF<FST<GS). The high variability in observed yields and predicted yields for NPK rates is illustrated by the large differences between their minimum and maximum values (Figure 4-2) and the standard deviation (Table 4-1 and Table 4-2).

QUEFTS is designed to provide an estimate of yield related to soil fertility, assuming that maize growth and development is not compromised by factors other than NPK, such as drought, lack of plot drainage, limited root depth and penetration, poor crop management, or other yield-reducing factors (Linneman et al., 1979; Janssen et al., 1990). Specifically, the QUEFTS model does not consider factors such as soil root depth, water-holding capacity, environmental temperature, physical soil characteristics, diseases, plant population, varietal choice, weed infestation, sowing time, or other crop management practice, despite their importance in determining crop yields at farm and station research levels. In the QUEFTS simulation, 0.5, 0.1, and 0.5 were used as the default for the maximum recovery fractions of N, P, and K; however, given the results revealed by QUEFTS, we can say that we do not obtain such a great value in practice in the fields. Previous studies have indicated that improved crop yields are obtained when soil fertility, soil depth, farm management, and climatic conditions (rainfall, temperature) are managed synchronously (Sadras and Calvino, 2001; Guilpart et al., 2017; Ravensbergen et al., 2021). This study has shown that the

QUEFTS model alone is not sufficient to explain actual maize yield response variability, either on-station or on-farm, and that soil fertility was therefore not the only factor limiting maize production. Similar studies in Africa (Mulder, 2000; Onduru and Du Preez, 2007; Njoroge, 2019) have also shown that QUEFTS-predicted yields are much higher than and correlate poorly with observed yields. Thus, improvement of maize yields in Ghana AEZs may require a more comprehensive model that considers covariates (including climate and physical soil properties and management) in addition to soil chemical and biological fertility. Despite this wide variability in yield response to fertilizer, we still know little about the spatio-temporal pattern of yields. In this study, soil chemical fertility status was variable across Ghana, as shown in Table 3-2 and Table 3-3. For example, soil organic carbon had a CV of 47% and 54%, and total nitrogen had a CV of 43% and 30% on-farm and on-station, respectively.

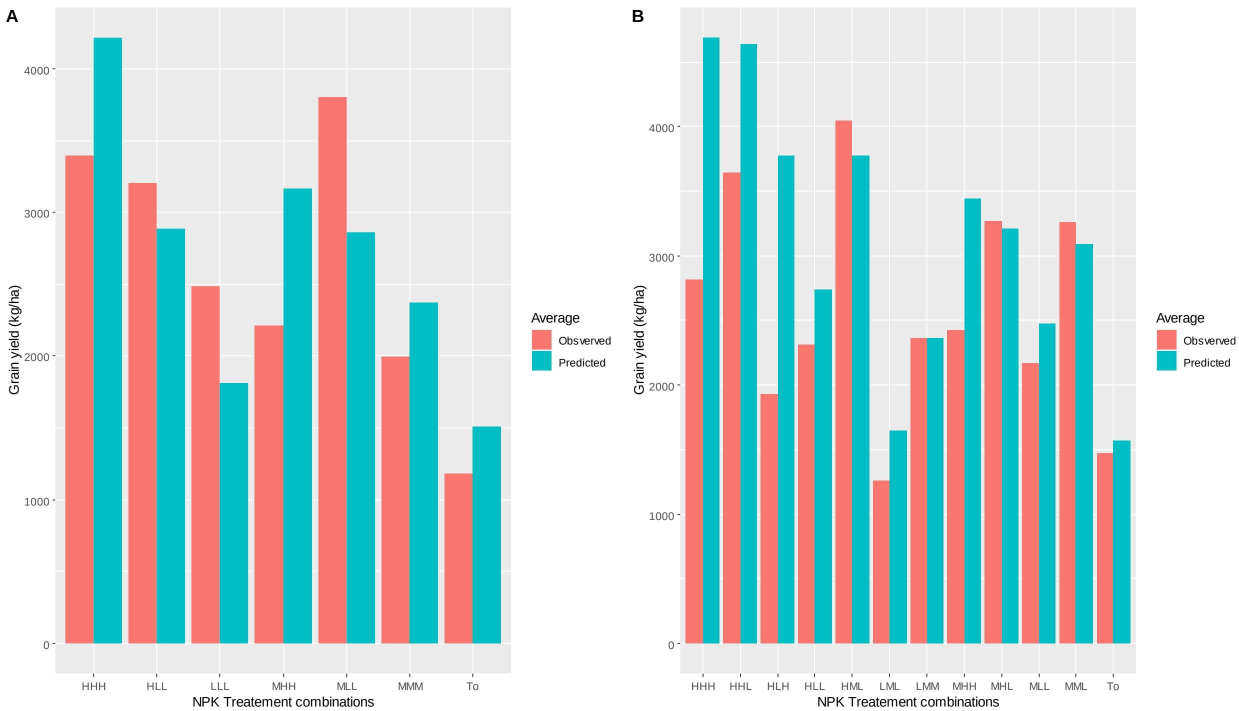


Figure 4-1. Average grain yield observed and predicted (A) on-farm and (B) on-station

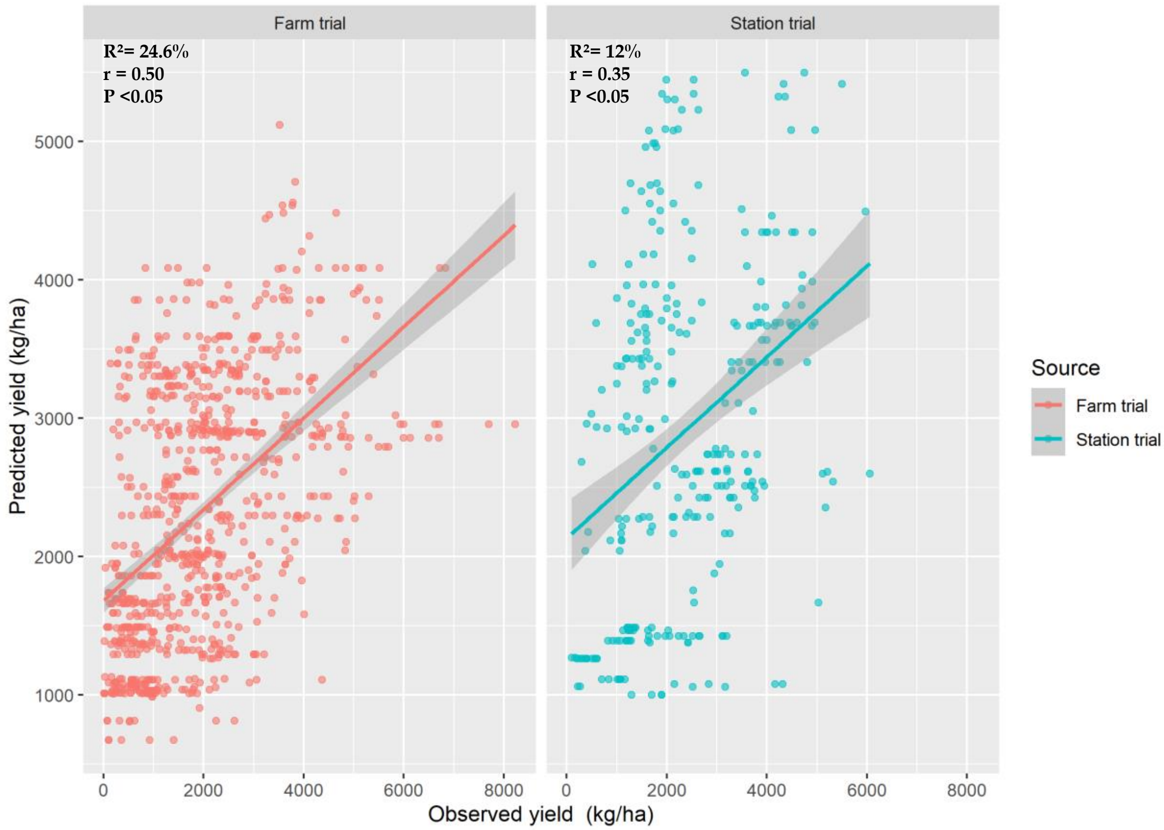


Figure 4-2. Comparison of grain yield predicted by the QUEFTS model and observed yield values

Contributions of three factors (organic carbon, pH, and available phosphorus) to 50 simulated data point yield uncertainties were expressed by model sensitivity analysis. Figure 4-3 shows a trend of variability in predicted yield that was due to the variation of those three factors. When organic carbon varied from 5 to 25 mg/kg, CV decreased from 38% to 9%. At organic carbon equal to 5 mg/kg, with other factors (pH, POlsen, KExch, Pt) constant, CV was 38% with QUEFTS yield predicted at 2,001 kg/ha. Beyond 25 mg/kg, the QUEFTS yield predictions declined. This was due to the C:N ratio. The larger the C:N ratio, the slower the availability of N to the plants (Singh, 2020). In other words, the slower the availability of N to the plants, the greater the C:N ratio and microbial activity immobilizing N. A ratio of 20-30 results in an equilibrium state between mineralization and immobilization (Brust, 2019); thus, the sensitivity of the QUEFTS model is weak. Hence, soil organic C is used as a proxy for soil fertility (Tittonell et al., 2008b).

The findings show that some parameters may not have high first-order sensitivity, yet they have a major influence on model outputs via interactions with other factors. Sensitivities to soil parameters dominate those for cultivar parameters in degraded soils and low-input cropping systems. A comparison of sensitivities among the two simulations (Appendices D and E) shows that organic carbon is a soil factor that limits yield the most strongly, followed by available phosphorus, in simulations. In low-input farming systems, other uncertainties that were not considered by the QUEFTS approach are likely to be dominant in some situations. In particular, biotic stresses caused by weed competition, plant diseases, and insect damage may greatly influence yield, and there are inherent uncertainties in the type, magnitude, and timing of biotic stresses due to the difficulties in measuring and modeling these yield-reducing factors.

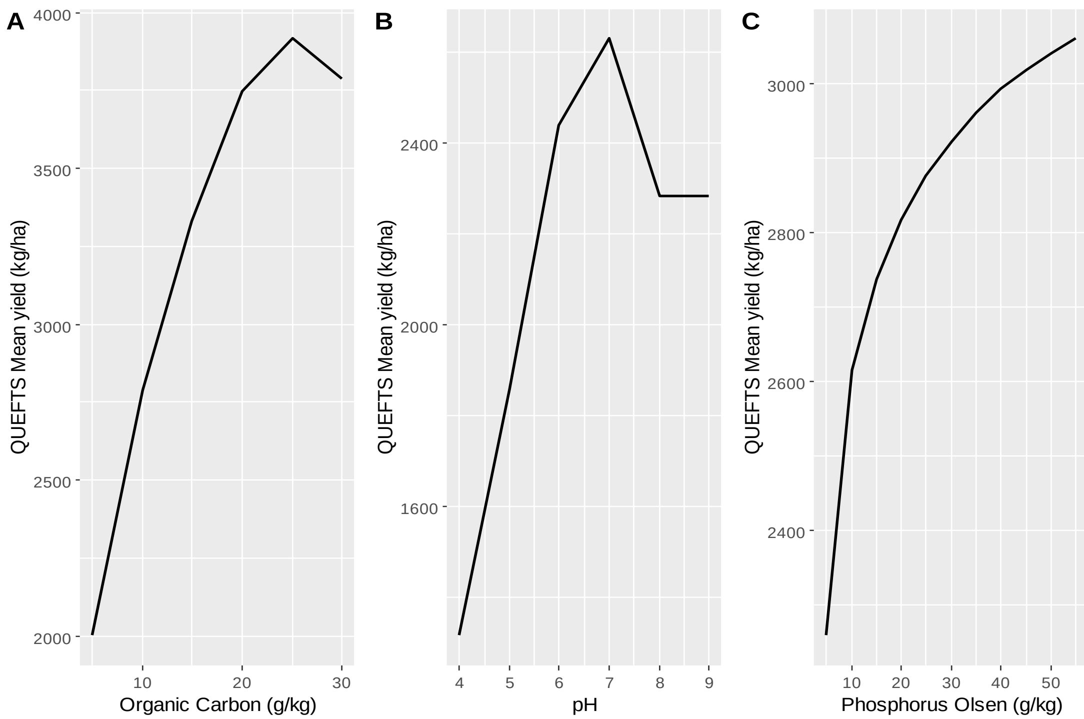


Figure 4-3. Impact of organic carbon, pH, and available phosphorus on yield estimates in QUEFTS

4.2 MLR-Simulated Maize Yield

The results of the prior analysis showed that QUEFTS models based on six soil chemical covariates (Corg, Norg, Pt, KExch, and pH) are not sufficient to grasp yield response variability either on farm or on station. Based on those findings, we went through several MLR models considering other covariates in addition to those used in the QUEFTS simulation.

Table 4-3 and Table 4-4 summarize all MLR models, their coefficient of multiple determination (R^2), and the adjusted coefficient of multiple determination (Adj. R^2).

Table 4-3. Summary of MLR models based on on-station covariate combinations, with MLR 7 including QUEFTS input variables

MLR	Independent variable																			Dependent variable	Model fit		
	Elev	Rain	Tmin	Tmax	Rzd	Rwhc	Sand	Clay	Silt	FN	FP	FK	CEC	pH	POlsen	Ex.K	Pt	Norg	Corg	YO	R^2	Adj.R^2	*p*
1		x	x	x																~	8%	7%	***
2	x	x	x	x																~	8%	7%	***
3					x	x	x	x	x											~	24%	24%	***
4	x				x	x	x	x	x											~	29%	28%	***
5										x	x	x								~	25%	25%	***
6	x									x	x	x								~	25%	25%	***
7													x	x	x	x	x	x	x	~	41%	41%	***
8	x												x	x	x	x	x	x	x	~	41%	41%	***
9	x	x	x	x	x	x	x	x	x											~	30%	27%	***
10	x	x	x	x									x	x	x	x	x	x	x	~	44%	42%	***
11					x	x	x	x	x				x	x	x	x	x	x	x	~	47%	45%	***
12	x	x	x	x	x	x	x	x	x				x	x	x	x	x	x	x	~	49%	46%	***
13	x	x	x	x						x	x	x								~	29%	27%	***
14					x	x	x	x	x	x	x	x								~	39%	37%	***
15	x	x	x	x	x	x	x	x	x	x	x	x								~	46%	43%	***
16										x	x	x	x	x	x	x	x	x	x	~	60%	59%	***
17	x	x	x	x						x	x	x	x	x	x	x	x	x	x	~	62%	60%	***
18	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	~	66%	63%	***
AIC-F[a]	x	x		x	x		x	x	x					x	x			x	x	~	48%	46%	***
AIC+F[b]	x	x		x	x		x	x	x	x	x	x		x	x			x	x	~	65%	64%	***

Table 4-4. Summary of MLR models based on on-farm covariate combinations, with MLR 3 including QUEFTS input variables

MLR	Independent variable																			Dependent variable	Model fit		
	Elev	Rain	Tmin	Tmax	Rzd	Rwhc	Sand	Clay	Silt	FN	FP	FK	CEC	pH	POlsen	Ex.K	Pt	Norg	Corg	YO	R^2	Adj.R^2	*p*
1		x	x	x																~	26%	26%	***
2	x	x	x	x																~	26%	26%	***
3					x	x	x	x	x											~	21%	21%	***
4	x				x	x	x	x	x											~	22%	21%	***
5										x	x	x								~	29%	29%	***
6	x									x	x	x								~	34%	34%	***
7													x	x	x	x	x	x	x	~	16%	16%	***
8	x												x	x	x	x	x	x	x	~	18%	17%	***
9	x	x	x	x	x	x	x	x	x											~	31%	30%	***
10	x	x	x	x									x	x	x	x	x	x	x	~	29%	28%	***
11					x	x	x	x	x				x	x	x	x	x	x	x	~	25%	24%	***
12	x	x	x	x	x	x	x	x	x				x	x	x	x	x	x	x	~	32%	31%	***
13	x	x	x	x						x	x	x								~	46%	46%	***
14					x	x	x	x	x	x	x	x								~	41%	40%	***
15	x	x	x	x	x	x	x	x	x	x	x	x								~	50%	50%	***
16										x	x	x	x	x	x	x	x	x	x	~	40%	39%	***
17	x	x	x	x						x	x	x	x	x	x	x	x	x	x	~	50%	49%	***
18	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	~	52%	51%	***
AIC-F[a]			x	x		x	x	x												~	31%	31%	***
AIC+F[b]			x	x		x	x	x		x	x			x						~	50%	50%	***

a. Akaike Information Criterion without fertilizer; b. Akaike Information Criterion with fertilizer; *** p = 0.00, which means the test is statistically significant.

On-station, MLR 1 performed with climatic parameters only explained 7% of yield response variability. Adding elevation (Elev) as a covariate does not change the explanation because R^2 stays unchanged. The R^2 could explain that elevation does not affect yield response in on-station trials. MLR 5 and 6 performed with soil physical parameters as covariates explained more yield response variability than MLR 1. The R^2 increases from 8% (MLR 1 and 2) to 24-29% (MLR 3, 5, 6, 4, and 13). In MLR 3, soil physical properties explained more yield variability (R^2 =24%) than environmental parameters (MLR 1,2) and adding the elevation covariate in MLR 3 increases variability explanation by 5%. However, elevation added as a covariate in MLR 6, in combination with fertilizer, does not increase R^2. The same result is found in MLR 7 and 8. Indeed, soil chemical properties explained up to 41% of yield variability, and combining soil chemical properties with elevation does not increase the R^2. These results could explain that interactions between elevation and soil chemical properties with fertilizer use do not affect yield in on-station trials. In addition, MLR 13, combining fertilizer (chemical or organic) and climatic parameters as covariates, explained yield variability at the same percentage as MLR 5 and explained more yield variability than MLR 1, 2, 3, 5, and 6. MLR 7, performed with soil chemical properties, explained a 41% yield response variability on-station. These findings show how soil chemical properties are fundamental in the maize production system. This result is corroborated by Braimoh and Vlek (2006), who showed that maize yields are overwhelmingly determined by soil quality. According to this research, under annual/repeated cropping in low-quality soils, the predicted maize yields are below 520 kg/ha. Thus, an understanding of the current status of land and its change over time is needed for promoting land management practices to maintain or improve land productivity and sustainable use of natural resources (Bindraban et al., 2000). In addition, MLR 7 (R^2 = 41%) explained more yield variability in the control plot (T_0) than the QUEFTS yield estimate (R^2 = 3%), and when fertilizer is used in those two prior models, the yield variability explanation in MLR 16 increased compared to QUEFTS. Indeed, MLR 16 explained 60% of yield response variability to fertilizer against 2.73% for QUEFTS. MLR 16, using the combination of soil chemicals properties and fertilizer applied as covariates, improved the comprehension of yield variability, since the R^2 is 60%. This result confirms those scholars who have said for several decades now that the low fertility of soils in Ghana is the most crucial problem with its impeded agriculture, keeping yields low (Kihara et al., 2016; Bationo et al., 2018; Bua et al., 2020). MLR 14, using fertilizer and physical soil properties as covariates, explained 39% of the variability on-station. Knowing that soil physical properties alone (MLR 3) explained yield variability at 24%, application of fertilizer reduces variability, as discussed in Section 4.1. This result supports the thesis that fertilizer application not only helps to increase yield, but also stabilizes yield fluctuation. As Figure 4-4 shows, in the control plots, the CV is very high, varying from 71% on-station to 83% on-farm, and when fertilizer is applied, the variation drops to an average of 43% on-farm and 36% on-station. The effect of fertilizer on grain yield can be altered or improved by soil physical properties, since using both as covariates explained more yield response variability than each did separately (MLR 3, 5, and 14). The combination of all covariates gave the highest explanation of yield response variability (MLR 18), as R^2 and Adj. R^2 were 66% and 63%, respectively. MLR 18 could be the optimal model to assess yield variability because it encompasses soil physical and chemical properties, the environment, and fertilizer and their effects on grain yield. However, covariates Tmax, Tmin, Elev, Corg, Pt, KExch, Rwhc, and CEC are not statistically significant ($p > 0.05$) when MLR 18 is used. The comparison of two MLR models from the AIC algorithm, one simulated with all covariates plus fertilizer and the other simulated with all covariates without fertilizer, allows us to highlight the positive effect of fertilizer on corn yield.

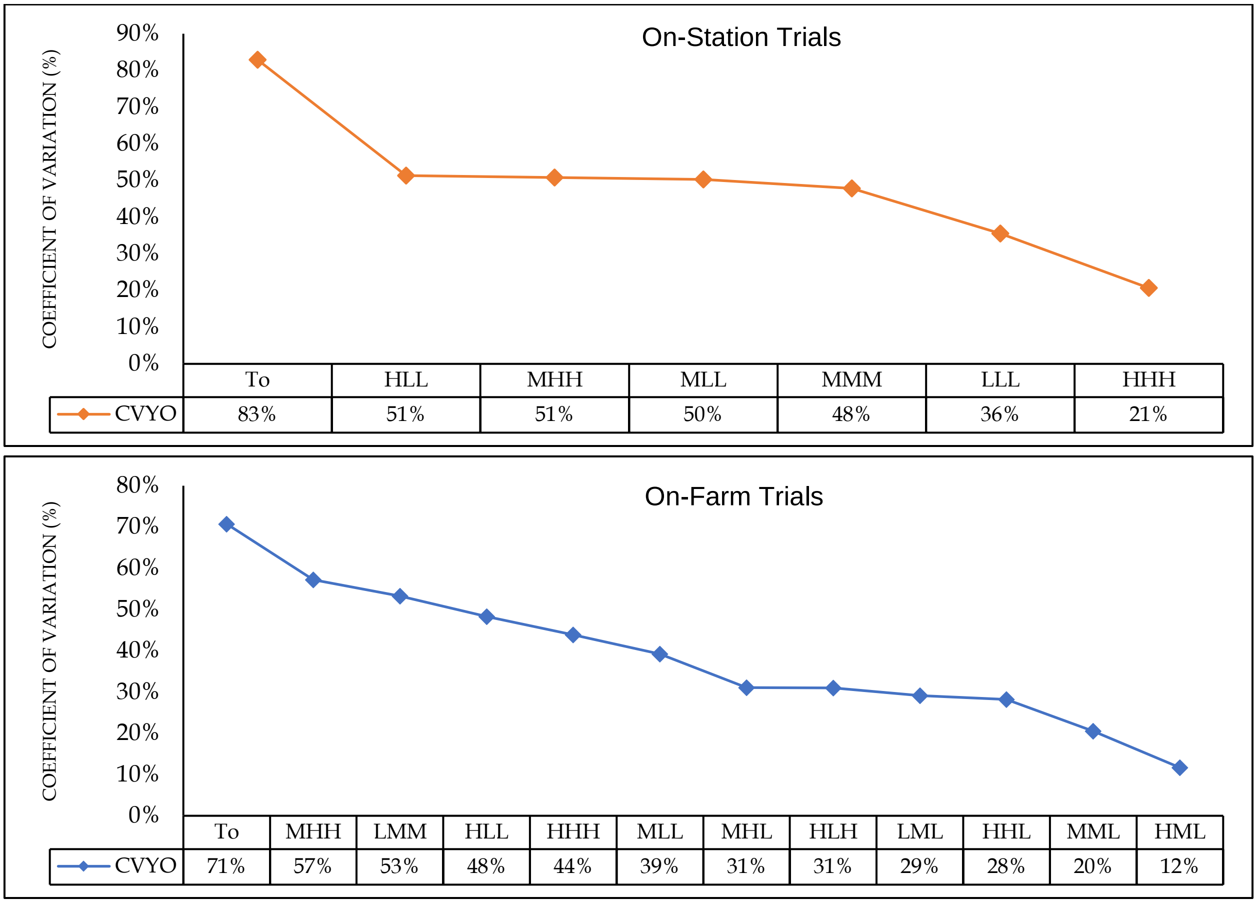


Figure 4-4. ***Coefficient of variation (%) of the observed yield (CVYO) in the control plots (T_0) and in the plots in which fertilizer was applied***

Without fertilizer in the set of independent variables, the best model from the dozens of scenarios produced by the AIC algorithm explained 46-48% of the variability, while the model from the scenarios produced by the AIC algorithm with fertilizer as a covariate in the set of independent variables explained 64-65% of the variability in the response of maize yield to fertilizer. Therefore, MLR AIC+F (Table 4-5) offers the most reliable way to predict yield and draw yield response variability on-station. Eq. 4-1 from MLR AIC+F is used to predict yield and compare observed yield (YO) and predicted yield.

$$\mathrm{YO} = f(\mathrm{Elev} + \mathrm{Rain} + \mathrm{Tmax} + \mathrm{Corg} + \mathrm{Norg} + \mathrm{POlsen} + \mathrm{pH} + \mathrm{Rzd} + \mathrm{Sand} + \mathrm{Clay} + \mathrm{Silt} + \mathrm{FN} + \mathrm{FP} + \mathrm{FK}) + \varepsilon \qquad (4\text{-}1)$$

Table 4-5. On-station best multiple regression model structure (MLR-AIC+Fertilizer)

Variable	Coefficient	Std_Error	t_value	Pr(>|t|)	
(Intercept)	-5885.349	8586.89	-0.685	0.493639	
Elevation	11.425	3.985	2.867	0.004439	**
Rainfall	5.19	1.217	4.266	2.68E-05	***
Temperature max	298.382	101.372	2.943	0.003505	**
Carbon organic	-398.328	38.849	-10.253	< 2e-16	***
Nitrogen organic	2424.193	795.697	3.047	0.002524	**
Phosphorus Olsen	824.131	128.283	6.424	5.32E-10	***
pH	-5155.176	758.177	-6.799	5.86E-11	***
Root zone depth	-17.086	3.895	-4.387	1.61E-05	***
Sand	252.306	50.447	5.001	9.80E-07	***
Clay	188.178	32.984	5.705	2.84E-08	***
Silt	189.089	47.992	3.94	0.000102	***
Fertilizer N	7.812	1.005	7.777	1.26E-13	***
Fertilizer P	11.155	1.987	5.613	4.59E-08	***
Fertilizer K	-7.327	2.123	-3.451	0.000641	***

Dependent variable is Yield Observed, F (14,294) = 39.48, p = 0.00, Std_Error = standard error.
Significance codes: ***0; **0.001. Residual standard error: 783.9 on 294 degrees of freedom.

Table 4-5 shows that maize yield significantly increases ($p < 0.05$) with elevation, rainfall, organic nitrogen, phosphorus, sand, clay, silt, and nitrogen and phosphorus fertilizer. This reveals that maize yield response variability on-station is dependent on a variety of factors. For example, in Kenya highland maize was planted at three different altitudes and showed large yield differences, with yield decreasing with decreasing altitude (Cooper, 1979). Ovalles and Collins (1986) and Kravchenko and Robertson (2007) conducted studies on a broad selection of soil chemical and structural properties, including pH, organic C, total P, coarse sand, medium sand, fine sand, very fine sand, total sand, silt, clay, and soil water retention content from topographic positions. They demonstrated that all of these selected soil properties had a significant dependence on the topographic position of the field. Increasing organic carbon tends to reduce observed yield. The same phenomenon is also observed when pH increases. Table 4-6 shows the negative correlation between pH and Corg (organic carbon), as well as Norg (organic nitrogen). The increase in Corg, leading to a decrease in pH, could immobilize the phosphorus in the soil, reducing its bioavailability, resulting in a reduction in yield via indirect effect through the negative coefficient of Corg in Eq. 4-1. According to Hong et al. (2019) and Zhou et al. (2019), the pH values of the topsoil are low because the topsoil is rich in organic matter and decomposition of the organic matter results in the production of more organic acids. This biogeochemical process is more pronounced in acidic soils, such as those of Ghana. On the other hand, increased pH also reduces yield. The negative correlation between Corg and pH could also be explained by this fact. Indeed, soil pH increases the solubility of soil organic matter by increasing the dissociation of acidic functional groups (Andersson et al., 2000) and reduces the bonding between organic constituents and clays (Curtin et al., 1998). Thus, dissolved organic matter content increases with soil pH and, consequently, mineralizable C and N (Curtin et al., 1998). This explains the significant effects of alkaline soil pH conditions on the leaching of dissolved organic carbon and dissolved organic nitrogen, thereby reducing maize yield, in Eq. 4-1. The negative correlation between rainfall and Corg and Norg, as shown in Table 4-6, confirms this assumption. Thus, when Corg or pH harms the yield response to fertilizer when it increases implies more of an indirect negative effect due to interactions between climate and soil variables. However, we must recognize that this phenomenon is a bit confusing, and beyond the discussion given, other more elaborate research must be done to understand why such negative correlations appear between pH, Corg, and Norg, which reduce drastically yield.

Table 4-6. On-station matrix of correlation

	Rain	Tmax	Corg	Norg	pH	Sand	Clay	Silt	YO
Rain	-					-			
Tmax	-0.68*	-	-						
Corg	-0.17*	-0.2*	-						
Norg	-0.06	-0.52*	0.69*	-					
pH	0	0.41*	-0.52*	-0.71*	-				
Sand	-0.22*	0.49*	-0.46*	-0.56*	0.66*	-			
Clay	-0.13*	-0.3*	0.31*	0.64*	-0.28*	-0.39*	-		
Silt	0.34*	-0.31*	0.31*	0.16*	-0.55*	-0.83*	-0.14*	-	
YO	0.21*	-0.28*	-0.29*	0.15*	-0.25*	-0.1	0.2*	-0.02	-

* denotes significance; see Appendix G for the full matrix.

Figure 4-5A, B, and C highlights the fact that temperature and rainfall have a significant effect on maize grain yield on-station. Those graphs are from MLR AIC+F (Table 4-5) of Eq. 4-1. Indeed, grain yield responses are positively affected by decreasing max and min temperature and by increasing rainfall amount. According to Barimah (2014), agricultural production in Ghana is expected to be negatively affected by the projected changes in rainfall regimes and increases in temperatures. The low level of total nitrogen observed (Table 3-3) indicates that this nutrient is a limiting factor for optimal maize production and that a response to nitrogen is expected (Danso et al., 2020). In Table 4-5, organic nitrogen is shown to have a significant positive effect on yield, and therefore its coefficient is very high, which balances the C:N ratio since the organic carbon coefficient is negative. A negative correlation between the root zone depth (Rzd) and silt ($r = -3.2$) could also explain why an increase in Rzd reduces maize yield (Table 4-5). The more the Rzd increases, the more the silt amount decreases and clay increases, which negatively affects maize yield response to fertilizer, knowing that soil particles such as silt are very important for soil texture quality (Adhinarayanan, 2017; Fang and Su, 2019; Scheiterle et al., 2019; Munialo et al., 2020). Table 4-4 shows that MLR 1 explained more yield variability on-farm than MLR 3 and 7. That means that on-farm environment variables (rainfall and temperature) have more influence than soil physical and chemical parameters since R^2 is 26% for MLR 1 and 16% and 18% for models with soil physical and chemical parameters as yield predictors. When elevation is added in MLR 1, 3, and 7, R^2 and Adj. R^2 do not increase significantly. This leads to the conclusion that the elevation covariate does not provide much of an explanation for the variability observed in the response of maize yield to fertilization when climate and soil physicochemical parameters are considered in a linear model. However, when elevation is added to the covariates (FN, FP, FK), it increases the R^2 of MLR 5, which considers only the fertilizers as an explanatory variable, from 29% to 34%. That is a 5% increase in explanation, which is not negligible in a field context where several variables interact. By comparing the R^2 of each set of covariates, the findings show that fertilizer application has the highest R^2 on observed yield compared with the other set of covariates. MLR 3 (R^2=21%), which has soil physical parameters as covariates, explained more of the observed variability in yield than MLR 7 (R^2=16%), which has soil chemical parameters as covariates. The difference in explanation of maize yield variability between MLR 7 (16%) and QUEFTS (13%) when only native soil fertility is considered as an explanatory variable is not significantly large, with only a 3% discrepancy. On the other hand, there is a large difference (37%) between the R^2 of MLR 16 and that of QUEFTS when fertilizer is added to soil chemical parameters as an explanatory variable. Also, the combination of soil physical and chemical variables in the MLR 11 model does not increase the R^2 much (25%) from the MLR 3 model (21%). This implies that soil chemical parameters do not contribute significantly to explain the variability in maize yield response to fertilizer, as highlighted by the QUEFTS model in Section 4.1. In contrast to the chemical soil parameters, fertilizer increases the R^2 considerably. The R^2 increases from 29% with MLR 5 to

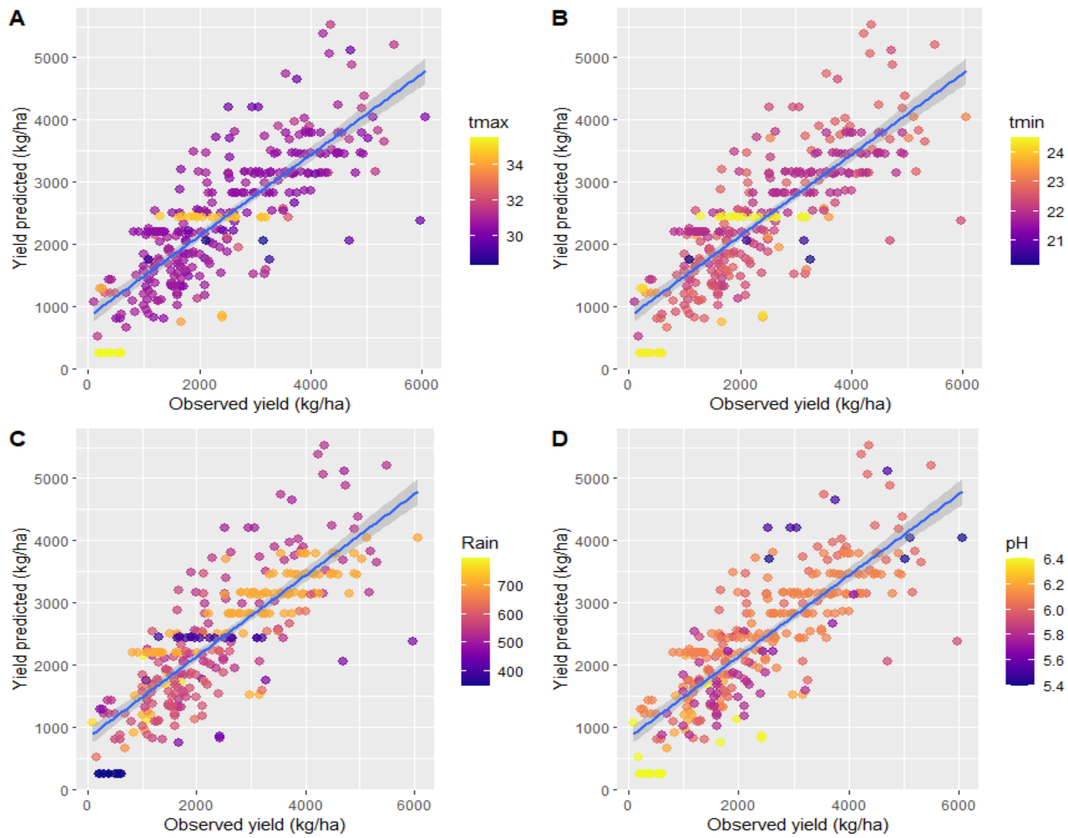


Figure 4-5. ***Scatter plot of observed yield and MLR-predicted yield as a function of (A) Tmax, (B) Tmin, (C) Rain, and (D) pH for on-station trials***

41% with MLR 14 when fertilizer is combined with soil physical parameters and then to 46% with MLR 13 when combined with climatic parameters. Furthermore, fertilizer application contributes to reducing yield variability both on-farm and on-station, as presented in Figure 4-4B. From T_0 via HHH, the coefficient of variation is seen to gradually decrease depending on whether a certain amount of nutrients is supplied or not. The CV decreases from 83% in the control (T_0) to an average of 43% when fertilizer is applied, then stabilizes. With an R^2 ranging from 16%, 21%, and 26% to 40%, 41%, and 46% when fertilizer is applied in combination with soil chemical properties, soil physical properties, and climatic variables, respectively, fertilizer application could be a proxy for maize yield increase in Ghana. However, despite fertilizer's importance ahead of all other covariates, high variability is observed in yield response to it, as confirmed by Bua et al. (2020), Bationo et al. (2018), and other scholarly researchers (Fosu-Mensah et al., 2012; MacCarthy et al., 2017; van Loon et al., 2019). The combination of chemical and physical soil covariates with fertilizer (MLR 17) explained more yield response variation since there is a high Adj. R^2 of 49-51%. MLR 14 with soil physical parameters and fertilizer as covariates expresses the role of soil texture and structure in fertilizer application and nutrient use efficiency. This result is highlighted by Zheng et al. (2003) and Martins et al. (2018), who stress the importance of considering soil texture when applying commercial fertilizers. MLR 18 considers all covariates, and its Adj. R^2 is 51%. Indeed, the combination of all those predictors helped explain 51% of yield response to fertilizer variability on the farm in Ghana.

The AIC algorithm was performed to compare two MLR models, one simulated with all covariates including fertilizer (AIC+F) and one simulated with all covariates except fertilizer (AIC-F). These simulations allow us to once again highlight the positive effect of fertilizer on maize yield. Without fertilizer in the set of independent variables, the best model from the dozens of scenarios produced by the AIC algorithm explained 31% of the variability in yield response to fertilizer, while the final model from the scenarios produced by the AIC algorithm including fertilizer as a covariate in the set of all independent variables explained 50% of the yield response variability on-farm. Between MLR 18 and MLR AIC+F, there is not a huge difference of variability explanation considering R^2 and Adj. R^2. However, in-depth there is a difference in covariate use to explain variability. All covariates in MLR AIC+F are statistically significant ($p < 0.05$), which is not the case with MLR 18. Therefore, MLR AIC+F offers the most reliable way to predict yield and draw yield response variability on-farm. Eq. 4-1 from MLR AIC+F is used to estimate yield and perform the comparison between observed yield and predicted yield. Table 4-7 presents Eq. 4-1 coefficients.

$$\mathrm{YO} = f(\mathrm{Tmin} + \mathrm{Tmax} + \mathrm{pH} + \mathrm{Rwhc} + \mathrm{Sand} + \mathrm{Silt} + \mathrm{FN} + \mathrm{FP}) + \varepsilon \quad (4\text{-}2)$$

Table 4-7. On-farm best multiple regression model structure (MLR-AIC)

	Estimate	Std_Error	t.value	Pr(>|t|)	
(Intercept)	19,628.76	1553.336	12.637	< 2e-16	***
Tmin	-250.883	71.772	-3.496	0.000497	***
Tmax	-346.267	61.232	-5.655	2.11E-08	***
pH	735.945	232.106	3.171	0.001573	**
Rwhc	-246.883	34.008	-7.26	8.54E-13	***
Sand	-39.691	9.832	-4.037	5.89E-05	***
Silt	-46.362	9.394	-4.935	9.57E-07	***
FN	15.327	1.436	10.672	< 2e-16	***
FP	7.3	2.747	2.657	0.008016	**

Dependent variable is yield observed, F (14,294) = 39.48, p = 0.00, Std_Error = standard error.
Significance codes: ***0 and **0.001; Residual standard error: 783.9 on 294 degrees of freedom.

In on-station trials, minimum and maximum temperature affect maize yield negatively (Table 4-7). Peprah (2012) conducted a study on rainfall and temperature correlation with maize yield in Ghana and concluded that temperature explains a larger portion of the maize yield variation than rainfall. This finding is also echoed in our data, as discussed previously. Figure 4-6 reveals that rainfall and temperature are also important factors in explaining the variability in the response of maize to fertilizer. Indeed, Figure 4-6A and B shows a negative correlation between temperature and yield. As the minimum and maximum temperatures increase, the yield tends to decrease despite the application of fertilizer. Lobell et al. (2008) indicated that each day with a temperature above 30°C would reduce the final yield by 1% under optimal rainfed conditions and by 1.7% under drought conditions. According to EPA (2000), even though other contributing factors exist, rising temperature and irregularity in precipitation are the major causes of the continuous reduction in maize yields. This confirms the finding depicted in Figure 4-6A and B, showing that the current temperature levels and evaporation rates in Ghana are high, particularly in the Guinea Savannah, Sudan Savannah, and Coastal Savannah zones, according to a study conducted by Ahene (2003) on the impact of climate change on maize production in Ghana between 1970 and 2002. Since agriculture in Ghana is rainfed, a scarcity of rainfall will harm the crop, but as Figure 4-7C shows, increasing rainfall tends to reduce the yield. This could be explained by Figure 4-7D, which highlights the fact that where rainfall is high, depth of soil is low (*less than 100 cm*); as shown in Table 4-8 of the correlation matrix, the coefficient of correlation between root zone depth and rainfall is negative (-0.14). The shallow depth increases the runoff of rainwater, which is then not available for maize. This runoff also causes leaching of fertilizer and soil nutrients, thus reducing

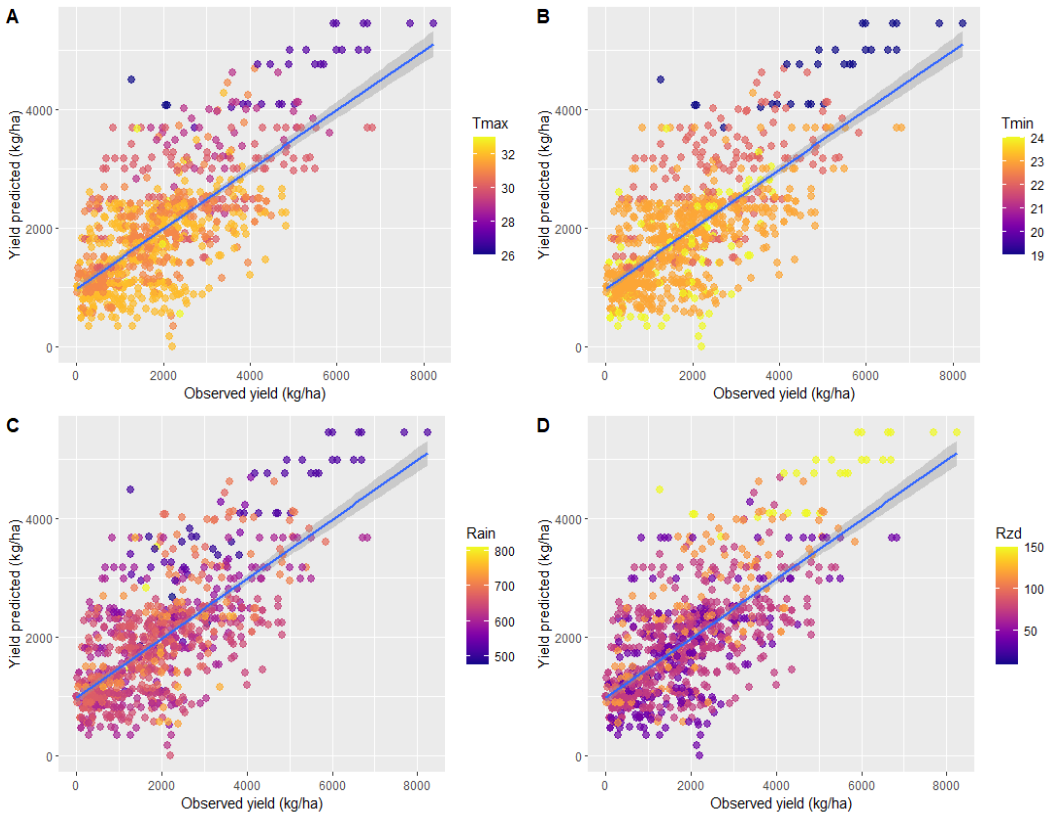


Figure 4-6. ***Scatter plot of observed yield and MLR-predicted yield as a function of (A) Tmax, (B) Tmin, (C) rain and (D) Rzd for on-farm trials***

the chemical and physical fertility of soils and could deteriorate surface water quality. The amount of moisture in the soil depends upon how much rainfall enters the soil and is stored (Shaxson and Barber, 2003). Under rainfed agriculture as in Ghana, the amount of water entering the soil depends on what percentage is diverted above the surface as runoff (Shaxson and Barber, 2003) as well as the soil's capacity to store water. Several scholars have highlighted the importance of soil depth in increasing yield in SSA (O'Halloran et al., 1985; Sadras and Calvino, 2001; Hengl et al., 2017; Leenaars et al., 2018a). According to Guilpart et al. (2017), SSA could become a *grain breadbasket* if rootable soil depths are comparable to those of other major breadbaskets, such as the U.S. Maize Belt and South American Pampas. In addition, the potential for SSA to become another major breadbasket is suggested by the fact that most of the existing SSA cereal cropland (where nearly all grain is produced under rainfed rather than irrigated conditions) receives abundant precipitation (≥*900 mm per year*), equal to or greater than all existing breadbaskets except for the humid tropical lowland rice areas in Asia.

Table 4-8. On-farm matrix of correlation

	Elev	Rain	Tmin	Tmax	pH	Rzd	Rwhc	Sand	Silt	YO
Elev	-									
Rain	-0.16*	-								
Tmin	-0.68*	0.25*	-							
Tmax	-0.54*	0.25*	0.83*	-						
pH	-0.19*	0.12*	0.33*	0.52*	-					
Rzd	0.51*	-0.14*	-0.56*	-0.46*	-0.34*	-				
Rwhc	-0.16*	0.35*	0.09*	0.13*	0.2*	-0.18*	-			
Sand	0.19*	-0.1*	-0.34*	-0.24*	-0.08*	0.33*	-0.22*	-		
Silt	-0.38*	0.08*	0.46*	0.46*	0.21*	-0.35*	0.23*	-0.81*	-	
YO	0.33*	-0.22*	-0.45*	-0.49*	-0.24*	0.27*	-0.25*	0.19*	-0.35*	-

* denotes significance. See Appendix 0 for the full matrix.

The main constraints to maize production in Ghana are drought during the critical early stages of crop growth, low soil nutrient levels (especially nitrogen and phosphorus), low soil pH, pests and diseases, and Striga (*Striga hermonthica*) infestations. *Striga hermonthica*, known as witchweed, is a parasitic weed that is a serious problem in many parts of the Guinea and Sudan Savannah zones of Ghana (Albert and Runge-Metzger, 1995; Sauerborn et al., 2003; Adu et al., 2014). However, it has been neglected according to Albert and Runge-Metzger (1995) and Scheiterle et al. (2019). Scheiterle et al. (2019) conducted a study using several covariates, such as soil properties, *Striga hermonthica*, and plot management, through three statistical models, which showed that the parasitic weed *Striga hermonthica* had a significant negative effect on maize yield in the Guinea Savannah zone of Ghana. In southwestern Kenya, a statistical analysis of the influence of *Striga hermonthica* on maize yields in fertilizer trials revealed that when the Striga is included in regression analysis, the percentage yield variation explained moves up to 55% and 65% (Smaling et al., 1991). Thus, Bua et al. (2020) reported that yield losses due to *Striga hermonthica* could be as high as 100%, depending on many factors. The major pests of maize in Ghana include stem borers, cutworms, grasshoppers, weevils, termites, and the larger grain borer, which have caused many troubles for maize yield response to fertilizer (Adu et al., 2014; Darfour and Rosentrater, 2016b). For example, research has shown that the estimated national mean loss of maize in Ghana is 45% (range 22-67%) due to damage done to the foliage by the younger larvae and consumption of the cob and kernels by the larger larvae that inhabit the whorls of older plants (Bhusal et al., 2020). Other limitations to maize production include poor management practices, such as low plant populations, inappropriate planting time, inadequate control of weeds, limited use of inputs (especially fertilizer and improved seeds) likely due to a lack of credit, and untimely application of adequate quantities of fertilizers (Adu et al., 2014; Darfour and Rosentrater, 2016b). In this study, variability of maize yield response to fertilizer was assessed using biophysical variables and, at a certain level, they explain the variability observed; however, factors other than those cited also contribute largely to maize yield response variability, mostly on-farm.

4.3 Spatial Variability Assessment

4.3.1 Geostatistical Analysis

According to Tobler (1970), "*Everything is related to everything else, but things that are close are more related than things that are far away.*" This law is the basis for the fundamental concepts of spatial dependence and spatial autocorrelation. In our context, this means that the value of maize yield observed at location "s" is more due to the actions and interaction of the covariates that are close to it and in which the trials are conducted. Here, the quantification of the spatial structure of the observed yields in on-farm and on-station trials gave a low positive degree of global spatial autocorrelation: on-farm, $I_m = 0.42$, and on-station, $I_m = 0.44$ (Figure 4-7 and Figure 4-8). To assess the significance of Moran's I's, we compared the observed Moran's I's with thousands (9999) of Moran's I's calculated using algorithms involving generating random permutations, i.e., random ordering of the numbers 1, 2, . . . , n, for some fixed "n" of the values among all other possible locations (Joost et al., 2017).

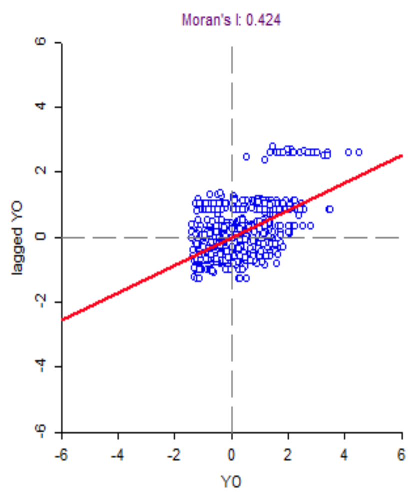


***Figure 4-7.** Moran's I scatter plot of on-farm trials*

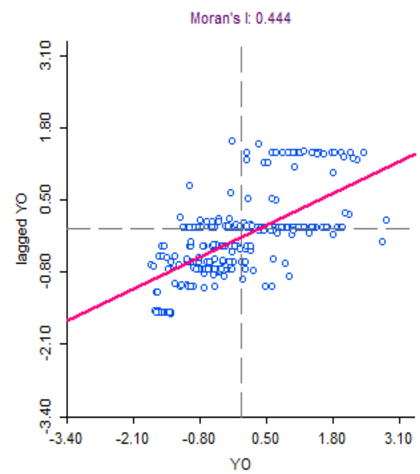


***Figure 4-8.** Moran's I scatter plot of on-station trials*

More precisely, we use stochastic computer simulation, often called Monte Carlo simulation, which includes some randomness in the underlying model rather than deterministic computer simulation (Rubinstein and Kroese, 2016). Figure 4-9A and B show pseudo-p-values that are less than 0.001. In addition, on the histograms, around the yield mean (*white bar*) the observed yield values from farms and stations do not resemble the average yields of their neighbors, and I_m values (*green bar*), both on-farm and on-station, are different from the rest of the distribution. Consequently, in this study, it can be concluded that the maize yield variables analyzed are significantly spatially autocorrelated and the risk of rejecting the null hypothesis (H_0) is low. However, the values of I_m not being more than 0.5 could be the result of some spatial gaps that existed in the dataset (Bua et al., 2020).

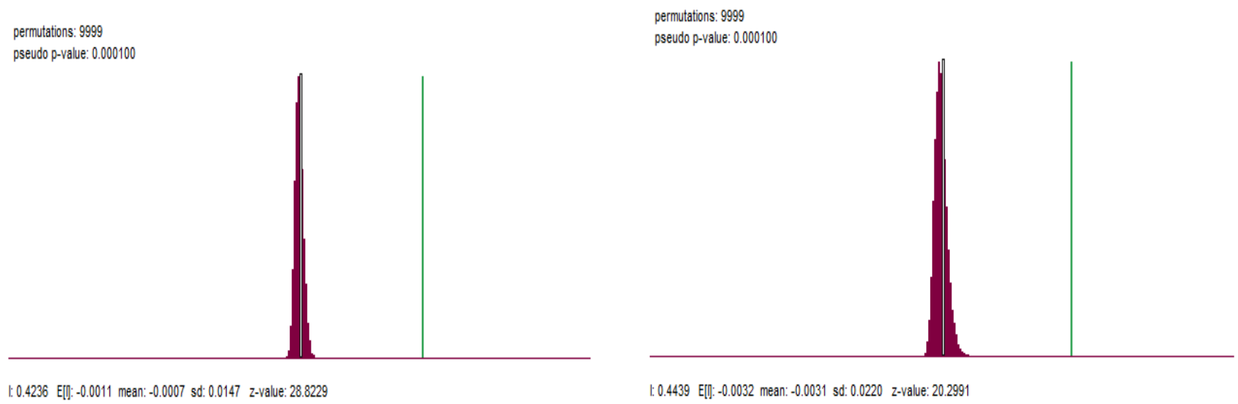


Figure 4-9. Histogram of permutation and p-value of observed yield and yield of neighbors on-farm (left) and on-station (right)

To predict the grain yields for all the unknown locations over the entire study area grid, the spatial structure in the data points of the grain yield on-farm and on-station were evaluated through their semivariograms (Eq. 3-17). Figure 4-10A shows the semivariograms for observed yield from on-station trials. As shown, the range of spatial dependence has a low and weak variation, 49 km. The semivariogram has 115 nuggets. Spatial continuity between neighboring points and the low range of spatial dependence indicates that this continuity disappears very fast. The semivariogram of the on-station yield trials is different from the on-farm yield trials. That of the on-farm trials shows a pure nugget effect.

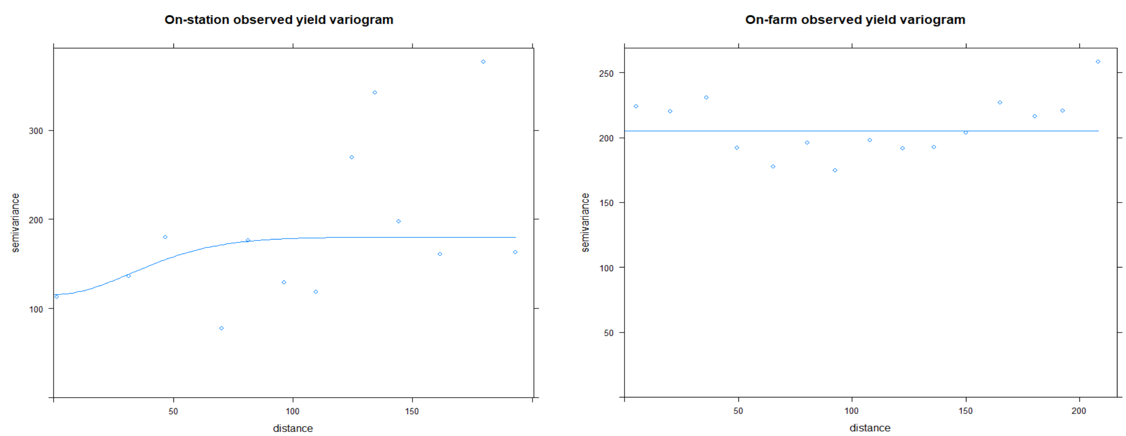


Figure 4-10. Variogram of observed yield (A) on-station and (B) on-farm

This could be explained by the fact that the factors affecting the variability of yield are not the same. The exhibited spatial autocorrelation through I_m analysis seems to be in contradiction with the variogram of Figure 4-10B. Bua et al. (2020) stated that the various sources of secondary data points could explain the nugget effect observed, and this effect might be misleading. In Figure 4-12 and Figure 4-13, a significant difference between NPK nutrient-containing soil spatial variation and NPK fertilizer applied semivariograms can be seen. Among the three nutrients, on-farm soil organic nitrogen contain has the highest range (145 km), compared with 72 km for on-station soil organic nitrogen-containing semivariograms. P- and K-containing soil semivariograms on-farm and on-station have a low range.

The semivariogram for soil NPK content shows almost zero nugget effect value. The zero nugget effect value indicates a very smooth spatial continuity between neighboring points. However, this smooth spatial continuity must make the application of fertilizers site-specific. In Figure 4-12 and Figure 4-13, semivariograms of NPK fertilizer application show the pure nugget effect. This means that fertilizer application, whether on-farm or on-station, is not applied as a function of soil NPK content dynamic. Remembering that the semivariogram is a result of the mean squared differences between the neighboring values (Vieira and Gonzalez, 2003), Figure 4-12 and Figure 4-13 show that values are not close and are not similar. The pure nugget effect observed in the NPK fertilizer spatial application again emphasizes the importance of the role of soil testing before fertilizer recommendation. The similarity between the soil exchangeable potassium and the potassium fertilizer application variograms could reveal that potassium is not a very limited nutrient in maize production in Ghana. Thus, there is a need for more research to verify these results.

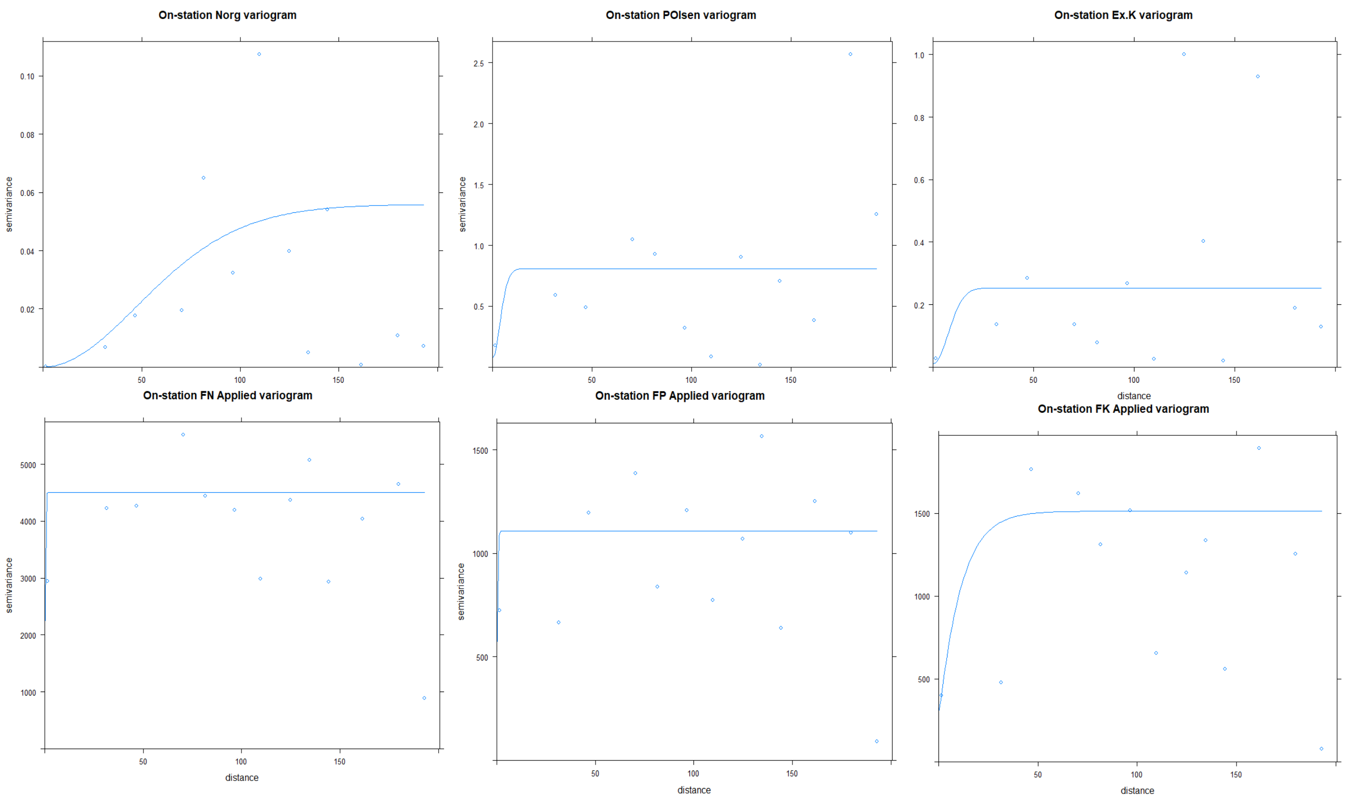


Figure 4-11. Comparison of soil NPK and NPK fertilizer application semivariograms on-station

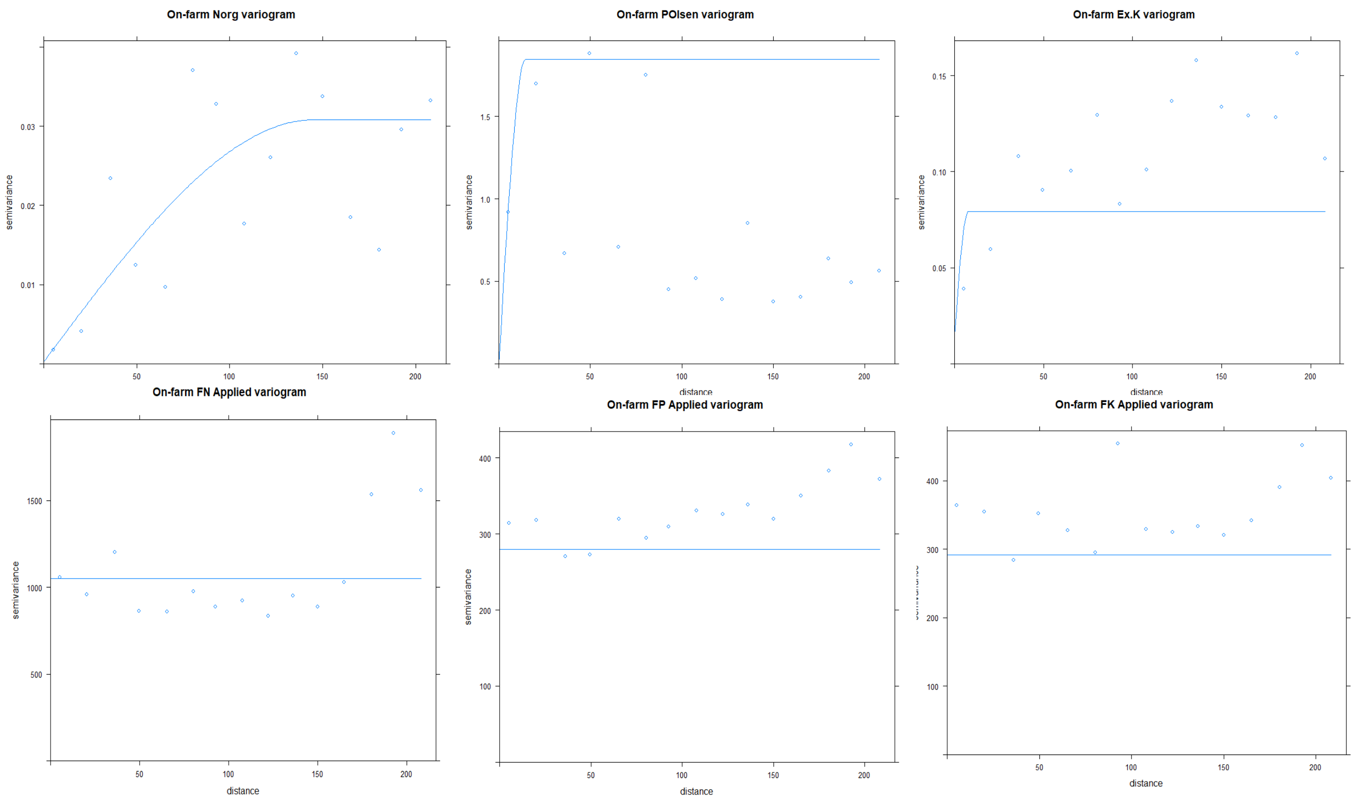


Figure 4-12. Comparisons of soil NPK and NPK fertilizer application semivariograms on-farm

4.3.2 Maize Yield Prediction Maps

The maps in Figure 4-13 are derived from several advanced statistical methods. They show a spatial response of the predicted yield distribution of maize to environmental conditions, physicochemical soil properties, and fertilizers. The more yellow the map, the higher the yield and the more the response to the covariates is positive, i.e., that it contributes to the increase in maize yield. However, the darker the blue, the lower the predicted return and the more the direct and indirect effects of the covariates tend to reduce the return and increase the yield gap.

RFsp seems to smooth the spatial pattern, which is possibly the result of averaging the trees in the random forest (Hengl et al., 2018). The on-farm DYPM shows a generally high yield in the southeast and southwest. The expected yields in the regions of these agroecological zones varies between 2,500 and 5,000 kg/ha. The expected yield starts to become low in Bono, in the northern and eastern regions, where it reaches more than 2,500 kg/ha. On-farm DYPMs from RF and MLR-AIC models are very similar. That of QUEFTS shows a slightly different distribution in staining; nevertheless, it predicts the yields following the same concentration gradient. Table 4-9 shows that, among the models from which on-farm DYPMs are derived, the MLR-AIC-RFsp map is the most accurate based on R^2. The large discrepancy between the R^2 of the RF-RFsp map, MLR-AIC-RFsp maps, and QUEFTS-RFsp map is due to the number of covariates used in the model. On-station, RFsp predicted yield up to 4,000 kg/ha in the two situations. The gradient of coloration starts from red-yellow-dark pink, which is more pronounced in the southwestern and southeastern part of the country, and becomes bluer going up to the north of the country (Guinea Savannah, Sudan Savannah), similar to the on-farm DYPM. This perfectly reflects what happens in reality in a general manner in Ghana. The increase of maize yield follows the rainfall pattern, even if in this study MLR-AIC showed that there is an inverse correlation between rainfall and observed yield; nevertheless, it remains an important factor in the variability of the response of maize. Other studies have underscored the importance of rainfall variability for yield stability (MacCarthy et al., 2017; Kyei-Mensah et al., 2019). Müller et al. (2011) reported that rainfall variability was the main cause of yield variability in SSA. These prediction maps created using biophysical variables show that there is still room to increase maize yields if fertilizer application rates are increased and adapted to the AEZ environmental conditions.

Table 4-9. Accuracy of DYPMs

Model	On-Farm		On-Station	
	RMSE	R^2	RMSE	R^2
QUEFTS-RFsp	821	22%	904	47%
MLR-AIC-RFsp	637	58%	562	72%

Table 4-9 shows that, on-station, the MLR-AIC-RFsp map is more accurate than QUEFTS-RFsp maps based on the R^2 and RMSE indicators. The MLR-AIC-RFsp map represents 72% of the variability in maize yield. This suggests that this map is the most accurate and closest to reality.

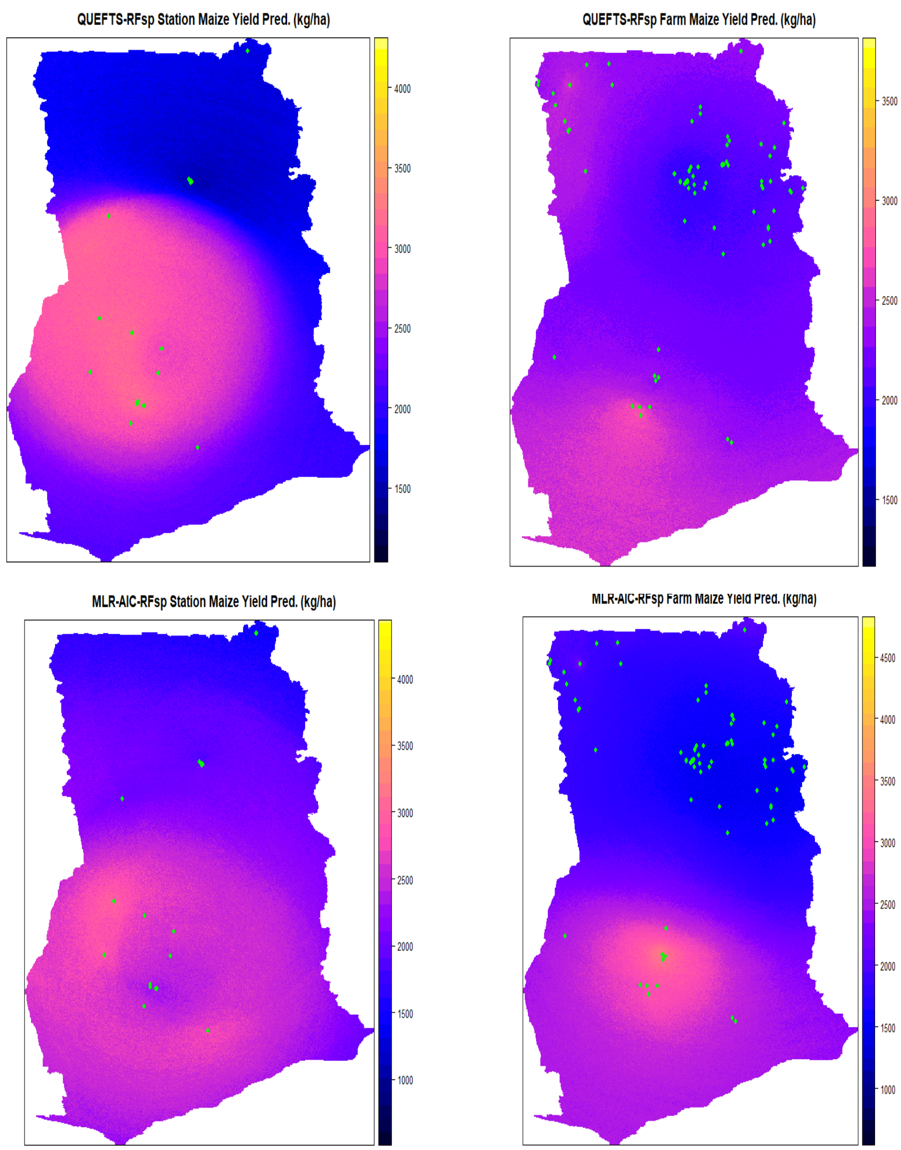


Figure 4-13. On-station and on-farm maize yield prediction maps

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

A total of 1,198 data points from old and peer-reviewed publications were used to assess and conduct a spatial analysis of the variability of maize yield responses to fertilizer application in Ghana in various regions, including Ashanti, East, North, Upper West, Upper East, and North East. The evaluation of the variability of maize yield was carried out using QUEFTS models based partly on empirical relations and partly on theoretical relations, complemented with Multiple Linear Regression. Spatial analysis was carried out through the calculation of the Moran's index and the construction of semivariograms from the observed maize yields; nitrogen, phosphorus, and potassium from the soil; and NPK fertilizers applied. This resulted in the creation of a predicted-yield map using the random forest for spatial prediction framework (RFsp), in which model-predicted yield point buffer distances are used as explanatory variables, thus incorporating geographic proximity effects into the map prediction process. The fitted measures and the coefficient of determination, as well as the mean root square error, were used to justify the level of explanation of yield variability by the prediction models.

The study shows that the QUEFTS soil fertility model was insufficient to explain on-station and on-farm trial yields. Thus, soil fertility was not the only limiting factor for maize production in Ghanaian AEZs. Other factors, identified through advanced statistical regression analysis and random forest algorithm, included root zone depth, water-holding capacity, rainfall, temperature, elevation, and soil texture and structure, which are as important as soil nutrients and fertilizer applied.

Multiple Linear Regression enhanced by Akaike Information Criterion led to the design of 13 models for on-station trials and 13 models for on-farm trials. MLR taking into account weather variables and soil physical and chemical properties explained over 50% of the variation for both the on-station and on-farm trials, with the AIC algorithm identifying the significant variables. Generally, the situation may improve with increasing pH. On-station trial pH values are not bad (around 6); however, the strong negative coefficient will need to be explored, especially since the pH range in the data may be narrow. The results show that observed yield variability with fertilization is explained by this comprehensive set of variables. These results suggest that attention should also be focused on factors such as temperature and root zone depth. Soil chemical properties alone explain observed yields poorly, using both QUEFTS and MLR. It can be inferred that climate change may heavily impact maize yield and variability.

Fertilizer use has a strongly significant effect on observed yield, and increased fertilizer rates tend to stabilize yield variation. Consequently, the Government of Ghana should pursue the campaign for fertilizer application on-field by farmers. However, blanket fertilizer application may no longer be the most effective fertilization strategy in Ghana, as revealed by spatial analysis variograms. Fertilizer formulations and rate of application should be at least AEZ specific because of climate and soil physical and chemical properties. Therefore, a soil-crop simulation model that takes the major factors influencing crop production into account, such as solar radiation, planting density, and soil nutrient and water dynamics, would be more appropriate for simulating yields in the study area and enhancing understanding of spatial-temporal variability of crop yields.

Spatial analysis showed significant spatial variability in yield within AEZs, offering the possibility of managing yield variability through modulation of applied fertilizer rates by considering the direct and indirect positive or negative effects that covariables identified in this study may have on physiological efficiency. On-station DYPM from MLR more accurately predicted the variability in maize yield. DYPM from MLR was the most realistic and could be used to predict maize yield. In two cases, the maps from the models showed a lot of similarity in the representation of the yield gradient of maize, which is higher in the AEZs of the west and south than in those of the center and north. For successful adoption of fertilizers by farmers, research institutions, the private sector, non-governmental organizations, and the Government of Ghana, specific recommendations must not only be based on soil tests but also must take into consideration DYPM. DYPM and *Digital Soil Prediction Mapping* are complementary decision-driven tools for farm management. Studying the variability of the yield response in Ghana can help farmers tailor fertilizer management by taking into account the direct and indirect effects of covariables on yield.

The applied methodology in this study demonstrates the potential to estimate yield with increasing accuracy for identification of location-specific fertilizer recommendations based on climatic characteristics and soil chemical and physical conditions. The converging results in the spatial maps using the different methods justifies the logic of the methods, and their results correspond to the reality of the field. However, an optimization method should be able to improve this and help in understanding all the inverse correlations in on-station trials, which undoubtedly hide what is hampering yield increasing in Ghana. In addition, the temporal and spatial logistical operations for fertilizer distribution, as well as the timely delivery of the right fertilizer, are key elements in increasing farm yields and incomes, even if all the covariates that were found to be important in this study are well managed.

REFERENCES

Abunyewa, A.A., Mercer-Quarshie, H., 2004. Response of maize to magnesium and zinc application in the semi-arid zone of West Africa. Asian Journal of Plant Sciences 3, 1-5.

Addo, A., Amponsah, S.K., 2018. Present status and future prospects of agricultural machinery industry in Ghana. AMA, Agric. Mech. Asia, Africa Lat. Am. 49, 87-94.

Adhinarayanan, R., 2017. Enhancing soil organic carbon through tank silt application. Global Symposium on Soil Organic Carbon, pp. 21-23.

Adu-Gyamfi, R., Agyin-Birikorang, S., Tindjina, I., Ahmed, S.M., Twumasi, A.D., Avornyo, V.K., Singh, U., 2019. One-time fertilizer briquettes application for maize production in savanna agroecologies of Ghana. Agronomy Journal 111, 3339-3350.

Adu, G.B., Abdulai, M.S., Alidu, H., Nustugah, S.K., Buah, S.S., Kombiok, J.M., Obeng-Antwi, K., Abudulai, M., Etwire, P.M. 2014. Recommended production practices for maize in Ghana.

AfricaFertilizer.org. 2021. AFO, Ghana, p. www.AfricaFertilizer.org.

Ahene, A.A. 2003. Effect of climate change on maize supply in Ghana, 1970-2002. University of Ghana, pp. 1-68.

Albert, H., Runge-Metzger, A., 1995. Management of natural resources: systems approaches to striga control in sub-Saharan Africa. Journal of Farming Systems Research-Extension 5, 91-106.

Andersson, S., Nilsson, S.I., Saetre, P., 2000. Leaching of dissolved organic carbon (DOC) and dissolved organic nitrogen (DON) in mor humus as affected by temperature and pH. Soil Biology and Biochemistry 32, 1-10.

Antwi, A., Duker, A., Fosu, M., Abaidoo, R.C., 2017. Simulation of major soil nutrients requirement for maize production using the QUEFTS model in the Northern Region of Ghana. Direct Research Journal of Agriculture and Food Science 5, 133-140.

Antwi, M., Duker, A.A., Fosu, M., Abaidoo, R.C., Pirasteh, S. 2016. Geospatial approach to study the spatial distribution of major soil nutrients in the Northern Region of Ghana. Cogent Geoscience 2, 1201906.

Asante, B.O., Temoso, O., Addai, K.N., Villano, R.A., 2019. Evaluating productivity gaps in maize production across different agroecological zones in Ghana. Agricultural Systems 176, 102650.

Atakora, W.K., Fosu, M., Marthey, F., 2014. Modeling maize production towards site specific fertilizer recommendation in Ghana. Global Journal of Science Frontier Research: D Agriculture and Veterinary 14, 71-82.

Banson, K.E., Nguyen, N.C., Bosch, O.J.H., 2014. Systemic intervention to tackle the constraints and challenges facing stakeholders and the performance of the agricultural sector in Ghana. 31-36.

Barimah, P.T., 2014. Impact of climate change on maize production in Ghana. A review. Journal of Agricultural Science and Applications 03, 89-93.

Bationo, A., Fening, J.O., Kwaw, A., 2018. Assessment of soil fertility status and integrated soil fertility management in Ghana. In: Bationo, A., Youl, S., Fening, J.O., Ngaradoum, D., Lompo, F. (Eds.), Improving the Profitability, Sustainability and Efficiency of Nutrients

Through Site Specific Fertilizer Recommendations in West Africa Agro-Ecosystems. Springer International Publishing AG 2018, pp. 93-138.

Bertolini, F., Galimberti, G., Calò, D.G., Schiavo, G., Matassino, D., Fontanesi, L., 2015. Combined use of principal component analysis and random forests identify population-informative single nucleotide polymorphisms: application in cattle breeds. Journal of Animal Breeding Genetics 132, 346-356.

Bertolini, F., Galimberti, G., Schiavo, G., Mastrangelo, S., Di-Gerlando, R., Strillacci, M.G., Bagnato, A., Portolano, B., Fontanesi, L., 2018. Preselection statistics and Random Forest classification identify population informative single nucleotide polymorphisms in cosmopolitan and autochthonous cattle breeds. Animal 12, 12-19.

Bevans, R., 2020. An introduction to the Akaike information criterion. Scribbr. https://www.scribbr.com/statistics/akaike-information-criterion/.

Bhusal, S., Chapagain, E., 2020. Threats of fall armyworm (Spodoptera frugiperda) incidence in Nepal and it’s integrated management-A review. Journal of Agriculture and Natural Resources 3, 345-359.

Bindraban, P.S., Stoorvogel, J.J., Jansen, D.M., Vlaming, J., Groot, J.J.R., 2000. Land quality indicators for sustainable land management: proposed method for yield gap and soil nutrient balance. Agriculture, Ecosystems and Environment, 103-112.

Bindraban, P.S., Verhagen, A.J., Uithol, P.W.J., Henstra, P., 1999. A land quality indicator for sustainable land management: the yield gap: The case of sub-Saharan Africa. AB, Wageningen, Rapport 106. AB, Wageningen, pp. 1-38.

Boulesteix, A.L., Janitza, S., Kruppa, J., König, I.R., 2012. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisciplinary Reviews: Data Mining Knowledge Discovery 2, 493-507.

Braimoh, A.K., Vlek, P.L.G., 2006. Soil quality and other factors influencing maize yield in northern Ghana. Soil Use and Management 22, 165-171.

Brust, G.E., 2019. Management strategies for organic vegetable fertility. Safety and practice for organic food. Elsevier, pp. 193-212.

Bua, S., El-Mejahed, K., MacCarthy, D., Adogoba, D.S., Kissiedu, I.N., Atakora, W.K., Fosu, M., S., B.P., 2020. Yield responses of maize to fertilizers in Ghana. IFDC FERARI No. 2, pp. 1-65.

Chapoto, A., Ragasa, C., 2013. Moving in the right direction ? Maize productivity and fertilizer use and use intensity in Ghana. IFPRI Discussion Paper 01314, 1-32.

Chapoto, A., Tetteh, F., 2014. Examining the sense and science behind Ghana's current blanket fertilizer recommendation. Washington D.C.: International Food Policy Research Institute (IFPRI), pp. 1-23.

Cicin-Sain, I., 2018 Population growth and food: A systems perspective. Food Systems. 01/07/2021, pp. https://sustainablefoodtrust.org/articles/population-growth-and-food-a-systems-perspective/

Cooper, P.J.M., 1979. The association between altitude, environmental variables, maize growth and yields in Kenya. 93, 635-649.

Curtin, D., Campbell, C.A., Jalil, A., 1998. Effects of acidity on mineralization: pH-dependence of organic matter mineralization in weakly acidic soils. Soil Biology and Biochemistry 30, 57-64.

Danso, I., Larbi, E., Andoh-Menash, E., Ribeiro, P.F., Adjarko, I.K., Agyarko-Mintah, E., 2020. Nitrogen fertilizer management strategy for oil palm-maize intercropping system in the semi-deciduous forest zone of Ghana. Ghana Journal of Agricultural Science 55, 75-86.

Darfour, B., Rosentrater, K.A., 2016a. Agriculture and food security in Ghana. 2016 ASABE Annual International Meeting.

Darfour, B., Rosentrater, K.A., 2016b. Maize in Ghana: An overview of cultivation to processing. 2016 ASABE International Meeting.

De Wit, C.T., Goudriaan, J., 1974. Simulation of ecological processes. Centre for Agricultural Publishing and Documentation, Wageningen, the Netherlands,.

Diao, X., Hazell, P., Kolavalli, S., Resnick, D., 2019. Ghana's economic and agricultural transformation: Past performance and future prospects. ISBN 978-0-19-884534-8 Oxford University Press.

Diest, V.A., 1980. Factor affecting the availability of potassium in soils. In: IPI (Ed.), Potassium in the Soil_Plant Root System (extracted from: Proc. 11th Congr. Int. Potash Institute p. 43-108 119781). IPI Research Topics, CH-3084 Worblaufen-Bern/Switzerland, pp. 37-60.

Dutta, S., Chakraborty, S., Goswami, R., Banerjee, H., Majumdar, K., Li, B., Jat, M.L., 2020. Maize yield in smallholder agriculture system—An approach integrating socio-economic and crop management factors. PloS One 15, 1-23.

EPA, 2000. Climate Change Vulnerability and Adaptation Assessment of Coastal Zone of Ghana. Final Report. Environmental Protection Agency (EPA) / Ministry of Environment Science and Technology, Accra.

EU, 2019. Global food supply and demand Consumer trends and trade challenges. EU Agricultural Markets Briefs. European Union, pp. https://ec.europa.eu/info/sites/default/files/food-farming-fisheries/farming/documents/market-brief-food-challenges-sep2019_en.pdf.

Ezui, K.S., Franke, A.C., Ahiabor, B.D.K., Tetteh, F.M., Sogbedji, J., Janssen, B.H., Mando, A., Giller, K.E., 2017. Understanding cassava yield response to soil and fertilizer nutrient supply in West Africa. Plant and Soil 420, 331-347.

Fang, J., Su, Y., 2019. Effects of soils and irrigation volume on maize yield, irrigation water productivity, and nitrogen uptake. Scientific reports 9, 1-11.

FAO, 2019. Promoting sustainable development for agricultural transformation, economic growth and rural development. FAO website. FAO.

FAO/OECD, 2018. Food Security and Nutrition: Challenges for Agriculture and the Hidden Potential of Soil. FAO, pp. 1-33.

FAOStat, 2021. Corporate Statistical Database, agriculture data. Food and Agriculture Organization of the United Nations, p. http://www.fao.org/faostat/en/ \# data/QC

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37, 4302-4315.

Fosu-Mensah, B.Y., MacCarthy, D.S., Vlek, P.L.G., Safo, E.Y., 2012. Simulating impact of seasonal climatic variation on the response of maize (*Zea mays L.*) to inorganic fertilizer in sub-humid Ghana. Nutrient Cycling in Agroecosystems 94, 255-271.

Fox, J., Weisberg, S., 2019. An R Companion to Applied Regression. Sage, Thousand Oaks CA.

Genuer, R., Poggi, J.-M., 2020. Random Forests with R. Springer.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food security: The challenge of feeding 9 billion people. Sci. 327(5967), 812-818.

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. The R Journal 8, 204-218.

GSS, 2019. Rebased 2013-2018 Annual Gross Domestic Product. April 2019 Ed. Ghana Statistical Service (GSS), Accra, pp. 1-11.

Guilpart, N., Grassini, P., van Wart, J., Yang, H., van Ittersum, M.K., van Bussel, L.G.J., Wolf, J., Claessens, L., Leenaars, J.G.B., Cassman, K.G., 2017. Rooting for food security in Sub-Saharan Africa. Environmental Research Letters 12, 1-8.

GYGA, 2021. Agriculture in Ghana. GYGA (Global Yield Gap Atlas). University of Nebraska-Lincoln and Wageningen University & Research, p. https://www.yieldgap.org/ghana

Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. International Journal of Climatology, 623-642.

Hengl, T., 2012. Welcome to Global Soil Information Facilities project! World Soil Information-ISRIC, pp. http://gsif.r-forge.r-project.org/

Hengl, T., Heuvelink, G.B., Kempen, B., Leenaars, J.G., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L., Tondoh, J.E., 2015. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. PLoS One 10, 1-26.

Hengl, T., Leenaars, J.G.B., Shepherd, K.D., Walsh, M.G., Heuvelink, G.B.M., Mamo, T., Tilahun, H., Berkhout, E., Cooper, M., Fegraus, E., Wheeler, I.a., 2017. Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. Nutr. Cycl. Agroecosystems, 1-26.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6 6, 1-49.

Heuvelink, G., Rossiter, D., 2021. Tutorial: Heavy metals in the Geul valley. In: Heuvelink, G., Rossiter, D. (Eds.), ISRIC Spring School. Version 1.0. Copyright © ISRIC - World Soil Information., Netherland, p. 14.

Hong, S., Gan, P., Chen, A., 2019. Environmental controls on soil pH in planted forest and its response to nitrogen deposition. Environmental research 172, 159-165.

Hsiao, T.C., Heng, L., Steduto, P., Rojas-Lara, B., Raes, D., Fereres, E., 2009. AquaCrop-the FAO crop model to simulate yield response to water: III. Parameterization and testing for maize. Agronomy Journal 101, 448-459.

Israelsson, J., Black, E., Neves, C., Torgbor, F.F., Greatrex, H., Tanu, M., Lamptey, P.N.L., 2020. The spatial correlation structure of rainfall at the local scale over southern Ghana. Journal of Hydrology: Regional Studies 31, 100720.

ISRIC, Africa SoilGrids nutrients. ISRIC, Wageningen, 6700AJ, Netherlands, p. https://data.isric.org/geonetwork/srv/eng/catalog.search#/search

Jamada, B., 2019. Validation of the QUEFTS model: A case study of field‒specific fertilizer recommendations to maize smallholder farmers in western Kenya. Plant Production Systems (PPS). Wageningen University, p. 42.

Janssen, B.H., Guiking, F.C.T., 1990. Modelling the response of crops to fertilizers. Plant nutrition - physiology and application, 699-703.

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

Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.M., Gerber, J.S., Reddy, V.R., Kim, S.H., 2016. Random Forests for Global and Regional Crop Yield Predictions. PLoS One 11, e0156571.

Jiang, P., Thelen, K.D., 2004. Effect of soil and topographic properties on crop yield in a North-Central corn–soybean cropping system. 96, 252-258.

Jones, J.W., 1993. Decision support systems for agricultural development. Systems approaches for agricultural development. Springer, pp. 459-471.

Jones, J.W., Tsuji, G.Y., Hoogenboom, G., Hunt, L.A., Thornton, P.K., Wilkens, P.W., Imamura, D.T., Bowen, W.T., Singh, U., 1998. Decision support system for agrotechnology transfer: DSSAT v3. Understanding options for agricultural production. Springer, pp. 157-177.

Joost, S., Soutter, M., Kouamé, F., Sall, A., 2017. Systèmes d'Information Géographique 2. Lausanne, Presses Polytechniques et Universitaires Romandes (PPUR), Lausanne.

Kang, F., Wang, Z., Xiong, H., Li, Y., Wang, Y., Fan, Z., Zhao, H., Kuang, D., Chen, Z., Wang, J., He, X., Chen, X., Shi, X., Zhang, Y., 2020. Estimation of watermelon nutrient requirements based on the QUEFTS Model. Agronomy 10, 10111776.

Kashaigili, J.J., Levira, P., Liwenga, E., Mdemu, M.V., 2014. Analysis of Climate variability, perceptions and coping strategies of Tanzanian Coastal Forest dependent communities. American Journal of Climate Change 03, 212-222.

Kassambara, A., 2018. Machine learning essentials: Practical guide in R. Sthda.

Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., 2003. An overview of APSIM, a model designed for farming systems simulation. European journal of agronomy 18, 267-288.

Kihara, J., Nziguheba, G., Zingore, S., Coulibaly, A., Esilaba, A., Kabambe, V., Njoroge, S., Palm, C., Huising, J., 2016. Understanding variability in crop response to fertilizer and amendments in sub-Saharan Africa. Agric. Ecosyst. Environ. 229, 1-12.

Krause, P., Boyle, D.P., and Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci., 89-97.

Kravchenko, A.N., Robertson, G.P., 2007. Can Topographical and Yield Data Substantially Improve Total Soil Carbon Mapping by Regression Kriging? Agronomy Journal, 12-17.

Kyei-Mensah, C., Kyerematen, R., Adu-Acheampong, S., 2019. Impact of rainfall variability on crop production within the Worobong Ecological Area of Fanteakwa District, Ghana. Advances in Agriculture 2019, 1-7.

Lamos-Díaz, H., Puentes-Garzón, D.E., Zarate-Caicedo, D.A., 2020. Comparison between machine learning models for yield forecast in cocoa crops in Santander, Colombia. Revista Facultad de Ingeniería 29, e10853-e10853.

Leenaars, J.G.B., Claessens, L., Heuvelink, G.B.M., Hengl, T., Ruiperez Gonzalez, M., van Bussel, L.G.J., Guilpart, N., Yang, H., Cassman, K.G., 2018a. Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa. Geoderma 324, 18-36.

Leenaars, J.G.B., Ruiperez González, M., Kempen, B., 2018b. Extrapolation of fertilizer nutrient recommendations for major food crops in West Africa; a proof of concept (with dataset). In: (draft), P.r. (Ed.). IFDC/USAID-West Africa Fertilizer Program, Accra, Wageningen, the Netherlands, pp. 7-81.

Leng, G., Huang, M., 2017. Crop yield response to climate change varies with crop spatial distribution pattern. Sci Rep 7, 1463.

Linneman, H., Hoogh, J.d., Keyzer, M.A., Heemst, H.D.J.v., 1979. MOIRA: Model of International Relations in Agriculture. Report of the project group "Food for a doubling world population. North Holland Publishing Comp., Amsterdam, Netherlands.

Liu, D., Wang, Z., Zhang, B., Song, K., Li, X., Li, J., Li, F., Duan, H., 2006. Spatial distribution of soil organic carbon and analysis of related factors in croplands of the black soil region, Northeast China. Agriculture, Ecosystems and Environment 113, 73-81.

Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L., 2008. Prioritizing climate change adaptation needs for food security in 2030. Science 319, 607-610.

Lovelace, R., Nowosad, J., Muenchow, J., 2021. Ecology. Geocomputation with R. CRC Press.

MacCarthy, D.S., Adiku, S.G., Freduah, B.S., Kamara, A.Y., Narh, S., Abdulai, A.L., 2017. Evaluating maize yield variability and gaps in two agroecologies in northern Ghana using a crop simulation model. South African Journal of Plant and Soil 35, 137-147.

Mack, C., 2016. From Data to Decisions: Measurement, Uncertainty, Analysis, and Modeling. University of Texas at Austin, p. http://www.lithoguru.com/scientist/statistics/course.html

Martins, J.D.L., Soratto, R.P., Fernades, A., Dias, P.H., 2018. Phosphorus fertilization and soil texture affect potato yield. Revista Caatinga 31, 541-550.

Masika, D., 2016. Using Dssat-Ceres Maize Model to Estimate District-Level Yields in Northern Ghana. Department of Soil Science. University Of Ghana, Ghana.

McCown, R.L., Hammer, G.L., Hargreaves, J.N.G., Holzworth, D.P., Freebairn, D.M., 1996. APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. Agricultural systems 50, 255-271.

Méneroux, Y., 2018-2019. Introduction à la Géostatistique. In: Méneroux, Y. (Ed.), Cours au Mastère Spécialisé DésiGéo. GEOLigne, Bibliothèque de ressources pédagogiques de l'ENSG, Ecole Nationale des Sciences Geographiques, pp. 1-189.

MiDA, 2010. Investment opportunity in Ghana: maize, rice, and soybean. Millennium Development Authority (MiDA) in conjunction with the United States Millennium Challenge Corporation, Accra, Ghana. , pp. 1-23.

MoFA, 2012. Ghana’s Private Sector Investment Plan for Agricultural Development. In: (SRID), R.a.I.D. (Ed.), Accra, Ghana, pp. 1-67.

MoFA, 2020. Meet the press; Presentation by the Minister For Food and Agriculture. Minist. Food Agric. Stat. Res. Inf. Dir. MoFA, Accra, Ghana.

Mohammed, Y.S., Mokhtar, A.S., Bashir, N., Saidur, R., 2013. An overview of agricultural biomass for decentralized rural energy in Ghana. Renew. Sustain. Energy Rev. 20, 15-25.

Moran, P.A., 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society: Series B (Methodological) 10, 243-251.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50, 885-900.

Mueller, N.D., Gerber, J.S., Johnston, M., Ray, D.K., Ramankutty, N., Foley, J.A., 2012. Closing yield gaps through nutrient and water management. Nature 490, 254-257.

Mugwe, J., Mugendi, D., Kungu, J., Muna, M.M., 2009. Maize yields response to application of organic and inorganic input under on-station and on-farm experiments in Central Kenya. Experimental Agriculture 45, 47-59.

Mugwe, J., Ngetich, F., Otieno, E.O., 2019. Integrated Soil Fertility Management in Sub-Saharan Africa: Evolving Paradigms Toward Integration. Zero Hunger. Encyclopedia of the UN Sustainable Development Goals, Springer, Cham.

Mulder, I., 2000. Soil Fertility: QUEFTS and Farmer's Perceptions. IIED, Amsterdam.

Müller, C., Cramer, W., Hare, W.L., Lotze-Campen, H., 2011. Climate change risks for African agriculture. Proceedings of the National Academy of Sciences 108, 4313-4315.

Munialo, S., Dahlin, A.S., Onyango M, C., Oluoch-Kosura, W., Marstorp, H., Öborn, I., 2020. Soil and management-related factors contributing to maize yield gaps in western Kenya. Food Energy Security 9, e189.

Nevavuori, P., Narra, N., Linna, P., Lipping, T., 2020. Crop Yield Prediction Using Multitemporal UAV Data and Spatio-Temporal Deep Learning Models. Remote Sensing 12, 4000.

Njoroge, K.S., 2019. Explaining variability in maize yield responses to nutrient applications in smallholder farms of western Kenya. C.T. de Wit Graduate School of Production Ecology and Resource Conservation (PE&RC). Wageningen University & Research, The Netherlands, p. 169.

Nkrumah, F., Klutse, N.A.B., Adukpo, D.C., Owusu, K., Quagraine, K.A., Owusu, A., Gutowski, W., 2014. Rainfall variability over Ghana: model versus rain gauge observation. International Journal of Geosciences 7, 673-683.

O'Halloran, I.P., Kachanoski, R.G., Stewart, J.W.B., 1985. Spatial variability of soil phosphorus as influenced by soil texture and management. Canadian journal of soil science 65, 475-487.

Oliver, M.A., 2010. Geostatistical Applications for Precision Agriculture. Springer, Dordrecht Heidelberg London New York.

Onduru, D.D., Du Preez, C.C., 2007. Spatial and temporal aspects of agricultural sustainability in the semi-arid tropics: a case study in Mbeere district, Eastern Kenya. Tropical Science 47, 134-148.

Ovalles, F.A., Collins, M.E., 1986. Soil-landscape relationships and soil variability in north central Florida. Soil Science Society of America Journal 50, 401-408.

Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., Athanasiadis, I.N., 2021. Machine learning for large-scale crop yield forecasting. Agricultural Systems 187.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences 30, 683-691.

Peprah, K., 2012. Rainfall and temperature correlation with crop yield: The case of Asunafo Forest, Ghana.

Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2009. AquaCrop—The FAO crop model to simulate yield response to water: II. Main algorithms and software description. Agronomy Journal 101, 438-447.

Rahman, A.N., Larbi, A., Kotu, B., Asante, M.O., Akakpo, D.B., Mellon-Bedi, S., Hoeschle-Zeledon, I., 2021. Maize–legume strip cropping effect on productivity, income, and income risk of farmers in northern Ghana. Agronomy Journal, 1-12.

Ravensbergen, A.P.P., Chamberlin, J., Craufurd, P., Shehu, B.M., Hijbeek, R., 2021. Adapting the QUEFTS model to predict attainable yields when training data are characterized by imperfect management. Field Crops Research 266.

Ren, T., Zou, J., Wang, Y., Li, X.K., Cong, R.H., Lu, J.W., 2015. Estimating nutrient requirements for winter oilseed rape based on QUEFTS analysis. The Journal of Agricultural Science 154, 425-437.

Ribeiro Jr, P.J., Diggle, P., 2006. geoR: Analysis of Geostatistical Data. The geoR package, version, 1-6.

Rubinstein, R.Y., Kroese, D.P., 2016. Simulation and the Monte Carlo method. John Wiley & Sons, Inc., Hoboken, New Jersey, Canada.

Sadras, V.O., Calvino, P.A., 2001. Quantification of grain yield response to soil depth in soybean, maize, sunflower, and wheat. Agronomy Journal 93, 577-583.

Safo, E.Y., 1990. Fertilizer requirement of some major crops in Ghana. 20.

Salima, B.A., Bellefon, M.D., 2018. Spatial autocorrelation indices. Handbook of Spatial Analysis: Theory Aplication with R, 51-68.

Sanchez, P.A., Shepherd, K.D., Soule, M.J., Place, F.M., Buresh, R.J., Izac, A.M.N., Uzo-Mokwunye, A., Kwesiga, F.R., Ndiritu, C.G., Woomer, P.L., 1997. Soil fertility replenishment in Africa: an investment in natural resource capital. Replenishing soil Fertil. Africa 51, 1-46.

Sanginga, N., Woomer, P.L., 2009. Integrated soil fertility management in Africa: Principles, practices, and developmental process. CIAT.

Sattaria, S.Z., van Ittersuma, M.K., Bouwmanb, A.F., Smitd, A.L., Janssena, B.H., 2014. Crop yield response to soil fertility and N, P, K inputs in different environments: Testing and improving the QUEFTS model. Field Crops Research 157, 35-46.

Sauerborn, J., Kranz, B., Mercer-Quarshie, H., 2003. Organic amendments mitigate heterotrophic weed infestation in savannah agriculture. Applied Soil Ecology 23, 181-186.

Sawyer, J.E., Mallarino, A.P., 1999. Differentiating and understanding the Mehlich 3, Bray, and Olsen soil phosphorus tests. Agronomy Conference Proceedings and Presentations 12.

Scheiterle, L., Birner, R., 2018. Assessment of Ghana's comparative advantage in maize production and the role of fertilizers. Sustainability 10.

Scheiterle, L., Häring, V., Birner, R., Bosch, C., 2019. Soil, Striga, or subsidies? Determinants of maize productivity in northern Ghana. Agricultural Economics 50, 479-494.

Shaxson, F., Barber, R., 2003. Optimizing Soil Moisture for Plant Production, The significance of soil porosity. Food and Agriculture Organization of the United Nations, Rome.

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

Shereen, A., Bernardin, S., Maikel, L.K.S., Haile, A., 2019. Assessing the Employment Effects of Processing Cocoa in Ghana. Int. Labour Off. Strengthen Publ. Ser.

Silva, J.V., Giller, K.E., 2021. Grand challenges for the 21st century: what crop models can and can't (yet) do. Journal of Agricultural Science, 1-12.

Singh, V., 2020. Environmental Plant Physiology: Botanical Strategies for a Climate Smart Planet. CRC Press.

Smaling, E., Stein, A., Sloot, P.H.M., 1991. A statistical analysis of the influence of Striga hermonthica on maize yields in fertilizer trials in Southwestern Kenya. Plant soil 138, 1-8.

Smaling, E.M.A., Janssen, B.H., 1993. Calibration of QUEFTS, a model predicting nutrient uptake and yields from chemical soil fertility indices. Geoderma 59, 21-44.

SRID/MoFA, 2011. Agriculture in Ghana-Facts and figures. Minist. Food Agric. (MoFA)-Statistics, Res. Inf. Dir. (SRID).

SRID/MoFA, 2017. Technical Report. Minist. Food Agric. (MoFA)-Statistics, Res. Inf. Dir. (SRID).

Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop—The FAO crop model to simulate yield response to water: I. Concepts and underlying principles. Agronomy Journal 101, 426-437.

Stockle, C.O., Martin, S.A., Campbell, G.S., 1994. CropSyst, a cropping systems simulation model: water/nitrogen budgets and crop yield. Agricultural systems 46, 335-359.

Stone, J.R., Gilliam, J.W., Cassel, D.K., Daniels, R.B., Nelson, L.A., Kleiss, H.J., 1985. Effect of erosion and landscape position on the productivity of Piedmont soils. Soil Science Society of America Journal 49, 987-991.

Tabi, F.O., Diels, J., Ogunkunle, A.O., Iwuafor, E.N.O., Vanlauwe, B., Sanginga, N., 2007. Potential nutrient supply, nutrient utilization efficiencies, fertilizer recovery rates and maize yield in northern Nigeria. Nutrient Cycling in Agroecosystems 80, 161-172.

Tang, S., Liu, Y., Zheng, N., Li, Y., Ma, Q., Xiao, H., Zhou, X., Xu, X., Jiang, T., He, P., Wu, L., 2020. Temporal variation in nutrient requirements of tea (Camellia sinensis) in China based on QUEFTS analysis. Sci Rep 10, 1745.

Tetteh, E.M., Opareh, N.O., Ampadu, R., Antwid, K.B., 2014. Impact of Climate Change: Views and Perceptions of Policy Makers on Smallholder Agriculture in Ghana. International Journal of Sciences: Basic and Applied Research (IJSBAR) 1, 79-89.

Tetteh, F., Larbi, A., Nketia, K.A., Senayah, J.K., Hoeschle-Zeledon, I., Abdulrahman, N., 2016. Suitability of soils for cereal cropping in Northern Ghana. Council for Scientific and Industrial Research - Soil Research Institute and International Institute of Tropical Agriculture, pp. 1-18.

Tetteh, F.M., Issaka, R.N., Ennin, S., Buri, M.M., 2008. Soil fertility initiative, Fertilizer update and recommendation trials. Soil Research Institute of Ghana, 33.

Tetteh, F.M., Quansah, G.W., Frempong, S.O., Nurudeen, A.R., Atakora, W.K., Opoku, G., 2017. Optimizing fertilizer use within the context of integrated soil fertility management in Ghana. In: Wortmann, C.S., Sones, K. (Eds.), Fertilizer Use Optimization in Sub-Saharan Africa. CAB International, Nairobi, Kenya, pp. 67-81.

Thornton, P.K., Jones, P.G., Alagarswamy, G., Andresen, J., 2009. Spatial variation of crop yield response to climate change in East Africa. Global Environmental Change 19, 54-65.

Timsina, J., Dutta, S., Devkota, K.P., Chakraborty, S., Neupane, R.K., Bishta, S., Amgain, L.P., Singh, V.K., Islam, S., Majumdar, K., 2021. Improved nutrient management in cereals using Nutrient Expert and machine learning tools: Productivity, profitability and nutrient use efficiency. Agricultural Systems 192.

Tittonell, P., Corbeels, M., van Wijk, M.T., Vanlauwe, B., Giller, K.E., 2008a. Combining Organic and Mineral Fertilizers for Integrated Soil Fertility Management in Smallholder Farming Systems of Kenya: Explorations Using the Crop-Soil Model Field. Agronomy Journal 100, 1511-1526.

Tittonell, P., Vanlauwe, B., Corbeels, M., Giller, K.E., 2008b. Yield gaps, nutrient use efficiencies and response to fertilisers by maize across heterogeneous smallholder farms of western Kenya. Plant and Soil 313, 19-37.

Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography 46, 234-240.

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

Van Diepen, C.A., Wolf, J., Van Keulen, H., Rappoldt, C., 1989. WOFOST: A simulation model of crop production. Soil Use and Management 5, 16-24.

van Loon, M.P., Adjei-Nsiah, S., Descheemaeker, K., Akotsen-Mensah, C., van Dijk, M., Morley, T., van Ittersum, M.K., Reidsma, P., 2019. Can yield variability be explained? Integrated assessment of maize yield gaps across smallholders in Ghana. Field Crops Research 236, 132-144.

Vanlauwe, B., Descheemaeker, K., Giller, K.E., Huising, J., Merckx, R., Nziguheba, G., Wendt, J., Zingore, S., 2015. Integrated soil fertility management in sub-Saharan Africa: Unravelling local adaptation. Soil 1, 491-508.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S. Springer.

Vieira, S.R., Gonzalez, A.P., 2003. Analysis of the spatial variability of crop yield and soil properties in small agricultural plots. SciELO Brasil, Bragantia 62, 127-138.

Wijayanto, Y., Prastyanto, E., 2012. A study of using QUEFTS model for establishing site specific fertilizer recommendation in maize on the basis of farmer fields. AGRIVITA, Journal of Agricultural Science 33, 273-278.

World Bank, 2017. Ghana Agriculture Sector Policy Note : Transforming Agriculture for Economic Growth, Job Creation, and Food Security. World Bank Washington DC. World Bank. https://openknowledge.worldbank.org/handle/10986/28394 License: CC BY 3.0 IGO.

World Bank, 2018. Third Ghana Economic Update: Agriculture as an Engine of Growth and Jobs Creation. World Bank, Washington DC.

Wright, M.N., Ziegler, A., 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software 77, 1-17.

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

Zheng, Z., Parent, L.E., MacLeod, J.A., 2003. Influence of soil texture on fertilizer and soil phosphorus transformations in Gleysolic soils. Canadian Journal of Soil Science 83, 395-403.

Zhou, W., Han, G., Liu, M., Li, X., 2019. Effects of soil pH and texture on soil carbon and nitrogen in soil profiles under different land uses in Mun River Basin, Northeast Thailand. PeerJ 7, e7880.

Zingore, S., Mutegi, J., Agesa, B., Tamene, L., Kihara, J., 2015. Soil degradation in sub-Saharan Africa and crop production options for soil rehabilitation. Better Crop. 99, 24-26.

APPENDIX

Appendix A. Chemical composition of the soil organic amendments used in the experiments

Organic Amendment	pH	% Organic Carbon	% N	P (mg/kg)	K (mg/kg)	Ca (mg/kg)	Mg (mg/kg)
Cow dung	9.25	37.83	1.92	3,610	18,750	17,085	24,502
Goat dropping	9.13	21.06	1.74	2,790	18,750	21,292	24,502
Sheep dropping	9.8	22.23	2.03	4,043	31,750	16,342	38,100
Poultry dropping	7.17	24.18	2.86	13,630	16,250	141,850	38,625
Compost	7.22	17.94	1.26	2,977	7,000	4,605	7,822
Town waste	9.82	2.34	0.32	1,552	9,250	42,650	21,035
Fertisol	8	14.04	1.97	12,795	18,000	83,475	35,575

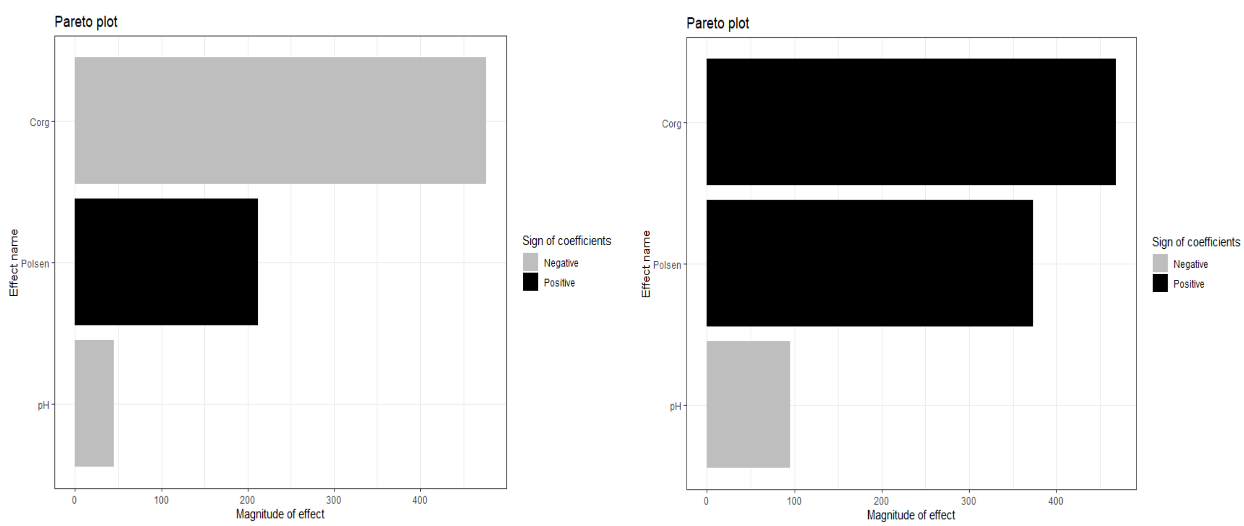
Appendix B. Nutrient contents of the plant residues used as green manure

Plant Material	N	P	K	Ca	Mg	C	C/N	Lignin	Polyphenol
	mg/g							%	
C. odorata	24.6	4.2	25.6	32.1	23.3	336.4	14.89	10.78	1.62
C. juncea	10.7	3.8	13.8	27.4	24.2	434.7	40.8	12.44	0.73
P. maximum	24.9	3.3	26.1	26	20.2	452.5	18.17	13.24	1.48

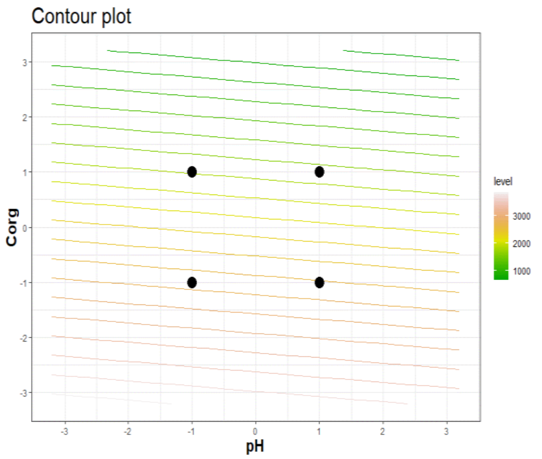
Appendix C. Chemical composition of biochar used in the experiment

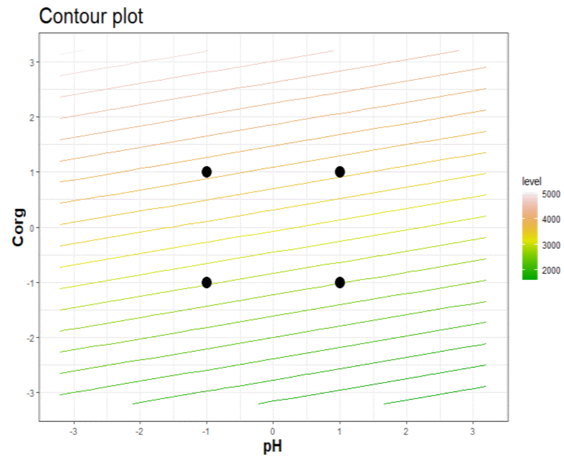
Nutrient	N	P	K	Ca	Mg	C	C/N	CEC
	g/kg							cmol/kg
Value	7.3 (0.1)	0.05 (0.0)	3.6 (0.2)	4.6 (0.1)	0.4 (0.1)	890.7 (3.2)	122 (0.3)	10.9 (0.2)

Appendix D. QUEFTS sensitivity analysis DOI Pareto plot



Appendix E. QUEFTS sensitivity analysis DOI contour plot





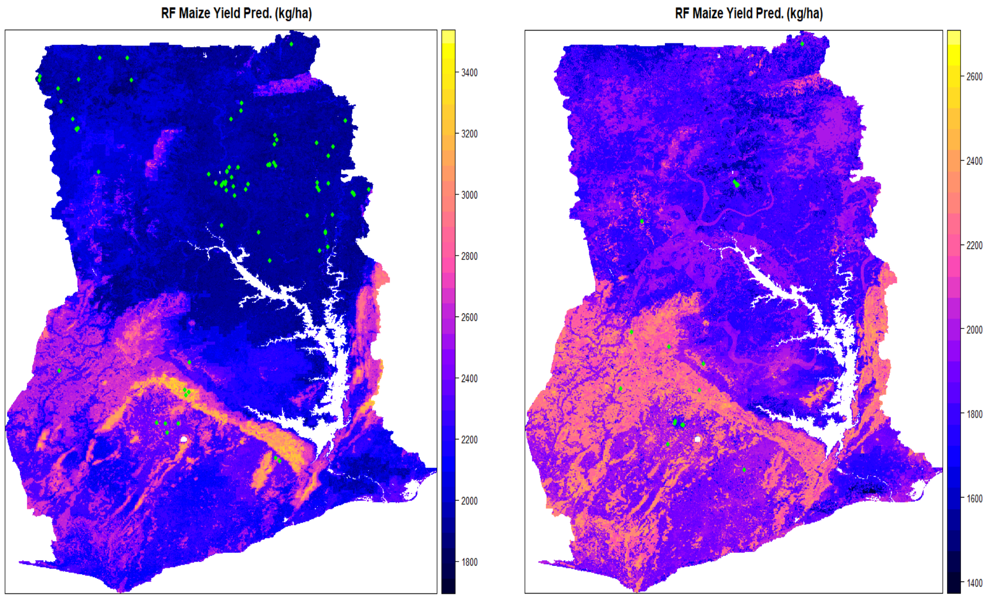
Appendix F. Full matrix of correlation for on-farm trial

	Elev	Rain	Tmin	Tmax	Corg	Norg	Pt	POlsen	Ex.K	pH	CEC	Rzd	Rwhc	Sand	Clay	Silt	YO
Elev	-	-0.16*	-0.68*	-0.54*	0.5*	0.53*	-0.03	0.23*	-0.12*	-0.19*	0	0.51*	-0.16*	0.19*	0.37*	-0.38*	0.33*
Rain	-0.16*	-	0.25*	0.25*	-0.11*	-0.1*	0.04	-0.21*	-0.11*	0.12*	-0.12*	-0.14*	0.35*	-0.1*	0.09*	0.08*	-0.22*
Tmin	-0.68*	0.25*	-	0.83*	-0.29*	-0.48*	-0.08*	-0.08*	-0.24*	0.33*	-0.25*	-0.56*	0.09*	-0.34*	-0.14*	0.46*	-0.45*
Tmax	-0.54*	0.25*	0.83*	-	-0.49*	-0.66*	-0.11*	-0.31*	-0.3*	0.52*	-0.28*	-0.46*	0.13*	-0.24*	-0.21*	0.46*	-0.49*
Corg	0.5*	-0.11*	-0.29*	-0.49*	-	0.73*	0.33*	0.72*	-0.09*	-0.3*	-0.02	0.04	-0.26*	-0.11*	0.52*	-0.28*	0.32*
Norg	0.53*	-0.1*	-0.48*	-0.66*	0.73*	-	0.32*	0.43*	0.14*	-0.61*	0.09*	0.27*	-0.17*	0.07*	0.42*	-0.47*	0.35*
Pt	-0.03	0.04	-0.08*	-0.11*	0.33*	0.32*	-	0.15*	0.29*	0.1*	0.05	-0.18*	0.21*	-0.29*	0.27*	0.01	0.01
POlsen	0.23*	-0.21*	-0.08*	-0.31*	0.72*	0.43*	0.15*	-	0.01	-0.18*	-0.03	-0.09*	-0.25*	0.02	0.21*	-0.24*	0.23*
Ex.K	-0.12*	-0.11*	-0.24*	-0.3*	-0.09*	0.14*	0.29*	0.01	-	-0.16*	0.2*	0.11*	0.35*	0.06	-0.08*	-0.1*	0.09*
pH	-0.19*	0.12*	0.33*	0.52*	-0.3*	-0.61*	0.1*	-0.18*	-0.16*	-	-0.15*	-0.34*	0.2*	-0.08*	-0.22*	0.21*	-0.24*
CEC	0	-0.12*	-0.25*	-0.28*	-0.02	0.09*	0.05	-0.03	0.2*	-0.15*	-	0.23*	0.14*	0.24*	-0.05	-0.15*	0.05
Rzd	0.51*	-0.14*	-0.56*	-0.46*	0.04	0.27*	-0.18*	-0.09*	0.11*	-0.34*	0.23*	-	-0.18*	0.33*	0.12*	-0.35*	0.27*
Rwhc	-0.16*	0.35*	0.09*	0.13*	-0.26*	-0.17*	0.21*	-0.25*	0.35*	0.2*	0.14*	-0.18*	-	-0.22*	0.03	0.23*	-0.25*
Sand	0.19*	-0.1*	-0.34*	-0.24*	-0.11*	0.07*	-0.29*	0.02	0.06	-0.08*	0.24*	0.33*	-0.22*	-	-0.34*	-0.81*	0.19*
Clay	0.37*	0.09*	-0.14*	-0.21*	0.52*	0.42*	0.27*	0.21*	-0.08*	-0.22*	-0.05	0.12*	0.03	-0.34*	-	0	0.08*
Silt	-0.38*	0.08*	0.46*	0.46*	-0.28*	-0.47*	0.01	-0.24*	-0.1*	0.21*	-0.15*	-0.35*	0.23*	-0.81*	0	-	-0.35*
YO	0.33*	-0.22*	-0.45*	-0.49*	0.32*	0.35*	0.01	0.23*	0.09*	-0.24*	0.05	0.27*	-0.25*	0.19*	0.08*	-0.35*	-

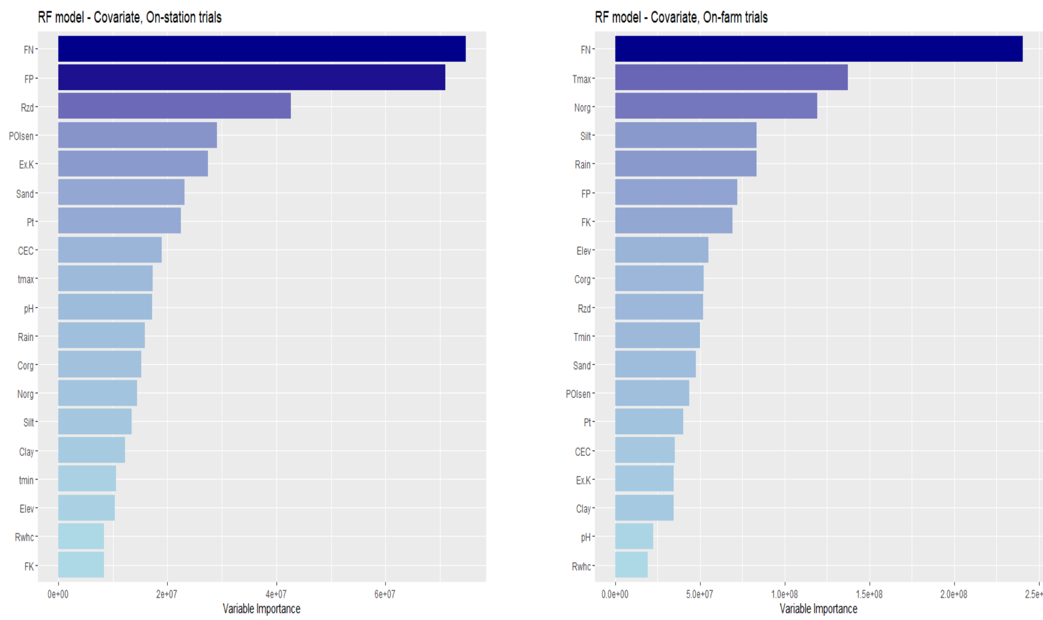
Appendix G. Full matrix of correlation for on-station trial

	Elev	Rain	tmin	tmax	Corg	Norg	Pt	POlsen	Ex.K	pH	CEC	Rzd	Rwhc	Sand	Clay	Silt	YO
Elev	-	-0.04	-0.57*	-0.48*	0.26*	0.65*	0.25*	0.46*	0.4*	-0.11*	0.2*	0.6*	0.11	-0.08	0.59*	-0.37*	0.12*
Rain	-0.04	-	-0.72*	-0.68*	-0.17*	-0.06	-0.18*	-0.08	-0.02	0	-0.03	0.38*	0.39*	-0.22*	-0.13*	0.34*	0.21*
tmin	-0.57*	-0.72*	-	0.9*	0.01	-0.3*	-0.06	-0.29*	-0.2*	0.13*	-0.07	-0.71*	-0.29*	0.26*	-0.22*	-0.11*	-0.24*
tmax	-0.48*	-0.68*	0.9*	-	-0.2*	-0.52*	-0.23*	-0.44*	-0.23*	0.41*	-0.12*	-0.61*	-0.35*	0.49*	-0.3*	-0.31*	-0.28*
Corg	0.26*	-0.17*	0.01	-0.2*	-	0.69*	0.09	0.52*	-0.08	-0.52*	0.67*	-0.24*	-0.17*	-0.46*	0.31*	0.31*	-0.29*
Norg	0.65*	-0.06	-0.3*	-0.52*	0.69*	-	0.4*	0.59*	0.28*	-0.71*	0.44*	0.29*	0.09	-0.56*	0.64*	0.16*	0.15*
Pt	0.25*	-0.18*	-0.06	-0.23*	0.09	0.4*	-	0.62*	0.54*	-0.36*	0.05	0.47*	0.38*	0.06	0.13*	-0.24*	0.29*
POlsen	0.46*	-0.08	-0.29*	-0.44*	0.52*	0.59*	0.62*	-	0.29*	-0.35*	0.22*	0.36*	0.33*	-0.33*	0.34*	0.05	0.09
Ex.K	0.4*	-0.02	-0.2*	-0.23*	-0.08	0.28*	0.54*	0.29*	-	-0.12*	-0.03	0.48*	0.41*	0.14*	-0.14*	-0.23*	0.16*
pH	-0.11*	0	0.13*	0.41*	-0.52*	-0.71*	-0.36*	-0.35*	-0.12*	-	-0.5*	-0.12*	0.04	0.66*	-0.28*	-0.55*	-0.25*
CEC	0.2*	-0.03	-0.07	-0.12*	0.67*	0.44*	0.05	0.22*	-0.03	-0.5*	-	-0.01	-0.38*	-0.29*	0.13*	0.25*	-0.18*
Rzd	0.6*	0.38*	-0.71*	-0.61*	-0.24*	0.29*	0.47*	0.36*	0.48*	-0.12*	-0.01	-	0.43*	0.08	0.29*	-0.32*	0.44*
Rwhc	0.11	0.39*	-0.29*	-0.35*	-0.17*	0.09	0.38*	0.33*	0.41*	0.04	-0.38*	0.43*	-	0.02	0.06	-0.08	0.34*
Sand	-0.08	-0.22*	0.26*	0.49*	-0.46*	-0.56*	0.06	-0.33*	0.14*	0.66*	-0.29*	0.08	0.02	-	-0.39*	-0.83*	-0.1
Clay	0.59*	-0.13*	-0.22*	-0.3*	0.31*	0.64*	0.13*	0.34*	-0.14*	-0.28*	0.13*	0.29*	0.06	-0.39*	-	-0.14*	0.2*
Silt	-0.37*	0.34*	-0.11*	-0.31*	0.31*	0.16*	-0.24*	0.05	-0.23*	-0.55*	0.25*	-0.32*	-0.08	-0.83*	-0.14*	-	-0.02
YO	0.12*	0.21*	-0.24*	-0.28*	-0.29*	0.15*	0.29*	0.09	0.16*	-0.25*	-0.18*	0.44*	0.34*	-0.1	0.2*	-0.02	-

Appendix H. Random forest map



Appendix I. Variable importance effect on maize yield observed (Left: A = on-station; Right = on-farm)



FERARI

FERTILIZER RESEARCH & RESPONSIBLE IMPLEMENTATION

FERARI is an international public-private partnership that builds science-based approaches to site-specific fertilization for widespread adoption by farmers in Ghana for improved food and nutrition security. This calls for a transformation of the fertilizer and food systems that must be driven by evidence-based agro-technical perspectives embedded in multi-stakeholder processes.

To support this transformation, the following institutions have partnered to implement the Fertilizer Research and Responsible Implementation (FERARI) program:

- International Fertilizer Development Centre (IFDC)
- Mohammed VI Polytechnic University (UM6P)
- OCP Group
- Wageningen University and Research (WUR)
- University of Liège (ULiège)
- University of Ghana (UG)
- University for Development Studies (UDS)
- Kwame Nkrumah University of Science and Technology in Kumasi (KNUST)
- University of Cape Coast (UCC)
- University of Energy and Natural Resources (UENR)
- Akenten Appiah-Menka University of Skills Training and Entrepreneurial Development (AAMUSTED) College of Agriculture Education
- Council for Scientific and Industrial Research in Kumasi (CSIR-SRI) and in Tamale (CSIR-SARI) and its subsidiary (CSIR-SARI-Wa)

FERARI operates in conjunction with the Planting for Food and Jobs program of the Government of Ghana (GoG) to embed development efforts into national policy priorities to reach impact at scale. It trains five Ph.D. and two post-doctoral candidates and dozens of master's-level students in building the evidence base for its interventions.

FERARI conducts hundreds of fertilizer response trials on maize, rice, and soybean, on-station and with farmers, and demonstrates them to farmer groups in the northern and middle belt of Ghana. It conducts surveys among farmers and actors in the value chain to understand the drivers for use of fertilizers and other inputs and the marketing of the produce to enhance farm productivity and income. It helps the GoG to establish a Fertilizer Platform Ghana, and is developing its soil mapping expertise toward an information platform.

The content of this report is the sole responsibility of the authors of the involved institutions portrayed on the front page.

Developing Agriculture from the Ground Up