Characterizing the variability of maize yield response to mineral fertilizer in Tanzanian smallholder systems

MSc Thesis Plant Production Systems

Thomas Delaune

September 2018

WAGENINGEN UNIVERSITY & RESEARCH

Characterizing the variability of maize yield response to mineral fertilizer in Tanzanian smallholder systems

---

MSc Thesis Plant Production Systems

Name:	Thomas Delaune
Student ID:	940306177070
Study:	MSc Plant Sciences, Specialization Natural Resource Management
Supervisors:	Dr. AGT (Tom) Schut & Dr. J (Joost) van Heerwaarden
Chair group:	Plant Production Systems
Course code:	PPS80436
Date:	September 2018
Examiners:	Dr. GWJ (Gerrie) van de Ven

Correct citation: Delaune, T. 2018, Characterizing the variability of maize yield response in Tanzanian smallholder systems, MSc Thesis Wageningen University, 51 p.

Contact office.pp@wur.nl for access to data, models and scripts used for the analysis.

WAGENINGEN
UNIVERSITY & RESEARCH

Table of content

Acknowledgement

First, I would like to acknowledge the work of the TAMASA project team, led by Peter Craufurd, that made this study possible. The TAMASA team implemented the experimental design and collected the data in Tanzania. I would also like to thank Jairos Rurunda, that shared with me the datasets and insights about the experimental design.

Moreover, I would like to express my gratitude to my supervisors, Tom Schut and Joost van Heerwaarden, for their advices and support throughout the course of this Thesis. I would like to thank them for their trust and the freedom they granted me to achieve this work.

Abstract

Soil fertility is a major constraint to agricultural development in sub-Saharan Africa. While intensification of crop productivity through mineral fertilizers is recognized as an option, large variability in yields has been observed on smallholders' fields. The formulation of fertilizer recommendations requires an understanding of the sources and scales of fertilizer response variability. During the 2015-2016 growing season, nutrient omission trials were performed on 296 farms to understand rainfed maize yield responses to nitrogen (N), phosphorus (P), potassium (K) and secondary macro- and micronutrients (M). Linear mixed effect and non-linear random forest models were used to analyze how yield and yield responses were related to a range of available geographical data with edaphic and climatic co-variates. The models were calibrated and tested using a holdout approach, at the farm and district level. Drought, temperature, pH, cation exchange capacity, soil organic carbon, P and K explained for at most 16% of the yield nutrient responses variation, after cross validation between farms. The variability of yield macronutrient responses was dominated by small-scale variation. To a much lower extent, also large-scale variation of yield responses was present for nitrogen and phosphorus. These findings highlighted that yield responses to macronutrient supply varied in response to factors that strongly differ between fields, even at small distances. However, yield variation in control plots were explained for 46% by the same covariates, highlighting the effect of large-scale climatic and edaphic factors. None of the tested statistical models had predictive power when testing between districts, highlighting lack of causality between covariates and yield response. Formulating refined fertilizer recommendations is not possible at the field scale for this dataset. The response variation, occurring at small-scale, requires further understanding and inclusion of local information, e.g. related to current and historical field management.

1. Introduction

1.1. Maize cultivation in sub-Saharan Africa

Low productivity in sub-Saharan Africa (SSA) has been attributed to the depletion of soil fertility (Sanchez, 2002). Severe nutrient depletion and declining soil fertility are widespread among African soils (Chianu et al., 2012; Haileslassie et al., 2005; Smaling et al., 1997). Continuous cropping without nutrient replenishment via crop residues, manure, mineral fertilizer or regeneration during fallow periods has led to nutrient mining and degradation of soil resources (Tittonell & Giller, 2013; Vanlauwe et al., 2014). As a result, many regions of SSA are affected by negative nutrient balances (Alley & Vanlauwe, 2009; Smaling et al., 1993). High deficiencies in macronutrients, particularly nitrogen (N) and phosphorus (P) (Chianu et al., 2012; Smaling et al., 1997) but also in micronutrients (Waddington et al., 1998) were observed in many cropping systems and limit crop productivity.

Rainfed maize is the most prevalent staple crop grown in SSA (van Ittersum et al., 2016), with half of the countries allocating more than 50 % of the total cultivated area to this crop (Tesfaye et al., 2015). Averages of actual dry maize grain yields range from 1.2 in Tanzania to 2.2 t ha^{-1} in Ethiopia for the 2003-2012 period (van Ittersum et al., 2016). Maize has one of the largest yield gaps among crops grown in SSA (Global Yield Gap Atlas, www.yield.gap.org) with actual yields representing only 15 to 27% of the water limited potential (van Ittersum et al., 2016). Increase in crop production has been achieved mainly through expansion of the cultivated area (Cassman et al., 2003) but little was attributed to an increase in crop yield, stagnating since the late 90' (Ray et al., 2012). In Tanzania, while the area under rainfed maize cultivation has been multiplied by four during the last decade, average annual yields have been stagnating between 1 and 1.5 t ha^{-1} (FAOSTAT, 2017).

Increasing production in Africa though area expansion has a limited scope in densely populated areas. Moreover, under the current agricultural practices, it is expected to occur at the expense of natural areas and hence associated with land resource degradation (Brink & Eva, 2009). Intensification of crop productivity to meet the increasing food demand (van Ittersum et al., 2016) is therefore desirable in SSA (Vanlauwe et al., 2014). Smallholder farmers, cultivating usually less than two hectares of land, are the first providers of staple crop in most African countries (AGRA, 2014). However, they generally have limited opportunities for intensification. In addition to poor soil fertility, they face socio-economic challenges including limited access to input, credit markets (Xu et al., 2009), and supporting services (Edmonds et al., 2009). These factors are recognized as important constraints that explain the low level of technologies and agrochemical inputs (Tittonell & Giller, 2013; Vanlauwe et al., 2014).

1.2. Crop intensification through mineral fertilization

Since the Africa Fertilizer Summit of 2006, increasing access to fertilizer by targeted subsidies has been a widely accepted intensification option for crop productivity and for addressing the challenge of declining soil fertility (IFDC, 2006). In other words, it aimed to increase fertilization rates in SSA from an average of 13 kg ha^{-1} (Minot & Benson, 2009) to 50 kg ha^{-1} (IFDC, 2006). Recommendations on nutrient management, through mineral fertilizers application, were promoted on uniform or "blanket" basis in relation to crop potential demand (Giller et al., 2011). They were formulated according to the potential yield within an agroecological zones (AEZ), for a given area, crop and dominant soil type (FURP, 1994) assuming homogeneity of production factors at landscape and farm level (Vanlauwe et al., 2015). Plus, these recommendations were derived from fertilizer response trials performed on research stations, often located on productive sites under researcher "best" practices (Vanlauwe et al., 2016) and were covering large areas, even entire countries (Vanlauwe & Giller, 2006). The drawbacks resulting from these generic recommendations were twofold. First, responses to fertilizers were often lower, and more variable on farmer's fields compared to the controlled conditions on research stations (Liverpool-Tasie et al., 2017; Tittonell et al., 2008). Second, these recommendations underestimated the large spatiotemporal heterogeneity of biophysical conditions that can substantially vary within a recommendation domain (Vanlauwe et al., 2015).

Despite a considerable increase of fertilizer use in many SSA countries as a result of fertilizer subsidy programmes (Sheahan & Barrett, 2017), the profitability of this option needs to be discussed (Jayne & Rashid, 2013). Adoption of such intensification option implies a risk as it represents a considerable investment for smallholder farmers (Bumb et al., 2011; Edmonds et al., 2009). In Tanzania, fertilizer application increased from 9 to 17 kg ha^{-1} per year since 2008 after the introduction of a fertilizer subsidy programme under the Agricultural Input Voucher Scheme (Senkoro et al., 2017; Mowo et al., 1993). These quantities remain far below the fertilizer recommendation based on the AEZ scheme proposed by De Pauw (1984). Moreover, the relatively low return on fertilizer of 7 kg of grain per kg on nitrogen applied (Mather et al., 2016) questions the efficiency of the blanket recommendations. Before their formulation, there is thus a need to understand the variability of the responses and its magnitude (Vanlauwe et al., 2016) as it translates directly the probability of success and determines farmer's willingness to adopt (Bielders & Gérard, 2015).

1.3. Yield response variability: magnitude and scale

1.3.1. Yield responses to fertilizer in SSA

Positive crop responses have been observed in farmers' fields with fertilizer application for several crops (Edmonds et al., 2009) such as maize in SSA (Kihara et al., 2016; Vanlauwe et al., 2011), and millet and soybean in West Africa (Buerkert et al., 2001). The response distribution and magnitude of variation is however more informative than the average response (Vanlauwe et al., 2016), but has only received interest recently. For instance, in a study by Buerkert et al., (2001) microdose fertilization of millet increased yield on average by 120% with a yield increase ranging from 0 to 2000 kg ha^{-1}. For NPKS (respectively 80-60-60-24 kg ha^{-1}) fertilization of maize plot trials in Kenya (Vanlauwe et al., 2016), average increase was 180% with a yield response ranging from the negative with – 2000 to 5500 kg ha^{-1}.

The wide range of responses can be illustrated by cumulative frequency curves that show the frequency of the yield increment with fertilizer. Vanlauwe et al., (2016) presented it for maize and bean (Fig. 1a,b). Moreover, it shows that for the same treatment, the potential gain from fertilizer application varies greatly across farms within similar AEZs. High variability in treatment response has been observed for several crops e.g. millet (Bielders & Gérard, 2015), soybean (Ronner et al., 2016), and maize (Kihara et al., 2016), with a wide range of responses at every level of control yield.

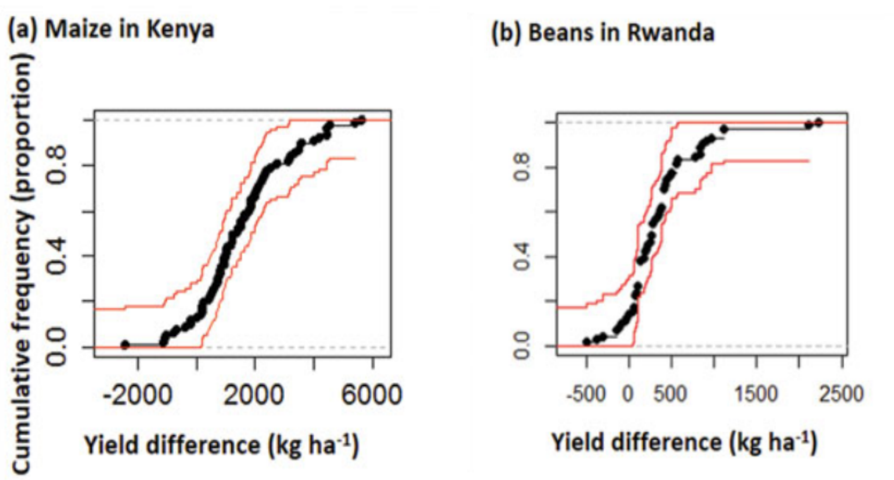

Figure 1. Cumulative frequency curves of maize response to NPKS fertilizer in Vihiga and Siaya district of Kenya (a) and bean response to NP fertilizer in Rwanda, from Vanlauwe et al. (2016)

1.3.2. Potential sources of variation at different scales

Under nutrient limiting conditions yield variability is related to the level of soil fertility and the farm management history (Zingore et al., 2007). When fertilizer is applied and nutrients are not limiting, the yield response variation is caused by factors that determine the components of nutrient use efficiency, i.e. capture and conversion efficiency (Tittonell et al., 2008). Moreover, the yield obtained with nutrient application varies according to geologic, pedologic (slope, soil type), and climatic (rainfall amount and distribution) factors that set permanent limitations and determine the attainable yield (Tittonell & Giller, 2013). Other processes, both biotic (pest and diseases) and abiotic stress (drought), can also occur during the growing period, reducing yield and nutrient uptake (Vanlauwe et al., 2015). The combination of these factors determines the crop growing conditions. They vary greatly in space and time, and contribute to yield variability at different scales.

At national and regional scale, yield variability may be attributed to different soil types, landforms, and climates. At that scale, a wide range of water limited yields for rainfed maize can already be observed. For Tanzania, they vary from 2400 to 8000 kg ha^{-1} (Global Yield Gap Atlas, www.yield.gap.org). Various agroclimatic zones co-occur in SSA (Voortman et al., 2003). They are characterized by temperature (min and max), rainfall (amount and distribution), solar radiation and the length of the growing season. It is also important to account for temporal variability of weather patterns that impacts the length of the growing season and rainfall, that substantially vary from year to year (Rowhani et al., 2011). In addition, different soil types with very heterogeneous distribution patterns can be observed at the regional level (ISRIC, www.soilgrids.org). Soil types are characterized by different inherent nutrient retentions and water holding capacities (Bationo et al., 2012) ; chemical properties ; and resilience to land degradation by erosion (Stocking, 2003). These properties, resulting from long-term processes by soil forming factors are correlated with plant growth and nutrient uptake constraints (Baligar et al., 2001).

At landscape and community/village scale, within similar agroclimatic conditions, yield response to fertilizer is influenced by factors varying on small distances. Within a Kenyan landscape, Njoroge et al. (2017) observed large spatial variability of maize yield responses to NPK fertilization. For similar fertilization treatments, Kihara et al. (2016) also reported large magnitudes of yield response to NPK (100-30-60 kg ha^{-1}) within several African landscapes. Detailed landforms can be recognized at that scale with different pedologic conditions varying along the toposequence. Field position in the landscape and topsoil texture are then important criteria to classify soil conditions (Deckers, 2002). Analyzing crop NDVI response to fertilization in a landscape of Mali showed that landscape position contributed significantly to the yield variation (Blaes et al., 2016). Differences between catena position can be characterized by distinct slopes, rootable depths, water holding capacities, drainage characteristics (Blaes et al., 2016; Deckers, 2002) and influence crop response to weather extremes (Bationo et al., 2012)

At farm scale, yield variability can be attributed to local conditions that are characterized by short range variation of soil properties in combination with crop and soil management (Tittonell et al., 2007, 2007; Zingore et al., 2007). For instance, Yemefack et al., (2005) showed that pH, clay content and available P were varying strongly at local level and illustrates the influence of land use on topsoil variability. In addition, historical management results in great heterogeneity between and within farms, evidence by the presence of strong fertility gradients (Tittonell et al., 2005). To a certain extent, practices leading to severe soil nutrient depletion lead to the occurrence of “non-responsive” fields, where commercially available fertilizer application is not followed by an increase in crop productivity (Zingore et al., 2007). They are characterized by complex nutrient imbalance, including macro- or micronutrient deficiency, often poor physical structure, and occasionally, nutrient toxicity (Kihara et al., 2016; Kurwakumire et al., 2014).

1.4. Targeting fertilizer recommendations: key challenges

There is in general agreement among researchers that crop intensification, in such heterogeneous context, cannot be achieved without locally adapted options, tailored to the productivity constraints of the farming systems (Giller et al., 2011). Spatially adapted fertilizer recommendations, in terms of quantity, blend and form (Muthoni et al., 2017), are expected to improve fertilizer agronomic efficiency (Vanlauwe et al., 2011). However, the challenge stands in the delimitation of recommendation domains, areas where farmers are facing a set of similar productivity conditions and yield responses are expected to be homogenous (Jauregui & Sain, 1992). Researchers have often treated the farm as a relevant level to evaluate performances of intensification options (Vanlauwe et al., 2015). Formulation of farm specific recommendations is however not yet available publicly. Then, this poses questions about the optimal scale for targeting fertilizer recommendations. What is feasible, according to the available data, and what is applicable, according to farmers access and awareness of input market?

A second challenge exists in capturing the heterogeneity of the growth conditions in the area of interest. Site specific studies do not give insight in the patterns that may vary gradually across the landscape. Controlled on-farm experiments, with consistent testing of treatments over a large area, are necessary to understand yield constraints across a population of heterogeneous farms (Lobell et al., 2009; Van Ittersum et al., 2013). An un-biased selection of experimental sites is needed to properly represent the area and farming conditions, and to capture the spatial heterogeneity in the agricultural landscape (Hochman et al., 2013; Vanlauwe et al., 2016).

As discussed by Vanlauwe et al. (2016) the problem of data collection and data quality needs to be addressed, as trials performed over large areas increase the number of operators and the probability of inconsistency in the measurements. The use of biophysical datasets, soil and climate related, can be challenging in such heterogeneous context (Waha et al., 2015). On one hand, measurements of soil properties can be inaccurate or not informative and show weak relationships with yield response (Njoroge et al., 2017). On the other hand, interpolated data predicting soil properties at coarse level may not capture short range soil spatial variability. As a result, the accuracy of the data needs to be taken into account when selecting variables to explain yield response (Grassini et al., 2015).

One of the key requirements for targeted recommendations consists in understanding the impacts of the biophysical environment on the yield response. Only few studies attempted to quantify the contribution of the factors discussed in section 1.3. and their interactions with the fertilizer performances. Ronner et al., (2016) succeeded to explain a substantial part of variability based on local management and environmental variables with a R^2 of 0.61 for soybean response to P fertilizer and rhizobium inoculants. However, the predictive power of their variables was limited for targeting recommendations as it decreased considerably when cross validating their model in new areas and other growing seasons. Bielders & Gérard, (2015) explained 20% of the overall millet yield variability by environmental and management factors and 27% by the effect of fertilization and manure application. In a more detailed selection of variables Burke et al., (2017) highlighted significant interactions between soil conditions and fertilizer response for maize. However, while numerous variables were used to explain variability, only 30% was explained by those covariates. The results of these studies highlight the challenge in the selection of explanatory variables and their accuracy. Most importantly, it highlights the difficulty of drawing reliable conclusions when explaining yield response variability within such heterogeneous contexts. The underlying components of yield response in SSA are currently not understood well enough to address fertilizer options at a scale that ensure farmers profitability.

2. Research objectives

Improving the understanding of the factors governing the yield response to fertilizer and their scale of influence are crucial information for the formulation of spatially targeted fertilizer recommendations. There is an increasing number of available geospatial datasets on climatic and edaphic conditions. However, it is unknown if these data, can improve our understanding of the nature of yield response variability and help to better define recommendation domains.

The aim of this study was to assess the sources of maize yield variability in response to mineral fertilizer application in Tanzania through the analysis of widespread on-farm testing of nitrogen (N), phosphorus (P), potassium (K) and secondary macro- and micronutrients (M). The analysis relies on the following axes of investigation.

The first step was to describe the variability of yield response variation to N, P, K and micronutrients application from the farm, to the regional scale. Second, spatial variability of yield response and presence of response patterns across the study sites were evaluated. Third, using climatic, topographic and edaphic variables derived from geospatial datasets into a statistical model to explore the relationships between biophysical context and yield response. Different modelling approaches were evaluated to predict yield response with the use of linear mixed model, spatial autoregression, geostatistics and machine learning. Fourth, the ability of the selected variables to predict yield response was evaluated by cross-validation and their relevance for improved targeted fertilizer recommendations was discussed.

3. Materials and methods

3.1. Study area

This study uses data from on-farm trials located in the Southern Highlands and Northern Zone of Tanzania (Fig. 2). The country has a complex landscape due to the East African Rift, resulting in topographic and climatic variability ranging from tropical savanna climates near the coast to warm semi-arid climates covering the Southern and Northern highlands (Rowhani et al., 2011).

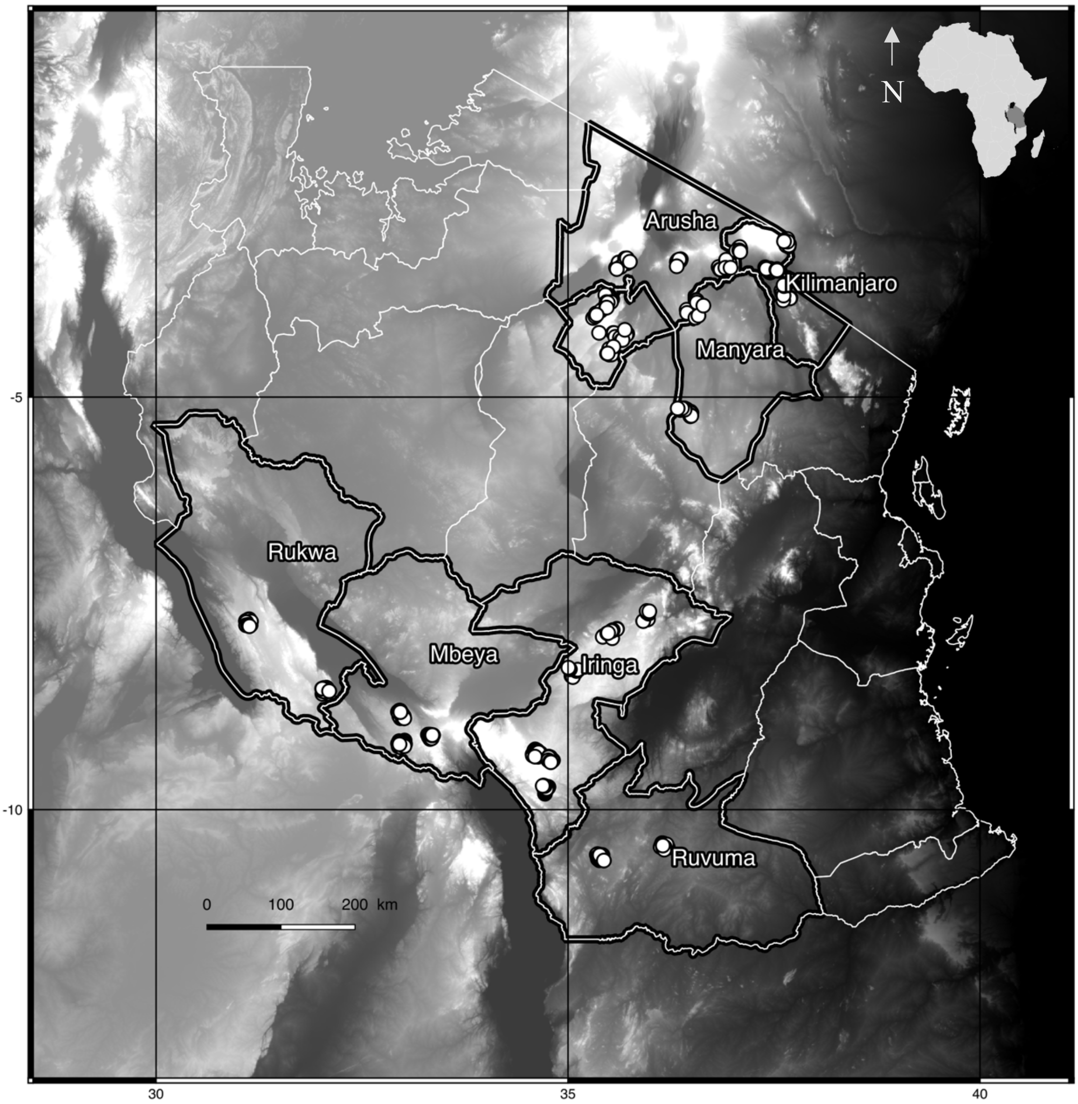

Figure 2. Geographic distribution of on-farm trials across the regions of Tanzania and a digital elevation model (DEM) GEMTED2010. Maize target regions are outlined in black.

The Southern Highlands zone is located between 7.5° to 10.5° S latitude, and 31° 36.5°E longitude, and it covers the regions of Rukwa, Iringa, Ruvuma, Mbeya. The Northern zone is located between 3° 5.5°S latitude, and 35° 38°E longitude, and includes the regions of and Manyara, Arusha and Kilimanjaro (Fig. 2). In terms of AEZ (De Pauw, 1984), the Northern Highlands (1000-2500 meter above sea level) are characterized by volcanic uplands with deep fertile loamy soils in the Kilimanjaro region, with a bimodal, but very variable rainfall (1000-2000 mm per year) pattern. Around the central part of the country, in Manyara and Northern Iringa (1000-1500 masl), well drained but not very fertile soils and saline soils are present with a unimodal, variable rainfall (500-800mm per year) pattern. In the Southern zone, the Highlands (1200-1500 masl) cover the regions of Mbeya and Iringa with moderate fertile clay soil sand low fertility sandy soils in the West of Rukwa. The Plateau region of Eastern Ruvuma is characterized by sandy plains. The Southern regions benefit from a more reliable unimodal rainfall when compared to the Northern Highlands, with a growing season lasting from about November to April (800-1400 mm per year) (Magehema et al., 2014).

Trials were implemented in areas of interest (AOI) defined by the TAMASA (Taking Maize Agronomy to Scale in Africa) project team. AOIs were defined as areas representative of major maize-based systems (50% of the crop in more than half the season is maize), where farmers have intensification as their aim (cf. TAMASA Sampling Strategy and site-selection SOP). A total of 296 on-farm trials ran in randomly selected locations within a selected set of 26 grids of 10^2 km during the maize growing season of 2016 (Fig. 2). Grids were selected to represent all major climate zones. In this study, “plot” referred to the 6 different treatments within a single “farm”. The term “farm” was used to name what is actually the field scale, as there was one field studied per fam. Farms were included in a 10 km^2 grid referred as “district”, each district contains between 6 and 15 farms. District being the closest scale to the grid cell scale, this term is used instead of “grid cell” for this study. The locations were further differentiated according “zones”, the Northern and Southern Highlands zone.

3.2. Experimental design

On-farm trials consisted of a nutrient omission trial (NOT) design with one replication per farm. Fields in the target areas were selected to represent the main types of farmers and soil types in the area. Only fields with uniform soil fertility, managed homogeneously in the past by the farmer were selected in order to avoid as much as possible confounding effect of within-field heterogeneity.

Each trial comprises 6 fertilization treatments and were performed on a single field: a control without fertilization, PK (-N), NK (-P), NP (-K), NPK, NPK with secondary macronutrients (Ca, Mg, S) and micronutrients added (Zn, B). Nutrient application rates varied according to rainfall and maize production potential of the location. Three levels of fertilization for N, P and K were applied with increasing rates, i.e. 100-30-30, 120-40-40 and 140-50-50 kg ha^{-1} (Table 1).

Table 1. Fertilization rate in kg ha^{-1} for N, P and K fertilization in NOT plots according to expected production potential

Production potential	Low			Medium			High		
	N	P	K	N	P	K	N	P	K
Control	0	0	0	0	0	0	0	0	0
PK	0	30	30	0	40	40	0	50	50
NK	100	0	30	120	0	40	140	0	50
NP	100	30	0	120	40	0	140	50	0
NPK	100	30	30	120	40	40	140	50	50

Fertilization rates were applied to achieve expected attainable yields without nutrient limitations. Expected attainable yields were determined based on rainfall (low, moderate and high) and agroecological potential with production potential ranging from 5-6 t ha^{-1}, 7-8 t ha^{-1} to 8-10 t ha^{-1} at a given location (Fig. 3). These production potentials were assigned according to expert knowledge.

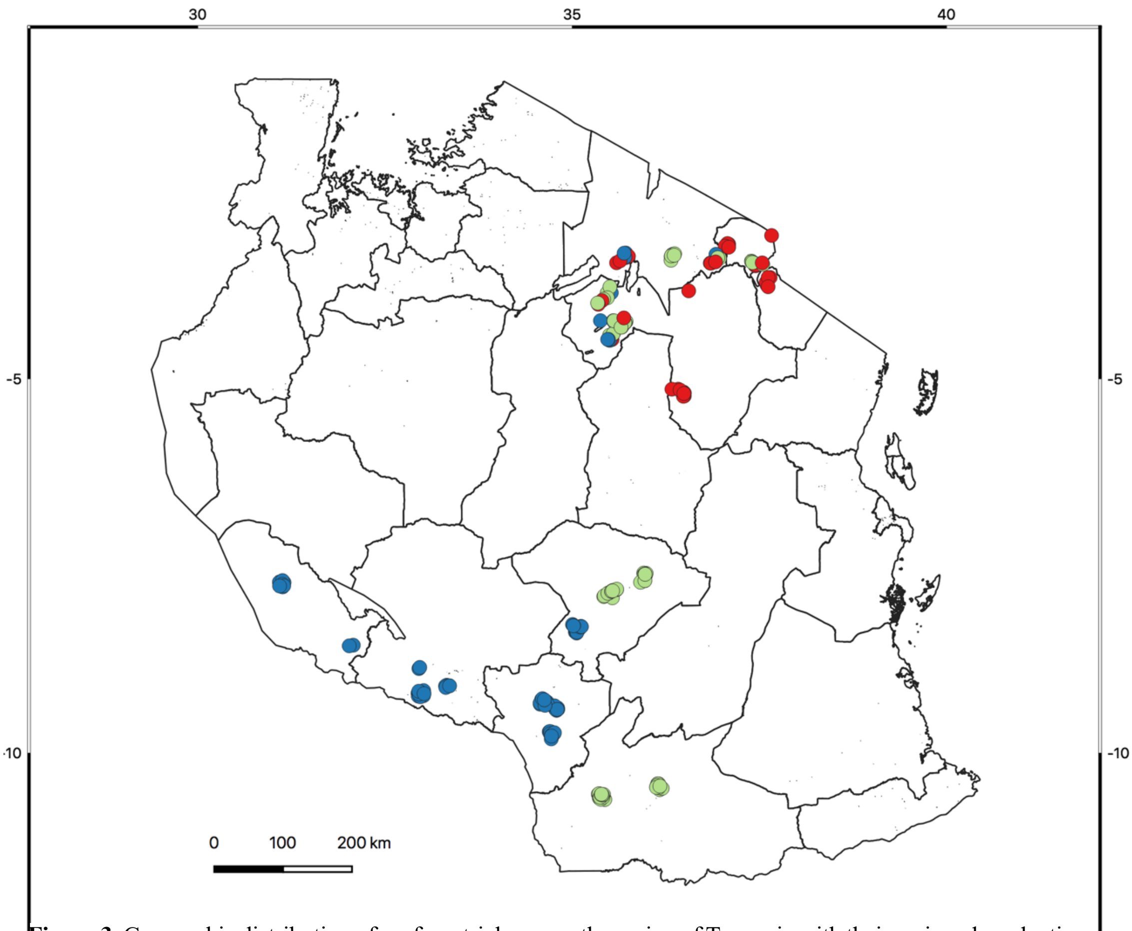

Figure 3. Geographic distribution of on-farm trials across the region of Tanzania with their assigned production potential classes, low (red), medium (green) and high (blue).

Maize (*Zea mays* L.) crops received N application in three equal splits: at planting, 1st and 2nd topdressing, other nutrients were all applied as basal application at planting (cf. Guidelines and protocols for implementation of on-farm nutrient omission trials as a basis for developing district-specific nutrient management practices). The fertilizer type for N, P, K application were urea, triple superphosphate and potassium chloride respectively.

Planting was performed at a spacing of 75 x 25 cm to ensure a density of at least 53 000 plants ha^{-1}. All fields were prepared with conventional tillage and weeded manually at least twice during the cropping season. Gross plot sizes were 25 m^2 for the trials in the Namtumbo district and 64 m^2 for the rest of the trials, corresponding to a net plot size of 16 and 36 m^2 respectively. Crops were grown on the gross plot area and harvested from the net plot. Detailed information about maize varieties used in this study were not available, a high yielding hybrid variety recommended for the specific district growing conditions was used.

3.3. Data collection

3.3.1. On-farm soil and crop data

Four soil samples were taken from each experimental field at 0-20 cm depth before planting for soil analysis. Data from two methods were available, near-mid infrared spectrometry (NIR-MIR) from CYMMIT (International Maize and Wheat Improvement Center) and wet chemistry from IITA (International Institute of Tropical Agriculture). Only the latter was used in this analysis due to the discussed reliability of the former (Njoroge et al., 2017; S1). Measured soil parameters from wet chemistry analysis are presented in Table 2. Each of the on-farm experimental fields, where the measurements were taken, was georeferenced with its latitude, longitude and altitude using a standard GPS device.

Table 2. Measured field soil chemical properties based on Mehlich-3 extraction

Soil nutrient	Abbreviation	Unit
Total Carbon Content	Org. CC	%
Total Nitrogen content	N	%
Phosphorus concentration	Meh. P	ppm
Potassium concentration	K	ppm
Sulphur concentration	S	ppm
Copper concentration	Cu	ppm
Exchangeable calcium concentration	Ca	ppm
Exchangeable manganese concentration	Mn	ppm
Zinc concentration	Zn	ppm
Exchangeable magnesium concentration	Mg	ppm
Exchangeable sodium concentration	Na	ppm
Iron concentration	Fe	ppm
Boron concentration	B	ppm
Soil water pH	Water pH	
Effective cation exchange capacity	ECEC	cmol kg^{-1}

Nutrient extracted by Mehlich 3 extraction (wet chemistry)

Crop data, including planting date, harvest date, fertilizer application and yield were collected by members of TAMASA project team. Six enumerators performed the data collection across Tanzania. Maize was harvested between May and October 2016 at physiological maturity. Yields were measured at harvest from the net plot area (avoiding border effect) as weight of fresh cobs per plot after stover removal. Five cobs were randomly selected in order to calculate the shelling percentage of the cobs. The following coefficient was used to convert cob fresh yield into grain fresh yield: Coef. Shelling = fresh grain weight (kg)/ fresh cobs weight (kg). Grain moisture content was measured in the field at harvest with a grain moisture meter. The following coefficient was used to convert fresh yield into yield adjusted at 12.5% moisture content: Coef. Adj. MC = (100 – Moisture Content)/87.5. Fresh cob weight per plot was then converted into kg ha^{-1} of grain at 12.5% moisture content.

3.3.2. Geospatial data

Several geographic open source datasets were included in this study to represent the climatic, edaphic and topographic conditions under which the trials took place (Table 3). These variables were selected as they are known to have an effect on maize growth and fertilizer use efficiency. Climatic variables were chosen to assess the extent of heat and water stress during the cropping period. Several rainfall characteristics (Beyer et al., 2016) were calculated using daily precipitation data from Climate Hazards Group InfraRed Precipitation (CHIRPS) (Funk et al., 2015). The characteristics included were: total rainfall; number of dry days; length of the longest drought, and precipitation seasonality. Temperature data were retrieved from MODIS Land Surface Temperature (Wan et al, 2015) 8-days aggregated. The third quartile of the temperatures and the temperature seasonality were chosen as variables. Soil physical properties provided by ISRIC (www.soilgrids.org) were used as an estimation of the soil texture parameters of the trial location such as sand content, bulk density and proportion of coarse fragments (Hengl et al., 2015). Root zone depth was also used in this study as an indicator of root zone plant available water holding capacity (Leenaars et al., 2018). In order to bring an estimation of the extent of land degradation and soil constraints, a set of land characteristic variables were added (Vågen et al., 2016, 2013). Furthermore, it can be assumed that vegetation characteristics of the surrounding environment may have impacted the extent of land degradation by erosion (Tully et al., 2015). Soil organic matter peaks after forest clearance and gradually declines during years of cultivation (Tittonell et al., 2007). As a result, a dataset of forest cover (Hansen et al., 2013) has been added, measuring the canopy closure of vegetation taller than five meters in 2000. To bring an indicator of land use, the percentage of cropland cover was retrieved from GlobeLand30 (Chen et al., 2014).

Table 3. Description of the biophysical geographic variables derived from geodata and their resolution

Variable name	Code	Unit	Res.	Reference
Topography				
Altitude	Altitude	m	20 m	GNSS
Slope	SLOPE	°	250 m	AfSIS, 2014
Climate				
CHIRPS Climate Hazard IR Prec.			5 km	Funk et al., 2014
Total Rainfall in GP	TotRainfall	mm		
Dry days (<0.5 mm prec.)	DryDay	day		
Longest Drought period (<0.5 mm prec.)	Maxdrought	day		
Precipitation seasonality	CVRain			
MODIS Land Surface Temperature - 8 days			1 km	Wan et al., 2015
Third quartile of temperature	TQTemp	°C		
Temperature seasonality	CVTemp			
Soil physical properties				
Sand content	SG.SAND.15	%	250 m	Hengl et al., 2015
Bulk density	SG.BULK.15	kg dm^{-3}		Hengl et al., 2015
Coarse Fragment	SG.COAR.15	%		Hengl et al., 2015
Root zone depth	RZD	cm	1km	Leenaars et al., 2015
Land characteristics				
Forest canopy cover in 2000	Tree2000	%	30 m	Hansen et al., 2013
Cropland area	Cropland.cov	%	30 m	Chen et al., 2014

3.4. Data processing

3.4.1. TAMASA datasets

Data processing and statistical analysis were performed in R version 3.4.3. Graphical outputs in R, are obtained using the *ggplot* package. Maps were produced with QGIS 3.02. A first dataset (296 farms) with georeferenced information on plant characteristics and yield was provided and background information of every plot and farm was carefully checked. In order to account for potential measurement errors of shelling and moisture content, cob fresh weight per plot was converted using farm medians of the shelling percentage and the moisture content adjusted at 12.5%. While this method overcame problems of missing data for shelling and moisture content, it was assumed that fertilization treatment did not affect these parameters. Plots with missing yield data (cob fresh weight or plant parameters), missing background information, missing fertilization rate or incomplete design (missing treatment) were discarded resulting in a dataset of 227 farms. A second dataset with analysis of soil properties was merged based on geographic coordinates due to incorrect background information for numerous farms. Background information was checked after merging through similarity between character strings using the Levenshtein distance. The package *stringdist* was used for this purpose. Two farms were removed from the process resulting in 225 farms.

The geographic distribution of data points is theoretically clustered in 10 km^2 grids. Using a hierarchical clustering model with the *hclust* function, new clusters were assigned. The aim was to combine two districts that were very close geographically in order homogenize the "district" scale (Fig. S2). The cutoff distance was chosen as trade-off between the total number of districts and the number of farms per districts. Finally, new districts containing less than three farms were discarded from the analysis resulting in a final dataset of 219 farms distributed in 20 districts (Table 4). A third dataset with planting and harvest date has been associated with the yield dataset, with data available for only 130 farms (Table 4). This subset was only used to predict yield response in function of environmental covariates while the whole dataset was used in the exploratory analysis.

Table 4. Farm distribution per districts for the whole dataset and the subset including planting and harvest date used for yield response predictions

Region	District	n (all)	n (subset)	AE Potential*
Southern Highlands				
Rukwa	Nkasi	12	0	HP
Mbeya	Mbozi	9	9	HP
	Mbeya Rural	10	4	HP
Iringa	Njombe	21	0	HP
	Ludewa	14	0	HP
Ruvuma	Songea Rural	15	0	MP
	Namtumbo	12	0	MP
Iringa	Mufindi	15	10	HP
	Iringa rural	12	12	MP
	Kilolo	10	10	MP
total		130	45	
Northern zone				
Kilimanjaro	Mwanga	6	6	LP
	Moshi Rural	10	10	MP
	Hai	6	6	LP
Arusha	Arumeru	9	9	LP
	Monduli	11	11	MP
	Karatu	8	8	HP/LP
Manyara	Mbulu	9	9	MP
	Babati	10	9	MP
	Hanang	12	9	HP
	Kiteto	8	8	LP
total		89	85	

*For the Southern Highlands, production potentials (low, medium and high) are homogeneous within a district, in the Northern zone, different potentials coexist within one district and only the dominant one is reported in this table.

3.4.2. Geographic open-sources data

MODIS Land Surface Temperature for temperature and CHIRPS for rainfall data were expressed in °C and mm respectively. In order to account for the potential effect of heat and water stress on fertilizer response at different physiological stage, each of the related variables (Table 3) were aggregated in three phases splitting the growing season. An example of this procedure can be found in Landau et al., (2000). Phase I represented the onset of the growing season, starting 10 days before sowing until the 15^{th} day after sowing. Phase II included the vegetative, early reproductive and anthesis growing stages, from the end of phase I and to two thirds of the total growing period. Phase III represented the grain filling and maturation growing stages from end of phase II to harvest. The overall growing period (GP) included all three phases.

3.5. Data analysis

3.5.1. Response variables and covariates assessment

First, a descriptive analysis was first performed to evaluated plant parameters, treatment yield, yield response to individual nutrient and agronomic efficiency using scatterplot and boxplot. Plant parameters included the proportion of harvested plants per plot (%), in comparison to the expected plant density at harvest, the number of cobs per plant, the individual cob weight (kg). Proportion of harvested plants was calculated as the ratio of the number of plants per plot at harvest and the number of plants per plot at planting, 192 plants per plot or 85 for the Natumbo district.

Yield absolute responses (kg ha^{-1}) to nutrient (Nitrogen (N), Phosphorus (P), Potassium (K), Secondary macro- and micronutrients (M)) were estimated using the following linear regression with farm and the interaction between farm and presence (binary variable) of the four nutrients as fixed effect. Interactions between nutrients were ignored in the model formulation as no field-specific interactions were assumed.

Model 0: yield ~ farm + farm:(N presence + P presence + K presence + M presence)

From the above model, the coefficients for farm:N, farm:P, farm:K and farm:M are the best unbiased estimates of the farm specific response to N, P, K and M. These coefficients ("nutrient presence" * "farm") were extracted and use as nutrient response variables. Exploratory analysis and yield response predictions were performed using these estimated nutrient yield responses. This method consisted in estimating farm nutrient responses by "comparing" all treatments with a given nutrient present to the other treatments were this nutrient was absent. It was considered more suitable than the observed yield response as it accounts for within farm variation for a given nutrient. Observed response being usually calculated as the difference between the yield in the NPK plot and the yield in the plot where the nutrient of interest is omitted. Response to NPKM was however calculated as the absolute difference between the control and the full fertilization (NPKM) plot. In addition, the within farm error was estimated by the root mean squared error of a simple linear regression between estimated and observed nutrient responses for each nutrient.

Agronomic efficiency (Vanlauwe et al., 2011) of N, P and K was calculated as the estimated nutrient yield response (e.g. yield response to N in kg ha^{-1}) divided by the quantity of input applied (kg of N applied). Treatment effect on yield was evaluated at the district level using the stability analysis approach from Raun et al., (1993) that consisted in plotting the district median treatment yields over the district median yield (all treatment confounded). The distribution of yield responses to nutrient was assessed graphically using cumulative frequency curves (Vanlauwe et al., 2016).

Before building predictive model assessment of the potential explanatory variables with summary tables was conducted where the district median of each covariate and their respective coefficient of dispersion were reported. Correlations between variables were further assessed with principal component analysis (PCA) with the *dudi.pca* function of the *ade4* package.

3.5.2. Linear mixed models

a. Theoretical justification

In the present case, the maize grain yield represents the response variable (**Y**) where **Y** is a vector of n observation from different locations. Fertilization treatments and/or environmental covariates are the potential explanatory variables (X) where X is a matrix of n observations and v variables. The linear relation (1) can be expressed with the following formula, where β is a coefficient for each X and ε N(0, σ^2) the n residuals of Y not explained by X:

$$Y = X\beta + \varepsilon \quad (1)$$

Due to the structure of the sampling distribution, it is reasonable to assume that observations from the same district, i.e. under similar biophysical conditions, are likely to be more related than observations between different districts. However, the assumption of independence as well as homogeneity may be violated. Mixed-effect modelling is then a better tool suited for clustered data where residuals are assumed independent and that have not constant variances (Zuur et al., 2009). Relation (1) become the following where the random term is $Z_i b_i$, Z being the random factor and b its coefficient, for i level of district.

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \quad (2)$$

The yield can then be modelled as a linear function of the fertilization treatments (fixed effect) where the intercept is allowed to vary per district and per farm within district (random effect). In this study, the purpose of the mixed model is to account and correct for district/farm specific yield differences. It also allows to quantify the variation associated with a particular level of data i.e. evaluate the sources of variability. As experimented by van Heerwaarden et al. (2017) and Tittonell et al. (2013), such multilevel model can be used to quantify variation, here of maize yield response (c.f. 3.5.1), at different spatial scales.

b. Practical use

Model 1: yield ~ treatment*(zone + fertilization rate), random = ~ 1 | district/ farm

Model 1 was used to check significance of the average treatment effect on yields and their interaction with the zone (Northern, Southern Highland) and fertilization regime (Low, Medium and High potential). The model was first used on the whole dataset after processing (c.f. 3.4) to detect main outliers from the treatment effect. From the analysis of the model, farms containing plot-level residuals higher than three times the standard deviation of the pooled residuals was excluded from the dataset. These farms were inspected individually and showed very strong heterogeneity in response to treatments, or a very high yield (12 to 13 t ha^{-1}). Consequently, 9 farms were removed and not included for further analysis. It was assumed that such variation was due to unexpected events and/or strong within farm heterogeneity, affecting plot performances.

Moreover, this model was used to estimate grain yield and yield components according to the fertilization treatment, fertilization level and zone. A separate analysis (model 1b) for the Northern and Southern Highlands was performed. This model had the same structure as model 1 without zone as fixed effect. The scale to which fertilization rate was applied differed between zones. In the Southern Highlands, there was no variation of rate within district giving a per district application, contrary to the Northern zone (Fig.3, Table 4). For this model, significance was computed with F statistics using the *anova* function directly on the model (Pinheiro & Bates, 2000). Fixed effects and their interactions were further analyzed with pairwise comparison using the *predictmeans* package.

Model 2: yield nutrient response ~ random = ~ 1 | zone/ district

Model 2 was used to analyze the contribution of geographic stratification (the scales), by estimating the variance components for the response to N, P, K and M (obtained from model 0, c.f. 3.5.1) with the zone, districts as random effect only. The contribution of a given scale on the total variation can be quantified with the intraclass correlation coefficient. To this end, the ratio between the variance of a given strata over the total variance was computed. For the analysis of the variance component, mixed models were fitted with REML (Fox et al., 2015)

Model 3: yield ~ treatment + environmental covariates, random = ~ 1 | district /farm

For the analysis of this mixed model, significance of fixed effects was assessed via likelihood-ratio test using the χ^2 distribution by comparing model models with and without a given fixed effect. P-value < 0.05 indicated significance differences between models and significance of the fixed effect.

Model 4

control or yield nutrient response ~ environmental covariates, random = ~ 1 | district

A final mixed model was used to predict the control yield and yield response for each nutrient (obtained from model 0, c.f. 3.5.1) according to a set of environmental covariates. This model was used for every nutrient response variable with the same set of environmental covariates.

For all models, normality and homoscedasticity of the residuals was verified visually (Zuur et al., 2009) by plotting residuals and their distribution against model fitted values. The *lme* function in the *nlme* package was used for model 1 and 1b while the *lmer* function from the *lme4* package was used for model 2, 3 and 4.

3.5.3. *Spatial regression and geostatistical models*

When analysing spatial data, spatial dependence is the case where the dependant variable at a given location is correlated with observations at other locations (Anselin & Bera, 1998) leading to the violation of independent and identically distributed errors. In order to detect spatial autocorrelation, the residuals of (1) were evaluated using Moran's index of autocorrelation and statically test using *lm.morantest* from the *spdep* package. If spatial correlation is detected, it can be accounted for in a regression model.

For this study, two spatial approaches were used, a geostatistical approach to spatial regression and a spatial autoregressive approach. Both methods have been used and compared for agronomic data (Colonna et al., 2004; Lambert et al., 2004) in order to account for spatially correlated residuals. These approaches imply the assumption of spatial stationarity and isotropy (Dormann et al., 2007). Spatial stationarity means that spatial autocorrelation and the effect of biophysical process is constant across the region. Isotropic spatial autocorrelation means, on the other hand, that the processes causing spatial autocorrelation acts in the same way in all directions.

The geostatistical approach is based on the generalized least squares (GLS) approach where the spatial covariance structure is directly modelled in the covariance matrix Σ of ε, $N(0, \Sigma)$. The mean function of the model (1) is estimated via ordinary least square (OLS) and the spatial structure of the residuals is assessed with an empirical semivariogram using the *variogram* function in the *gstat* package. The semivariogram $\gamma(h)$ is a function that measures the spatial dependence between two districts separated by a distance h and characterized by its parameters (range, nugget, sill). Its parameters are used to model the "spatial" covariance matrix Σ characterized by the range, sill, and nugget of the semivariogram (Lambert et al., 2004). Parameters of Σ are estimated by restricted maximum likelihood (REML) and the coefficient of the regression model are re-estimated, this time adjusted for spatial autocorrelation, by replacing the elements of Σ by its estimates (Gelfand et al., 2010). Using the *georob* package in R the experimental semivariogram can be fitted to the appropriate variogram model by weighted non-linear least squares and include in the linear model via the *georob* function by Gaussian REML (Papritz, 2018).

Alternatively, the spatial autoregressive approach aims to correct for spatial autocorrelation, not by directly including the spatial covariance structure in Σ but by modelling the process generating spatial correlation with neighbourhood matrices. Here, the values of the residuals at a given location is modelled as a function of the values at adjacent locations (Dormann et al., 2007). The spatial error specification for this type of model correspond to a situation where the correlation in the residual is due to the omission of a spatial explanatory variable leading to inefficient estimates, and correct for it (Lambert et al., 2004). For this model the error term is expressed as $\varepsilon = \lambda W\varepsilon + \xi$ where W is a weight matrix defining the neighbourhood structure, λ is the autoregressive coefficient and ξ is non-spatially correlated error term. First, a neighbourhood criterion was selected based on a radius distance around a central point and a spatial weight matrix was created to identify neighbors. The criterion used here was a radius distance of 10 km that is the shortest distance between points allowing all point to have neighbors. After testing for spatial dependency in the residuals the spatial error regression was modelled with the *errorsarlm* function from the *spdep* package.

3.5.4. Random forest

The last approach for predicting yield response is based on regression trees, that make no assumption about the linearity of the variable to predict in their relation to the predictors. Alternatively, to linear regression, the data can be partitioned into "boxes" defined by threshold values of each predictors. The main advantage of this approach over the linear regression is there is no assumption that the functional form is the same throughout the range of predictors. However, regression trees are subjected to high variances. For improving the stability of this method, random forests, developed by Breiman, (2001) can be used. It consists in the building of a large number of regression trees using random sets of observations, or bagging, that have their prediction averaged over the number of trees performed. The package randomForest in R was used to model yield response to nutrient according to a given set of covariates.

3.5.5. Model evaluation and validation

Concerning regression models with fixed and/or random effect, residuals ($\boldsymbol{\varepsilon}$) distribution was evaluated graphically. First homogeneity was checked by plotting the residuals against the fitted values of the model, normality was checked using quantiles plot function and independence was evaluated by plotting residuals against explanatory variable values.

Moreover, the yield response prediction performances were evaluated by cross validation using two resampling methods. The data was first divided (50/50 split) into training and validation set. The parameters of the training model are used to predict unknown yield values in the validation set. The first resampling method described as "farm" cross validation was performed by randomly splitting the data in two at every district. The purpose here was to maintain an equal number of farms of training and validating within each district. It allowed to evaluate prediction on farms being in similar conditions to the training set. The same approach (50/50 split) was used at the district level, this time by training the model on a set on randomly selection districts and testing on the rest of them. The prediction performances of the model were tested on conditions that are likely to be more heterogeneous than the testing set.

3.5.6. *Workflow summary*

A summary of the procedure for analyzing yield response, based on the approaches describe above, is presented in figure 4.

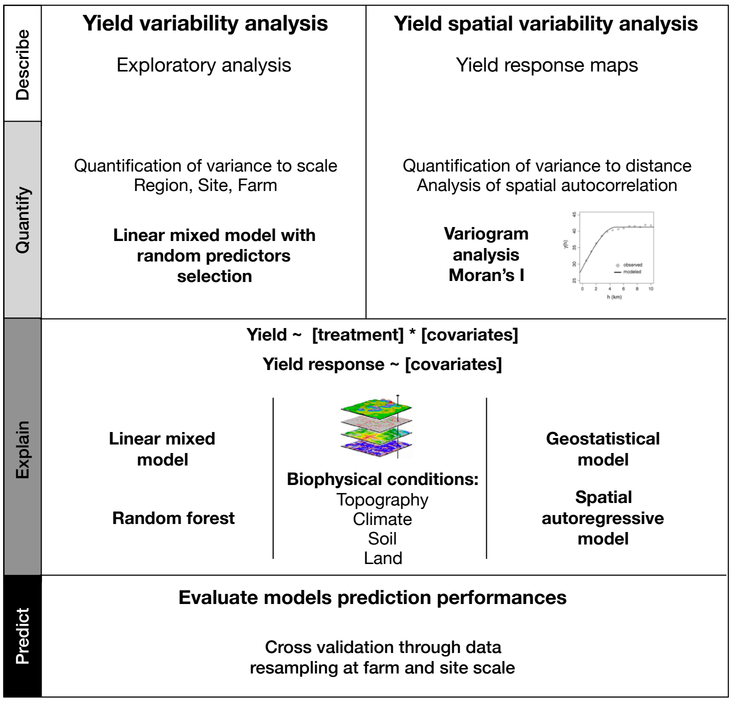

Figure 4. Scheme of the data analysis procedure

4. Results

4.1. Effects of fertilization treatments on maize yield and its components

In this section, plant parameters and grain yield of maize are analyzed under the effect of the 6 fertilization treatments and the 3 fertilization rates. In addition, the plot conditions at harvest and their relations to grain yield were investigated through a number of plant parameters. These parameters were the following: number of plants harvested, number of cobs per plant, cobs individual fresh weight.

4.1.1. Plant parameters

Large variability was found in the proportion of plants harvested per plot within and to a lower extent between district when comparing to the expected number of plants from sowing density. The proportion of harvested plants should reflect the proportion plant established from sowing and is likely to vary according the plot growing conditions. District median proportion of harvested plants (Fig. 5) ranged from 58.7% (Kilolo) to 109 % (Hanang). The presence of values above the 100% threshold indicated that numerous plots were having more plants than it was expected from the sowing density, leading to a potential overestimation of the grain yields.

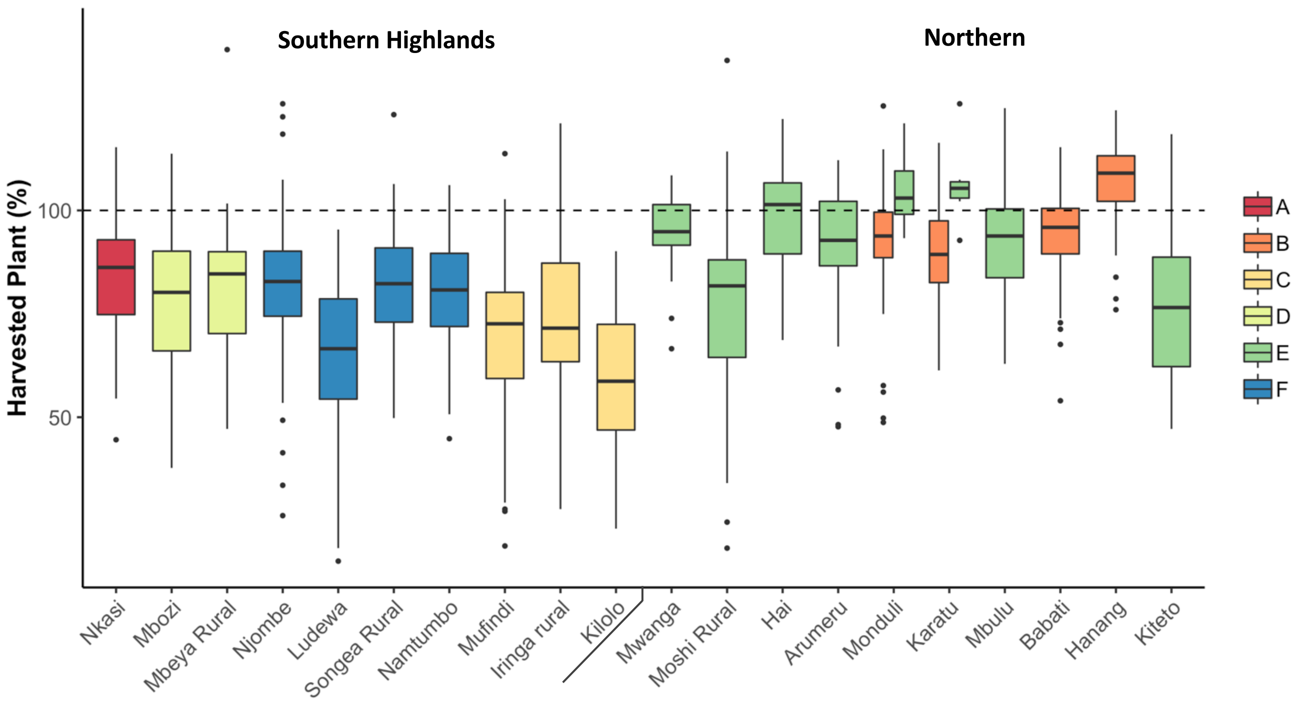

Figure 5. Boxplot representation of the proportion of harvested plants (%) on the total number of plants at sowing in a given plot for the different district. Districts on the left of the axis break correspond to the Southern Highlands zone and to the Northern zone on the right. Letters (A, B, C, D, E, F) and their respective colour stand for the different enumerators.

The relationship between maize grain yield and proportion of plants harvested differed between the Northern and the Southern Highlands zone (Fig. 6a) with significant higher values of harvested plants per plot in the North (model 1, + 11%, P = 0.001). Moreover, larger variation in grain yield was found for this zone (Fig. 6a) with a higher proportion of harvested plants while in the Southern highland, variability in yield increased with increasing proportion of harvested plants. Significant interaction between zones and treatment was found on the proportion of harvested plants (model 1, P < 0.001) because of lower values for NK treated plots only (Fig. S3a), observed in the Mufindi and Kilolo districts of the Southern Highlands zone. The effect of fertilization rates was not significant (model 1, P=0.068), neither its interaction with treatment (model 1, P=0.133). In our case, the proportion of harvested plants was not affected by the treatment, indicating that within a farm, the number of plants established was not affected by the fertilization.

The number of cobs per plant in a given plot was affected differently for the fertilization treatments when comparing the two zones. The significant interaction zone*treatment (model 1, P < 0.001) indicated that cobs per plant was highly responsive to every fertilization treatment in the Southern highlands but was unaffected in the Northern zone (Fig. S3b). Estimated means ranged from 0.52 for the control up to 0.80 cob per plant in the full fertilization (NPKM) plot in the Southern zone (Fig. S3b). Every fertilization treatment was significantly different from the control, however, N and P fertilized plots (NP, NPK and NPKM) were not different from each other (avg. LSD = 0.08, P < 0.05). The average number of cobs per plant were significantly higher in the North (model 1, P < 0.001). The number of cobs per plant showed a strong correlation with grain yield (Fig. 6b) up to the value of one cob per plant in the Southern Highlands. For the Northern zone, most values were clustered around one cob per plant with grain yield showing high variation from about 2000 to 10 000 (kg ha^{-1}).

Individual cob fresh weight showed a strong linear relationship with grain yield (Fig. 6c), however, for similar individual cob weight, maize yield was higher in the Northern zone. Variability of cob weight was deemed higher in the Southern highlands with very low values associated with low yields (Fig. 6c). The interaction of the treatments with zones was significant (model 1, P < 0.001) as well as the effect of fertilization rate alone (model 1, P = 0.023). No interaction was found between fertilization rate and treatment (model 1, P= 0.600). Cob weight was significantly higher in the NP fertilized plots compared to the control (Fig. S3c) in both zones. However, no significant difference between zones was observed for these plots (LSD = 0.02, P < 0.05). Moreover, PK and NK fertilization treatment had significant higher yield than the control in the Southern Highland, while only the NK application was significantly higher in the Northern zone. This indicated a stronger response to P fertilizer in the Southern Highlands, even with the absence of N fertilization.

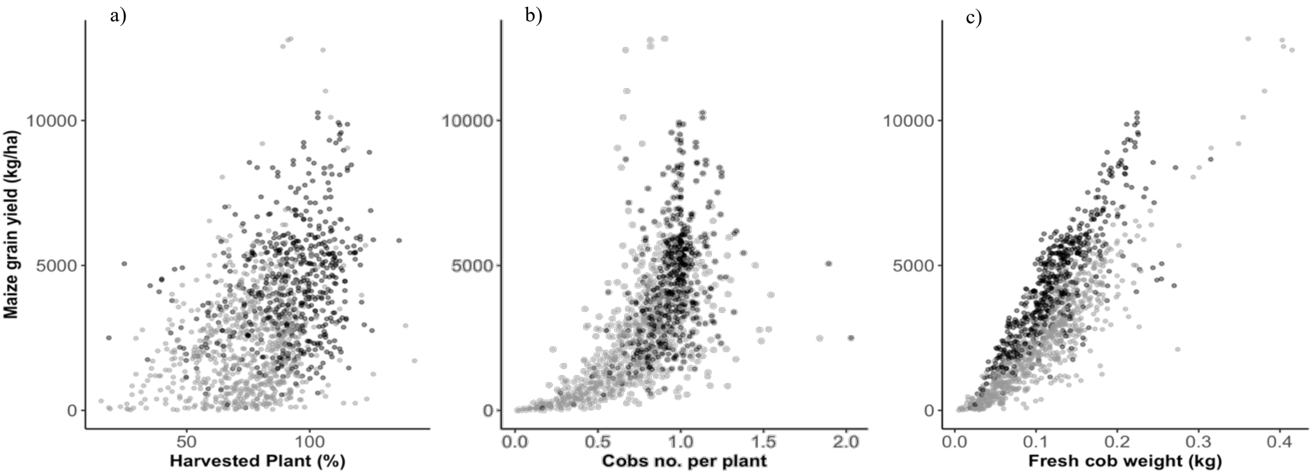

Figure 6. Relation between maize grain yield (kg ha^{-1}) at 12.5 moisture content for all plots and (a) proportion of harvested plants per plot (%), (b) plot average of number of cobs per plants, (c) plot average individual fresh cob weight (kg). Black and grey colours stand for the Northern zone and the Southern Highlands zone respectively.

4.1.2. Average yields

For all farms, average maize grain yield was strongly affected by fertilization treatment (model 1, $P < 0.001$) with a positive increment of 1074 kg ha^{-1} for the NK plots up to 1575 kg ha^{-1} for the NPK plots, compared to the control. There was no significant difference between the control and the PK treatment, neither between the N and P fertilized plots (NP, NPK, NPKM) (avg. LSD = 182.3, $P < 0.05$). However, the significant interaction between treatment and zones (model 1, $P < 0.001$) indicated a stronger response to P fertilization in the Southern Highlands (Fig. 7) even without N application (avg. LSD = 657.6, $P < 0.05$). Grain yield was further increased by the combination of N and P application up to 2763 kg ha^{-1} (avg. LSD = 657.6, $P < 0.05$). For the Northern zone, only the application of N increased yield significantly (Fig. 7).

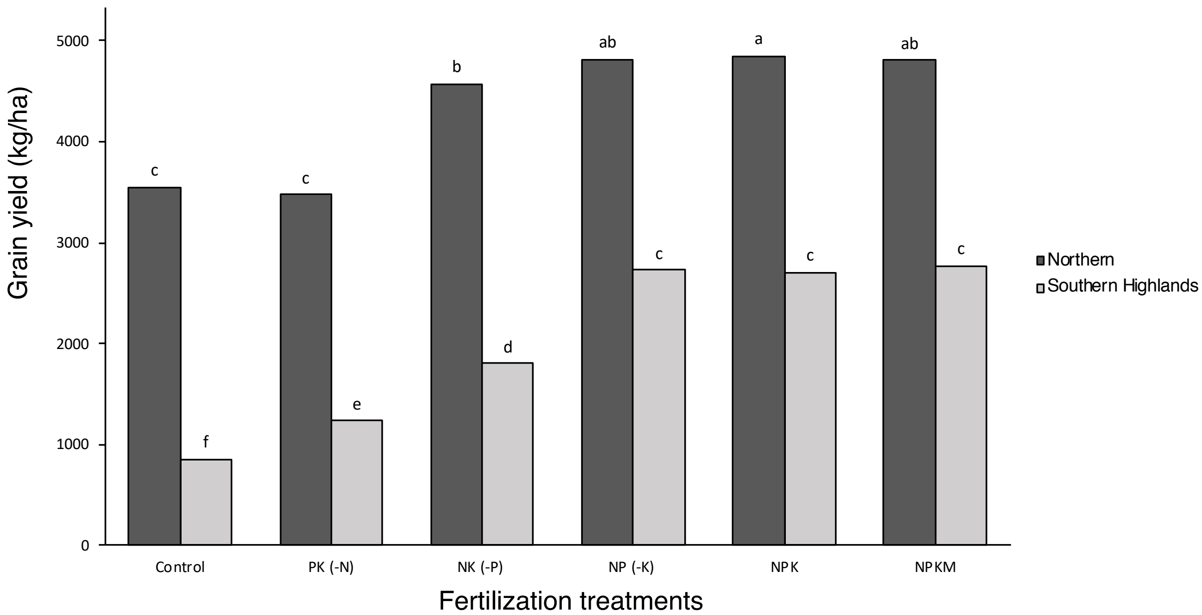

Figure 7. Predicted means of maize grain yield (kg ha^{-1}) for all farms for each fertilization treatment, model 1. Different letters stand for significant difference between treatments*zones, by pairwise comparison ($P < 0.05$).

When analyzing the two zones separately to study the effect of the fertilization rates (Fig. S3f, S3g), only the Southern Highlands showed a significant interaction between treatment and fertilization rate (model 1b, $P < 0.001$, S3f). Indeed, response to NK fertilization was significantly higher (+ 991 kg ha^{-1}, $P < 0.001$) for the medium rate compared to the high rate. This indicated a lower response to P fertilization in the following districts: Songea Rural, Namtumbo, Iringa Rural and Kilolo. However, fertilization rates were confounded with the district locations and it was not possible to conclude if the differences in average NK responses was due to the rates or the locations.

4.1.3. Agronomic efficiency

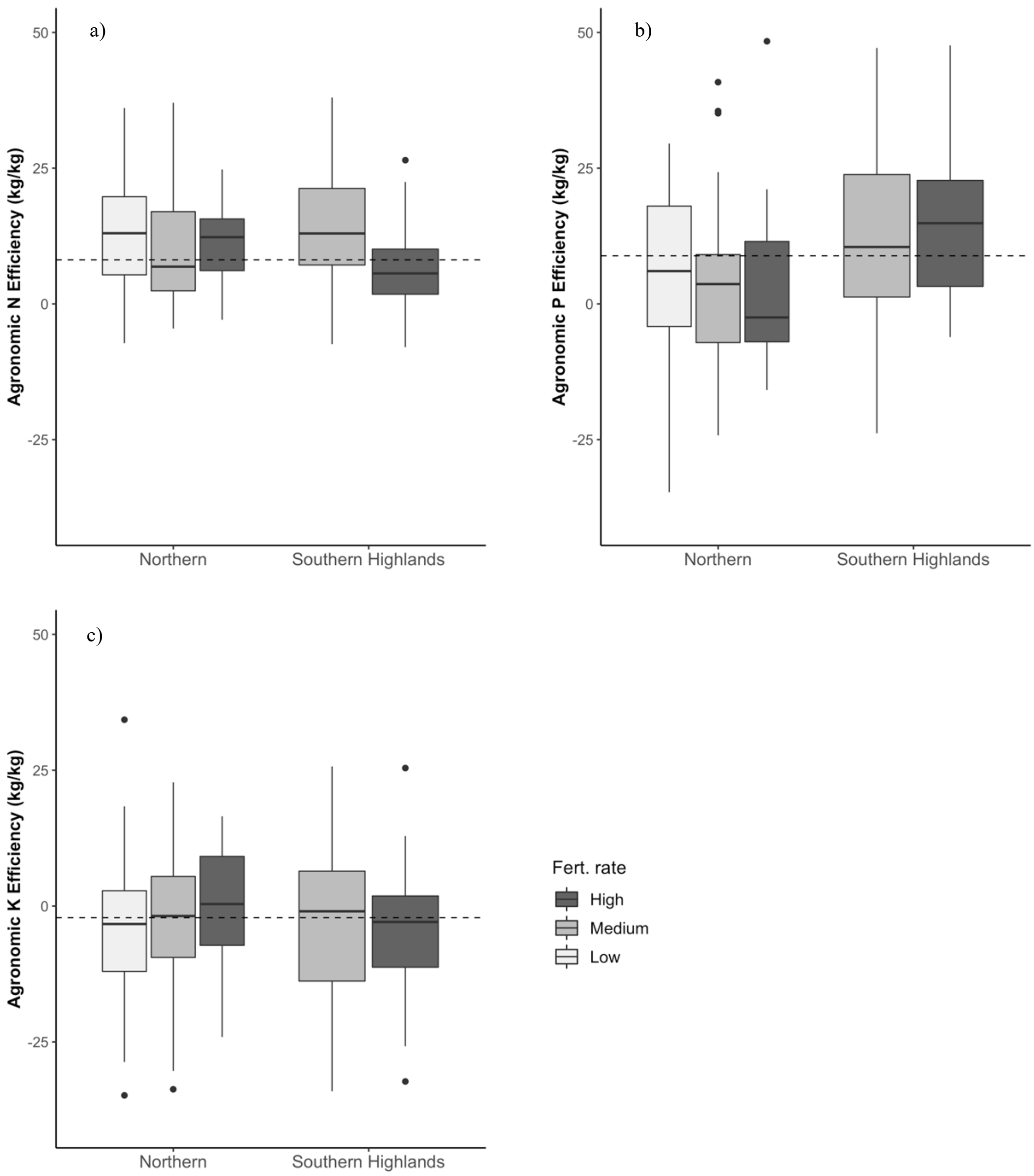

Figure 8. Boxplot representation of agronomic efficiency (kg kg^{-1}) of N (a), P (b) and K (c) for the Northern and Southern zones according for low, medium and high rate of fertilization. Dashed lines represent the median agronomic efficiency of all farm and are 8.11, 8.86 and -2.13 for N, P and K respectively. Different grey scales stand for the fertilization rates.

While fertilization rate was found to have only an effect in the Southern Highlands, the agronomic efficiency of the application of N, P and K was assessed to highlight the profitability of these different rates. Application of increasing rates was associated with increasing target yields. Agronomic efficiency was low for N and P with a median of 8.11 and 8.86 kg kg^{-1} respectively (Fig. 8).

Large variability of agronomic efficiency between farms with identical fertilization rate was observed (Fig. 8). For the Northern zone, agronomic P efficiency tended to be lower with increasing fertilization rate while the opposite can be observed in the Southern Highlands, highlighting the importance of P fertilization in the latter. Median agronomic K efficiency was negative (-2.13 kg kg^{-1}) with a large proportion of farms showing a negative response to K application. However, increasing rates of K fertilization in the Northern zone were associated with increasing agronomic efficiency (Fig. 8).

4.2. Yield variability in fertilization treatments

Best average yields were obtained under the combined application of N and P. However, responses to nutrients were highly variable between farms within districts (Fig. S4). In this section, variations of yield between fertilization treatment are analyzed, as well as variations resulting from the individual responses to N, P, K and secondary nutrients.

4.2.1. Treatment yields variability at the farm scale

Grain yields from fertilized plots varied substantially between farms for a wide range of control yields, from 0 to 8000 kg ha^{-1} (Fig. 9). PK application yielded the same as the control, supporting the importance of N fertilization for all farms. Highest variation in PK yields was observed for low values of control yield (0-2500 kg ha^{-1}) and numerous farms did not benefit from this application. Farms associated with low control, clustered in the bottom left corner (Fig. 9), mainly concerned plots located in the Southern Highland zone, in the regions of Rukwa, Mbeya and Southern Iringa (Fig. S4).

Yield variability was larger when N was applied in absence of P. Treatments with NK fertilization increased yields up to 5500 kg ha^{-1} for control yield between 0 and 3500 kg ha^{-1} and response to NK tended to level off after than range. It was observed that when P and K were applied in addition of N (NP, NPK and NPKM), high variation in yield was observed for a broader range of control yield, up to 5000 kg ha^{-1}. The combination of N and P fertilization resulted in a high number of plots yielding more than 1000 kg ha^{-1} for a range low control yields (0 to 1000 kg ha^{-1}) of the Southern Highlands. Fewer farm were below the 1:1 line when N and P were applied (Fig. 9). Plus, NP, NPK and NPKM yields tended to be less clustered in the bottom left corner (Fig. 9) supporting the evidence of N and P deficiency for these farms. In addition, while a constant increase in yield for N and P fertilized plots in the Northern zone was observed (Fig. 9), plots in the Southern Highlands leveled off at high control yields (Fig. 9). This indicated a lower attainable yield in the Southern Highlands.

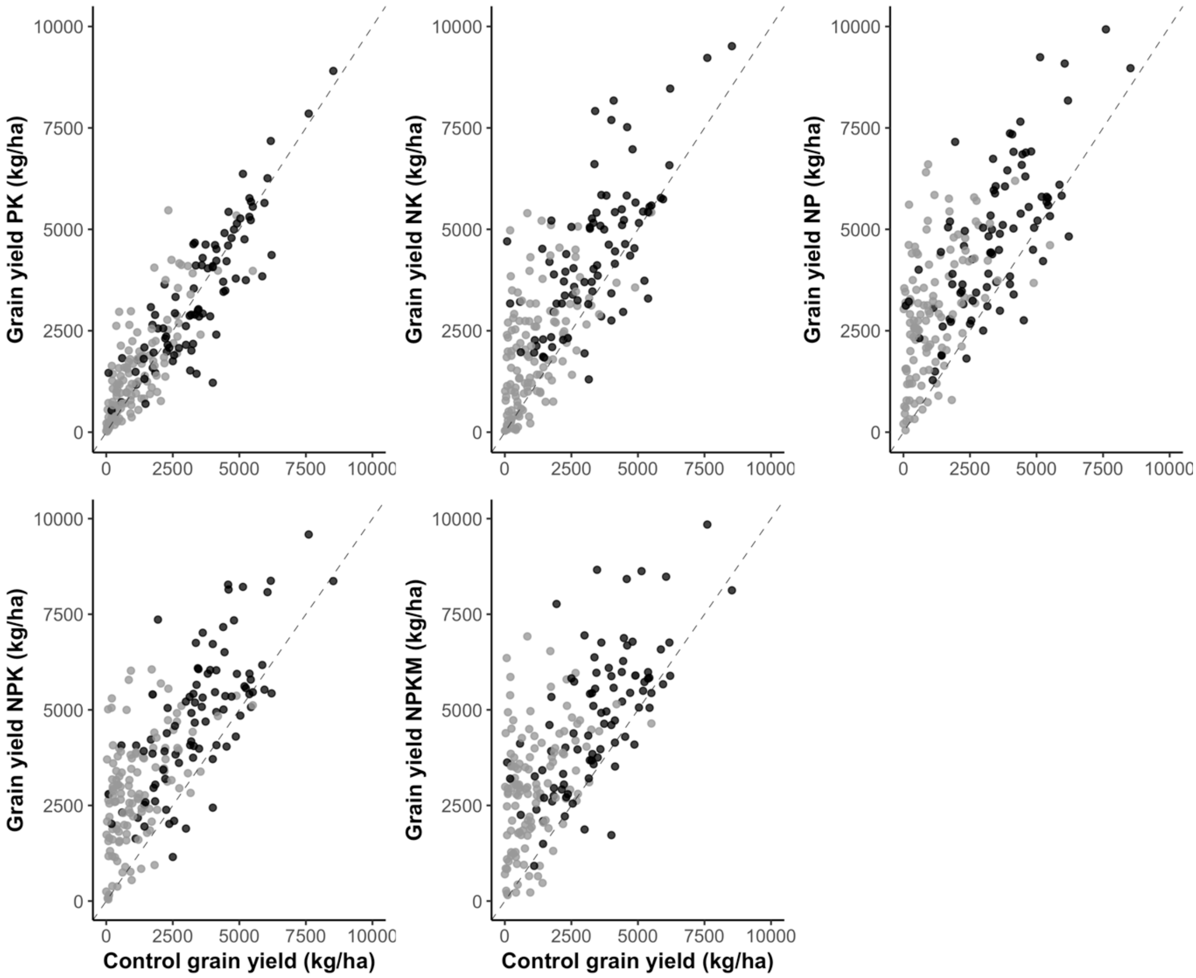

Figure 9. Yield (kg.ha^{-1}) of individual farms in fertilized plots over control yield. The dashed diagonal represents the 1:1 line where control and fertilized plot yield are equal. Black and grey colours stand for the Northern zone and the Southern Highlands zone respectively.

4.2.2. Stability analysis at the district scale

From the stability approach used by Raun et al., (1993) districts treatment median yields are compared with the districts median yields (all treatment confounded). Districts median yields resulting from P and K application (-N) were very close to the control, indicating little impact from application of these nutrients alone (Fig. 10). However, the intercept of the linear regression was higher for the PK treatment and pointed out a more positive response to PK in districts with very low yields.

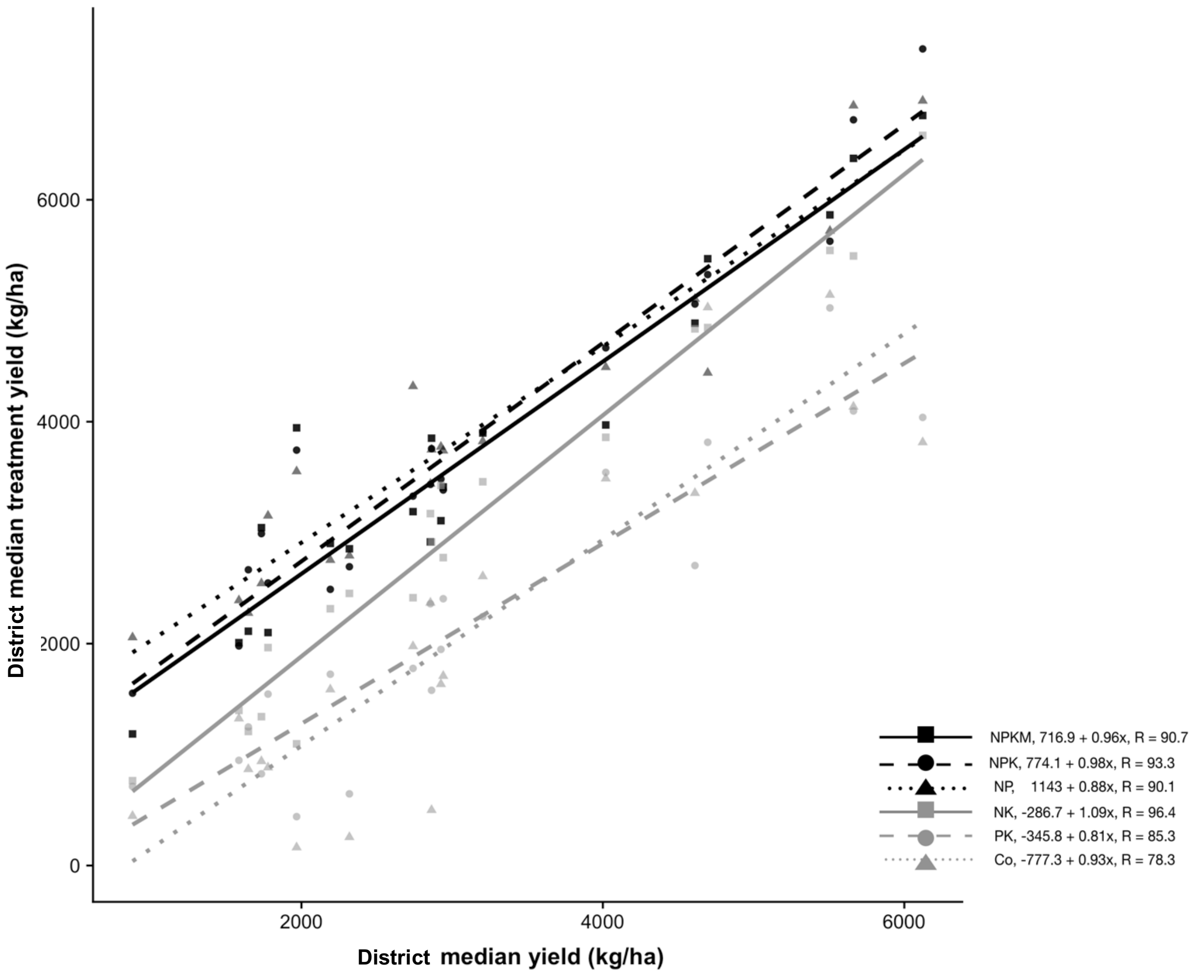

Figure 10. Stability analysis of treatment median yield over district median yield (all treatments) for the 20 districts. Regression equation of every fertilization treatment and adjusted coefficient of determination are displayed on the bottom right.

From the interpretation of Shehu et al., (2018), the slopes of the regression lines translate the sensitivity of a treatment to the environmental conditions. NK treatment had the highest response in the high yielding districts and the highest slope (1.09) compared to other nutrient combinations indicating a higher sensitivity to the environmental conditions, soil available phosphorus for instance. It can be assumed here that low yielding districts are more P deficiency than high yielding districts. Thus, response to NK is likely to be lower in the low yielding districts.

When N and P were applied (Fig. 10, black regression lines), little differences were found in term of intercept and slopes between the different treatments. NP treatment had a higher response in low yield districts and a slightly lower slope (0.88) in comparison to NPK (0.98) and NPKM (0.96) treatments. These results pointed out that responses to K and M are likely to be slightly higher in high yielding districts.

4.3. Yield variability in nutrient responses

4.3.1. Variability of yield nutrient responses at the farm scale

At the lowest scale of variability, the estimation of yield nutrient responses (model 0) allowed a quantification of the within farm variability. The root means square error of the relation between observed and estimated farm responses to nutrient was 617.4, 439.7, 468.5 and 469.9 for N, P, K and M respectively.

The yield responses to N, P, K and secondary macro- and micronutrients did not show a strong relation with the control yield (Fig. 11). As observed before with yields in fertilized plots (Fig. 9), the yield responses for all nutrients were varying strongly for the entire range of control yields. Difference in response to P were again observed between the two zones, with decreasing P response for high control yield in the Southern Highlands. In the Northern zone, the variation in P response increased with high control yield, with more farms being negatively affected by P application. High responsive districts, Nkasi, Mbozi and Mbeya rural, showed large variation within districts with an interquartile range (1st quartile – 3rd quartile) situated between 835 to 1194 kg ha^{-1} (Fig. S5). Such variability was also observed in low responsive districts such as Mbulu, characterized by an interquartile range of 1863 kg ha^{-1} (Fig. S5).

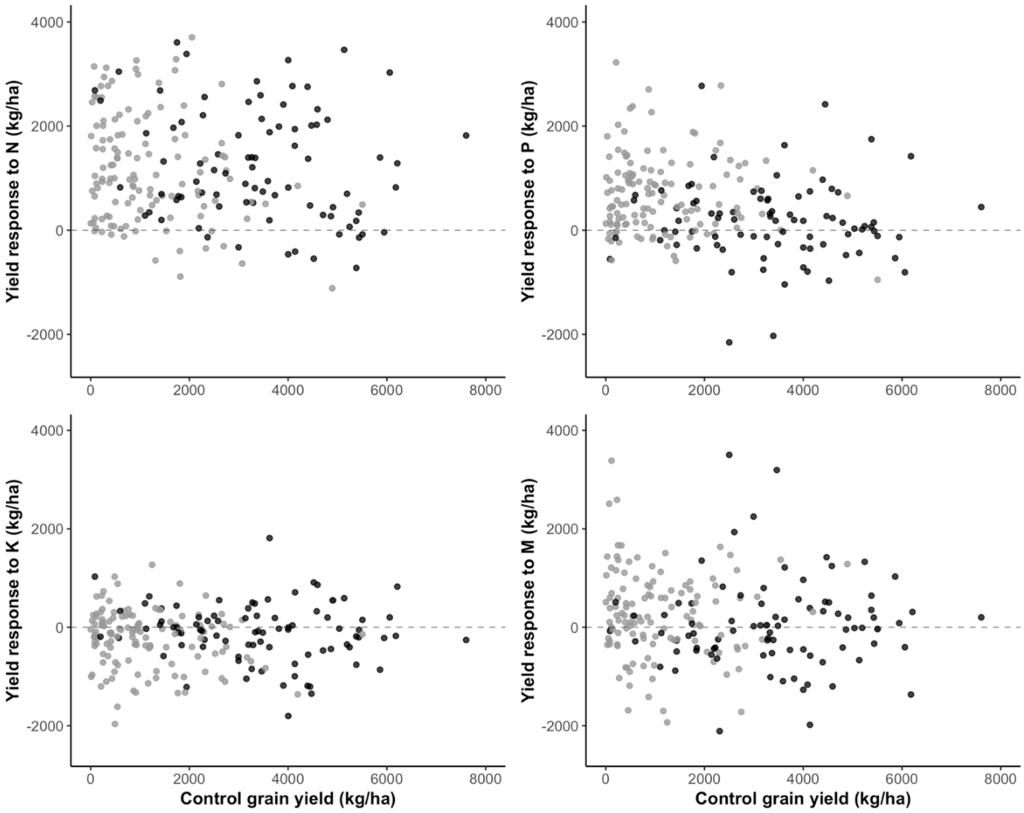

Figure 11. Yield (kg.ha^{-1}) responses (estimated from model 0) on individual farms to N, P, K and secondary macro and macronutrients over control yield. Dashed horizontal line indicates no response. Black and grey colours stand for the Northern zone and the Southern Highlands zone respectively.

N responses varied greatly between districts and less between farms within the same district, with the exception of Mufindi, Songea Rural (Southern Highlands) and Karatu, Hanang (Northern) (Fig. S5). High responsive (> 1900 kg ha^{-1}) districts such as Mbulu, Babati, Karatu and Mbozi showed high variation with and interquartile range between 819 and 1690 kg ha^{-1} and this range of response was even high in low response districts (< 500 kg ha^{-1}) such as Njombe, Hanang and Moshi rural (Fig. S5). Responses to K and M was relatively low and often negative, for the entire range of control yields (Fig. 11).

Responses to K showed less variability within districts and only a few districts (Mbulu, Hai and Mwanga) had a consistent gain from K application while most of the districts showed none, or consistent negative responses. The latter concerned the Mbozi, Mbeya rural, Ludewa, Iringa Rural, Kilolo in the Southern Highland and Babati, Kiteto in the Northern zone (Fig. S5). However, responses to secondary nutrients was very inconsistent. Yield responses were varying greatly between districts but particularly within districts (Fig. S5) such as Nkasi, Mbozi (Southern Highlands), Mbulu, Babati (Northern) were the response range reached 4000 kg ha^{-1}.

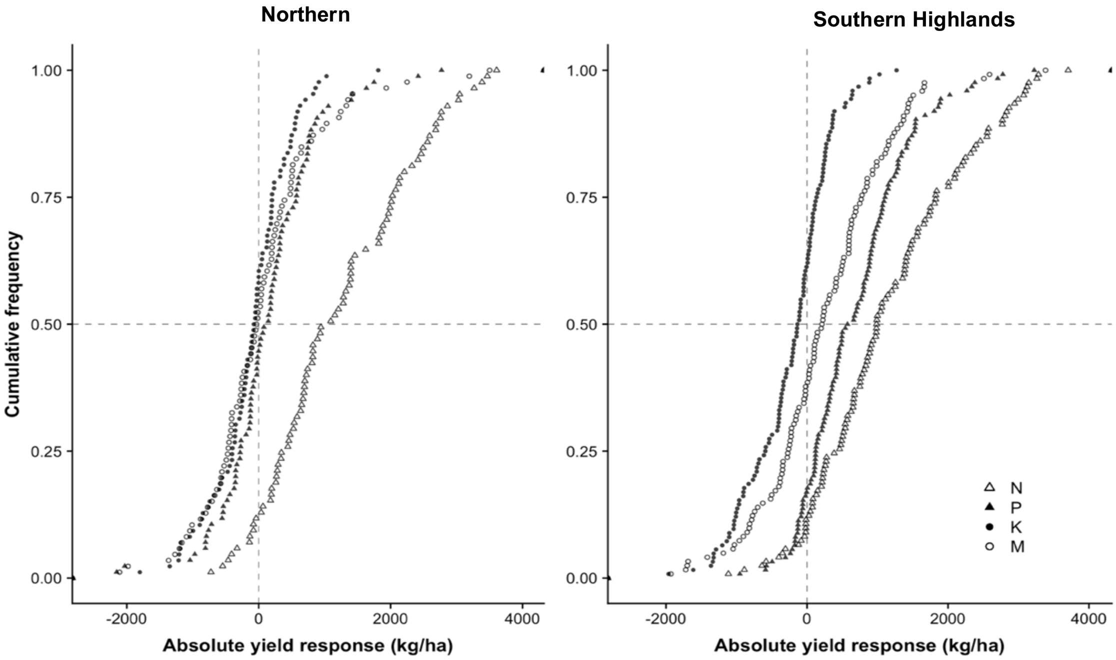

Figure 12. Cumulative frequency of yield (kg.ha^{-1}) responses (estimated from model 0) on individual farms to N, P, K and secondary macro and macronutrients. The left panel represents farms located in the Northern zone, the right panel represents the Southern Highlands zone. Dashed lines marks the response of 0 kg.ha^{-1}

The range of yields can be interpreted with the cumulative frequency of yield responses (Fig. 12). It indicated the proportion of farm achieving more or less a given response to a particular treatment, in comparison to the control. These curves allow to draw important conclusions regarding the performances of a given nutrient application and the risk associated with it. For both zones, it was observed that more than 90% of the farms benefited from the application of N while for P application, only 50% found benefits in its application in Northern zone compared to 80% in the Southern Highlands. In this zone, 50% of the farmers had a yield increase higher than 650 kg ha^{-1} from P application and higher than 1000 kg ha^{-1} from N application (Fig. 12 right panel). In the Northern zone, 50% of the farmers had an N response higher than 1100 kg ha^{-1} but only 110 kg ha^{-1} from the application of P (Fig. 12 left panel). Such trends support the importance of P application in the Southern Highlands. In both zones application of M was profitable for slightly larger proportion of farms than application of K. However, more farms (60%) in the Southern Highlands were benefiting from M application compared to the Northern zone (50%). A more horizontal curve, as observed for high N responses, indicated a less predictable response for this nutrient.

4.3.2. Nutrient response and soil fertility

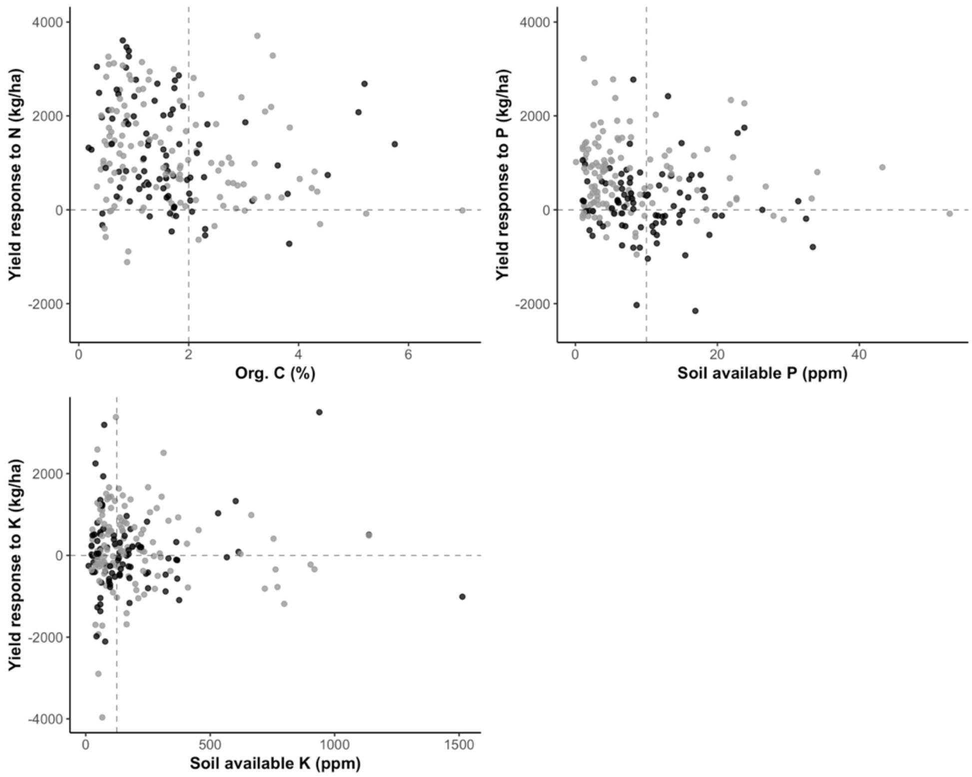

Figure 13. Yield (kg ha^{-1}) responses (estimated from model 0) on individuals farms to N, P and K over soil organic carbon, soil available P and soil available K respectively. Dashed horizontal line indicates no response and dashed vertical lines represent critical threshold for each soil properties (2% for org. C, 10 ppm for soil P and 125 ppm for soil K) Black and grey colours stand for the Northern zone and the Southern Highlands zone respectively.

Response to N were not related to soil organic matter content represented by the percentage of organic carbon (Fig. 13) most of the farms were below the threshold of 2% (Musinguzi et al., 2013). Within the 0-2% range, high variation of response was found, varying between 0 and 3 500 kg ha^{-1}. Maximum P response were found in farms below the 10 ppm threshold of soil available P and tended to decreased after this limit. It can also be observed that P deficient farms are mainly located in the Southern Highlands and showed the highest response to this nutrient. Yield response to K were highly variable and low for a wide range of soil available K (0-400 ppm) and did not show any correlation with soil available K.

4.3.3. Scales and nutrient responses variability

In view of the wide variability observed above, it seemed that variation in yield responses differed between nutrients, but also between geographic scales. The correlations between observations at a given scale gives an indication of the scale at which the observations may vary (Fig. 14).

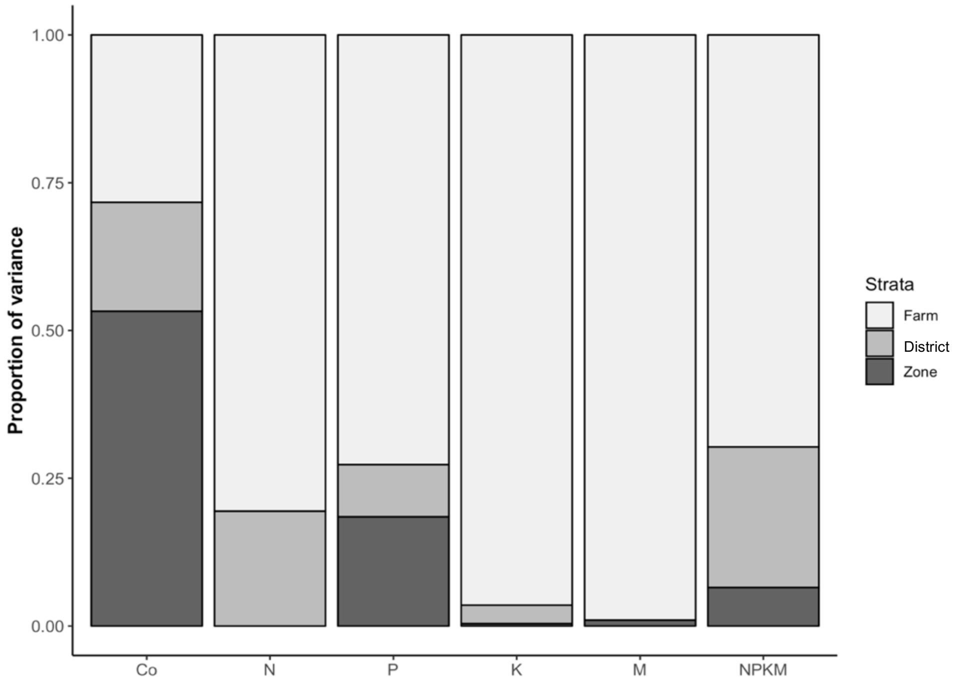

Figure 14. Contribution of the geographic scale (strata) on the variance components related to control (Co) yields and yield responses to N, P, K, M (estimated from model 0) and their combination (NPKM**).** Different colours stand for the different strata. Variance components were extracted from model 2, per nutrient.

For the control yield, 56.3% of the total variation is accounted for zones, and 26.0% for districts. Variance in the response to N was mainly attributed to unexplained variation at farm scale with 75% of the total variance. The 25% of variance left was accounted for between-district variations and might indicate the presence of factors with some consistent effect within districts. Similarly, variation in response when all nutrients are applied (NPKM) are accounted for 25% by differences between districts. Response to P, with zone differences accounting for 22% of the total variance, are likely to be under the influence of large-scale processes impacting the response to this specific nutrient. This supports again the extent of P deficiency in the Southern Highlands. Variation of K and M responses was due to the unexplained farm variation, 94.0 and 99.6 % respectively.

4.4. Spatial dependence of yield response

4.4.1. Detection of spatial dependency

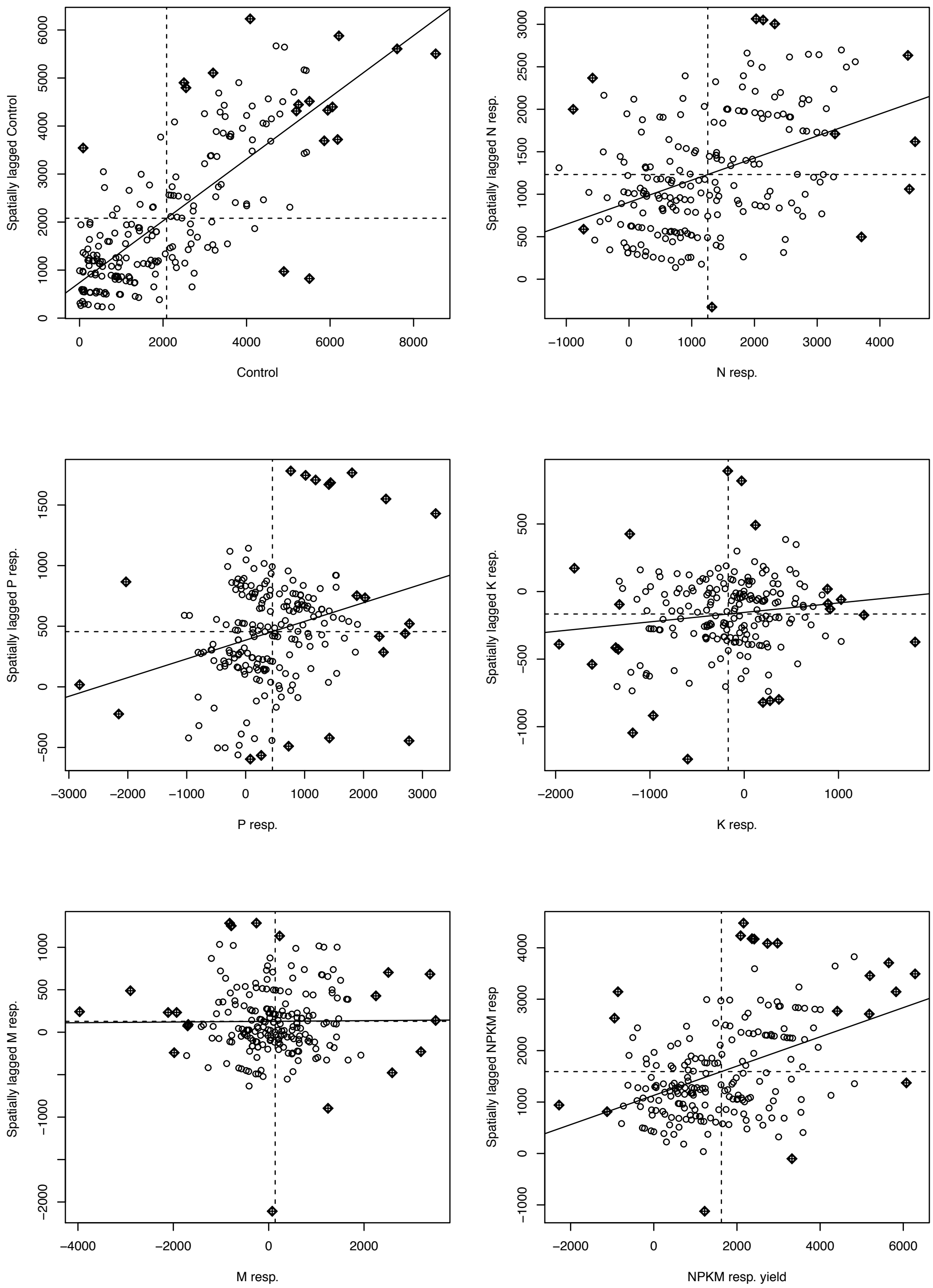

Figure 15. Moran scatterplot for nutrient responses. Diamonds data points represent high influence measures. Spatial lag represents the weighted average of neighbouring values defined by the neighbour criterion (here 10 km).

Spatial autocorrelation was assessed using Moran's index on control yield and predicted nutrient responses (N, P, K, M, NPKM). Control yield, response to N, response to P and response to NPKM showed a significant positive spatial autocorrelation ($P < 0.001$) when testing Moran'I on the entire dataset. Control yield showed the strongest spatial autocorrelation with an index of 0.642. In figure 15, it is clear that, low values of control yield tended to be surrounded by low values (lower-left quadrant). A similar trend can be observed with the response to N. However, it was observed that numerous points were located in the lower-right and upper-left quadrant. These quadrants correspond to negative spatial autocorrelation and can be interpreted as dissimilar values at neighboring locations. Regarding control yield and NPKM responses together, the Moran scatterplot indicated a stronger positive correlation for low values of yield in comparison with high yield values (upper-right). K and M responses did not show any spatial autocorrelation ($P > 0.05$) indicating a random distribution.

Table 5. Moran'I of nutrient responses and significance for all locations, Northern zone and Southern Highlands

Variables	Moran'I					
	All	p-value	Northern Zone	p-value	Southern Highlands	p-value
Control	0.643	**< 0.001**	0.475	< 0.001	0.175	**< 0.001**
N response	0.256	**< 0.001**	0.419	< 0.001	0.149	**< 0.001**
P response	0.154	**< 0.001**	-0.017	0.9409	0.073	0.053
K response	0.070	0.057	0.060	0.320	0.056	0.128
M response	0.004	0.822	-0.024	0.857	0.008	0.697
NPKM response	0.284	**< 0.001**	0.248	< 0.001	0.261	**< 0.001**

A closer examination of each zone indicated that spatial autocorrelation of control yield and N response was stronger in the Northern zone compared to the Southern Highlands (Table 5). Responses to P were not significantly spatially correlated within zones (Table 5). However, it was the case when testing on the entire datasets and may indicate that similarities between neighboring values may be observed at larger scales when including both zones in the analysis.

4.4.2. Quantifying spatial dependency over distance

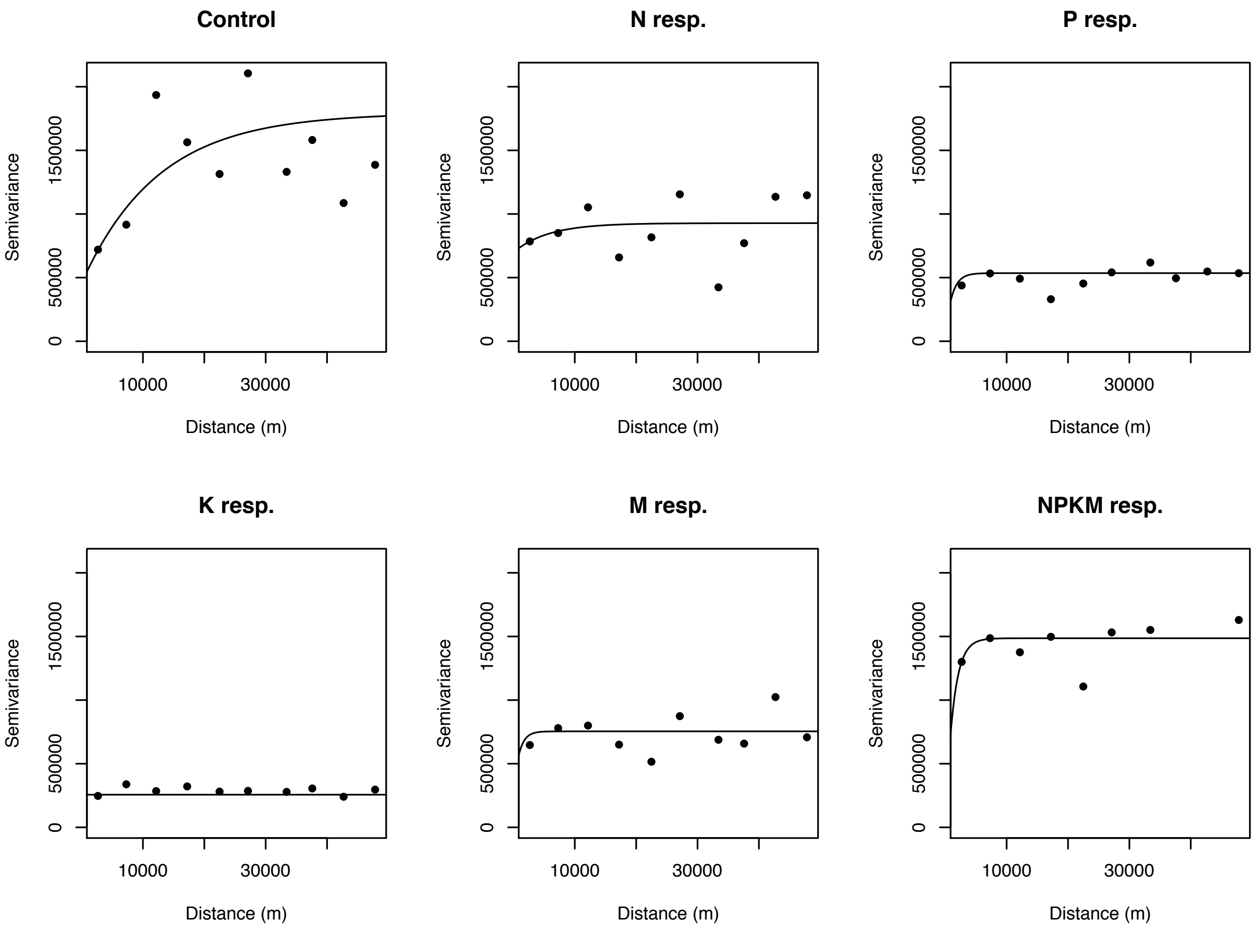

Figure 16. Nutrient responses empirical variogram (dots) and with a fitted exponential variogram model (lines)

Spatial dependency was illustrated with an exponential variogram model to estimate the nugget, range and sill of the empirical variogram (Fig. 16). Every variogram showed erratic behavior. This was assumed to be the result of a low number of sample points and the very clustered structure of the data points. The separation distance in which pair of points are included to estimate the semivariance was 50 km. Lower values did not result in interpretable variograms and larger ones did not show clear increase of the semivariance. However, a clear spatial dependency was found for the control yield and the response to N. These variograms showed dependence to 12.4 km and 5.7 km respectively according the range giving by the model. However, the empirical variogram for the control showed a larger range of about 20 km. Moreover, N response and control yield were observed to have a similar short distance variance (nugget) but the semivariance associated with control yield increased greatly over distance and reach a higher sill than the response to N. Other nutrient responses were difficult to model due to unstable parameter values. Response to P, K, M and NPKM, did not show spatial dependency.

4.5. Characterizing growing conditions for variable selection.

4.5.1. Heterogeneity of soil properties

Table 6. Soil physico-chemical properties of the districts

District	Sand (%)		B.D. (kg dm^{-3})		C.F. (%)		RZD (cm)		Org. C (%)		N (%)		Meh.P (ppm)		S (ppm)	
S. Highlands	62.0	4.4	1362.0	86.0	4.0	2.2	143.5	9.6	1.1	0.8	0.1	0.1	5.9	4.8	8.1	3.7
Mbozi	59.0	2.2	1335.5	36.3	6.5	2.2	145.0	5.9	1.4	0.7	0.1	0.0	**1.5**	1.0	8.4	2.5
Mbeya Rural	51.0	3.0	1304.5	1.5	6.0	0.7	**150.0**	0.0	1.9	0.5	0.2	0.0	3.3	0.6	12.0	0.8
Mufindi	61.5	2.2	1327.5	26.7	4.0	0.7	**150.0**	0.0	1.6	0.4	0.2	0.1	5.6	1.5	11.7	3.9
Iringa rural	**67.0**	7.4	1436.0	36.3	**2.0**	1.5	139.0	11.9	0.5	0.1	0.1	0.0	10.1	8.4	**5.9**	1.9
Kilolo	63.5	3.7	**1453.5**	20.8	2.5	3.0	119.0	9.6	0.7	0.3	0.1	0.0	9.3	5.1	6.7	1.0
Northern	41.5	8.9	1345.5	50.4	5.0	3.0	141.0	13.3	1.5	0.8	0.1	0.1	9.2	5.3	11.1	3.4
Mwanga	54.5	3.0	1329.5	57.1	**8.0**	1.5	146.0	5.2	1.4	0.6	0.1	0.1	8.9	5.5	9.3	3.3
Moshi Rural	38.0	3.7	1382.5	17.8	5.0	1.5	**40.0**	48.2	1.8	0.4	0.2	0.0	14.3	3.2	10.4	1.7
Hai	40.5	1.5	**1285.0**	23.0	7.5	0.7	**150.0**	0.0	**4.1**	2.1	**0.4**	0.2	6.5	4.7	10.6	5.0
Arumeru	40.0	1.5	1347.0	41.5	7.0	1.5	115.0	0.0	1.4	1.0	0.1	0.1	12.2	12.6	12.7	4.8
Monduli	36.0	3.0	1324.0	11.9	7.0	1.5	95.0	29.7	1.6	0.2	0.1	0.0	9.2	2.3	**15.0**	4.7
Karatu	**23.0**	4.4	1294.5	40.8	7.0	0.7	130.0	22.2	1.5	0.8	0.2	0.0	7.5	6.8	10.6	3.6
Mbulu	55.0	7.4	1392.0	29.7	3.0	1.5	141.0	5.9	0.9	0.1	0.1	0.0	8.2	3.0	10.8	3.8
Babati	42.5	5.9	1351.0	25.2	3.5	1.5	**150.0**	0.0	1.3	0.6	0.1	0.1	**15.4**	3.8	10.8	1.7
Hanang	56.0	7.4	1345.0	4.4	4.0	0.0	**150.0**	0.0	1.1	0.6	0.1	0.1	6.9	1.2	13.3	2.3
Kiteto	**67.0**	3.7	1382.5	53.4	4.0	0.0	**150.0**	0.0	**0.4**	0.1	**0.0**	0.0	6.8	6.9	9.9	2.8
CD (%)	21.4		3.0		40.0		13.2		39.2		41.7		45.8		25.4	

B.D.: bulk density, C.F.: Coarse fragments, RZD: Root zone depth

Values presented are median (left) and median absolute deviation (right) per district of each soil properties. Max and min values are highlighted in bold

Table 7. Soil physico-chemical properties of the districts

District	pH		Ca (ppm)		Mg (ppm)		K (ppm)		Na (ppm)		ECEC (cmol kg^{-1})	
S. Highlands	6.1	0.4	432.0	358.8	105.1	86.5	101.7	92.8	16.1	3.4	3.4	2.3
Mbozi	**5.8**	0.3	496.0	237.2	173.7	91.9	281.5	40.6	**20.7**	0.0	4.7	2.0
Mbeya Rural	6.3	0.1	1000.0	231.3	**286.1**	90.1	**713.6**	176.8	17.3	3.4	9.4	2.6
Mufindi	5.9	0.4	512.0	456.6	95.4	62.1	62.6	34.8	18.4	3.4	3.4	2.1
Iringa rural	6.1	0.1	284.0	258.0	46.8	27.0	54.7	17.4	16.1	3.4	2.0	1.4
Kilolo	6.4	0.8	364.0	252.0	117.2	54.0	142.7	60.9	13.8	3.4	3.1	1.3
Northern	6.8	0.4	1158.0	790.2	95.4	54.9	117.3	89.9	18.4	3.4	7.0	4.5
Mwanga	7.6	0.3	1044.0	410.7	185.3	70.3	121.2	49.3	16.1	1.7	7.6	2.0
Moshi Rural	**8.2**	0.5	1467.0	470.0	125.8	45.9	584.5	568.1	16.1	17.0	9.6	2.5
Hai	6.6	0.1	**1826.0**	154.2	145.8	18.9	320.6	69.6	**20.7**	1.7	**11.0**	0.4
Arumeru	6.5	0.6	1694.0	284.7	104.5	36.0	179.9	179.7	**20.7**	6.8	10.2	2.3
Monduli	7.3	0.7	1656.0	198.7	99.6	9.0	164.2	69.6	18.4	3.4	9.8	0.8
Karatu	6.9	0.1	1202.0	194.2	94.2	6.3	105.6	37.7	18.4	1.7	7.0	1.0
Mbulu	**5.8**	1.2	384.0	115.6	54.7	19.8	58.7	5.8	16.1	3.4	2.8	0.5
Babati	6.8	0.1	948.0	136.4	84.4	45.9	62.6	37.7	13.8	0.0	5.9	1.8
Hanang	6.8	0.1	1036.0	717.6	81.4	36.0	117.3	104.3	13.8	3.4	6.4	4.2
Kiteto	7.0	0.2	**224.0**	65.2	**24.3**	11.7	**27.4**	5.8	18.4	0.0	**1.5**	0.4
CD (%)	6.9		58.3		43.5		59.6		20.0		53.9	

Values presented are median (left) and median absolute deviation (right) per district of each soil properties. Max and min values are highlighted in bold

District aggregated soil properties give an indication of the heterogeneity that can be found between the study districts. Soils in the region of Manyara (Mbulu, Hanang, Babati, Kiteto) and in the Southern Highlands tend to have a higher sand content than regions in the Northern zone. This was associated with a lower organic carbon content and higher bulk density in this zone. Between districts, available soil P varied strongly but with a lower range and median for the Southern Highlands (5.9) compared to the Northern zone (9.2). The lowest values were found the region of Mbeya (Mbozi and Mbeya rural) (Table 6). Soil K showed very high variation between and within districts (Table 7). Often, variation of soil K was associated with other cations (Ca and Mg), as illustrated by the districts Mbeya Rural, Mbozi, Moshi Rural and Hai. Soil water pH tended to be lower in the Southern Highlands with values ranging from 5.8 to 6.4 than the Northern zone with 5.8-8.2. Districts near the Kilimanjaro mount (Moshi Rural, High, Arumeru) showed medium to high pH (up to 8.2) and were associated with high values of Ca, Mg and K (Table 7).

Table 8. Soil micronutrient content of the districts

District	Zn (ppm)		Cu (ppm)		Mn (ppm)		Fe (ppm)		B (ppm)	
S. Highlands	4.6	2.2	2.3	1.1	45.4	27.3	68.3	24.6	0.1	0.1
Mbozi	5.3	1.3	2.0	0.8	35.0	10.9	**35.2**	11.8	0.16	0.04
Mbeya Rural	11.8	3.9	2.1	0.8	**9.5**	4.9	42.5	1.4	0.17	0.10
Mufindi	4.6	2.2	2.3	1.1	43.9	26.1	88.7	24.6	0.06	0.04
Iringa rural	**3.1**	1.1	2.6	0.5	46.0	26.9	77.2	7.6	**0.04**	0.04
Kilolo	5.3	2.2	2.3	0.5	67.6	26.9	60.6	15.1	0.08	0.07
Northern	6.3	2.6	2.1	1.5	114.5	49.3	92.4	34.0	0.2	0.1
Mwanga	6.1	0.0	2.6	1.1	100.9	40.3	68.3	15.1	0.09	0.03
Moshi Rural	9.8	4.1	3.7	1.1	104.4	31.4	99.3	37.8	**0.33**	0.14
Hai	**15.5**	7.4	**0.5**	0.0	130.4	24.0	**134.6**	20.9	0.21	0.16
Arumeru	10.6	5.9	0.8	0.4	81.1	93.4	99.3	20.9	0.25	0.13
Monduli	4.6	2.2	2.1	0.9	111.9	7.6	75.2	25.5	0.22	0.13
Karatu	4.6	2.2	3.1	0.5	**163.6**	15.2	72.3	25.5	0.10	0.06
Mbulu	4.6	2.2	1.5	0.9	89.4	50.0	63.7	17.0	0.07	0.03
Babati	6.8	3.3	**5.0**	1.9	120.6	52.3	126.7	46.7	0.16	0.08
Hanang	9.0	6.6	2.7	0.9	85.3	34.9	115.3	25.5	0.18	0.06
Kiteto	5.3	1.1	2.1	0.5	119.1	50.0	109.6	12.7	0.19	0.06
CD (%)	32.75		39.26		43.06		27.91		50.00	

Values presented are median (left) and median absolute deviation (right) per district of each soil properties. Max and min values are highlighted in bold

Soil micronutrients showed relatively little variation between zones than macronutrient (Table 8) with the exception of soil Mn and Fe. The median values of these nutrient being higher in the Northern zone.

4.5.2. Heterogeneity of climatic conditions

Table 9. Length of the growing period, third quartile of average daily temperatures and total precipitation for the total growing period (GP) and during maize physiological stages

District	LGP (days)		T. GP (°C)		T. I (°C)		T. II (°C)		T. III (°C)		Prec. GP (mm)		Prec. I (mm)		Prec. II (mm)		Prec. III (mm)	
S. Highlands	207.0	11.1	27.9	2.0	29.6	2.9	27.8	1.8	26.1	1.9	1124.9	191.1	193.3	17.4	898.7	161.0	14.4	15.9
Mbozi	211.0	3.0	28.2	0.3	32.1	2.2	28.4	1.0	27.8	0.5	1197.2	21	195.6	20	957.3	46	52.8	10
Mbeya Rural	215.0	0.0	27.0	0.5	**26.9**	0.6	27.0	0.4	26.3	0.1	1076.6	8	193.3	3	856.3	10	25.0	0
Mufindi	**222.5**	2.2	**25.4**	0.4	28.3	0.9	**25.7**	0.5	**23.9**	0.6	**1376.7**	96	194.0	8	**1093.7**	74	**71.5**	34
Iringa rural	199.0	7.4	27.6	1.2	29.5	2.0	27.9	0.9	25.7	0.9	991.0	119	175.5	39	799.4	111	8.5	5
Kilolo	201.0	1.5	30.0	0.9	36.1	2.0	29.9	0.6	26.8	0.7	990.3	15	**203.5**	19	782.4	15	6.4	1
Northern	162.0	10.4	31.8	3.6	34.0	5.3	29.9	2.8	28.9	2.8	523.5	159.1	79.9	68.1	348.5	123.1	12.6	18.7
Mwanga	**131.0**	0.0	28.4	3.0	33.8	5.1	26.0	3.2	27.3	2.7	354.4	117	167.4	32	**157.3**	58	8.1	5
Moshi Rural	166.0	4.4	33.2	1.8	**42.6**	2.4	31.1	1.0	30.3	1.5	449.9	128	62.3	14	340.8	141	19.2	9
Hai	163.0	1.5	**35.5**	3.1	41.2	0.7	26.8	1.0	29.7	0.5	608.2	0	69.6	0	538.6	0	**0.0**	0
Arumeru	159.0	4.4	34.9	1.6	40.5	1.7	**34.5**	1.5	29.6	0.7	438.1	87	79.3	7	349.2	74	**0.0**	0
Monduli	154.0	3.0	34.3	0.9	35.8	2.3	34.2	1.5	**30.7**	0.9	**301.4**	0	35.8	7	259.2	17	**0.0**	0
Karatu	160.5	2.2	33.2	0.7	35.0	1.6	32.6	1.5	25.9	1.4	326.3	15	**27.6**	1	299.0	12	**0.0**	0
Mbulu	172.0	11.9	28.9	0.9	30.0	2.0	29.1	0.7	25.9	1.9	612.2	48	80.8	15	486.3	27	59.7	23
Babati	193.0	1.5	28.3	1.0	33.1	1.4	28.2	1.3	25.7	1.8	628.2	37	129.7	16	476.1	45	23.6	6
Hanang	184.0	1.5	28.7	0.5	29.0	1.5	29.0	0.9	27.1	0.5	555.7	4	171.4	1	365.5	6	18.8	8
Kiteto	161.0	3.0	30.0	0.6	31.6	0.9	30.7	1.8	29.0	0.7	527.0	11	201.3	27	325.6	18	25.5	5
CD (%)	11.4		7.7		11.0		6.3		6.5		40.7		46.1		44.9		97.4	

Values presented are median (left) and median absolute deviation (right) per district of each soil properties. Max and min values are highlighted in bold
I: Onset of the growing period, from 10 days before sowing to 15 days after sowing
II.: Vegetative, early reproductive and anthesis growing stages, from end of onset to two third of the total growing period
III.: Grain filling and maturation growing stages, from end of II to harvest

The growing period was relatively longer in the Southern Highlands, that benefited from high amounts of rainfall (990 to 1376 mm) with most of the rainfall occurring during phase II (Table 9). A similar pattern was observed for the Northern Highlands. The lowest amounts of rainfall were associated with the districts situated in the Kilimanjaro and Arusha region with Mwanga, Monduli and Karatu that received less than 300 mm. Temperature regimes showed high values particularly during phase I in the Arusha and Kilimanjaro regions (Table 9).

4.5.3. Selecting variables for predicting maize yield responses

In view of the heterogeneity of growing conditions and the low correlations between these variables and the maize response yields under the different fertilization treatments, the selection of the main predictors was done using agronomic knowledge. The initial stepwise procedure of model 3 failed to select variables as a clear overfitting of the data was present when the model was fitted with farms nested into districts as random effects. As a result, the addition of any environmental covariates yielded no significant differences against the basic model. The variable selection was performed on the same model by removing the farm random component of model 3.

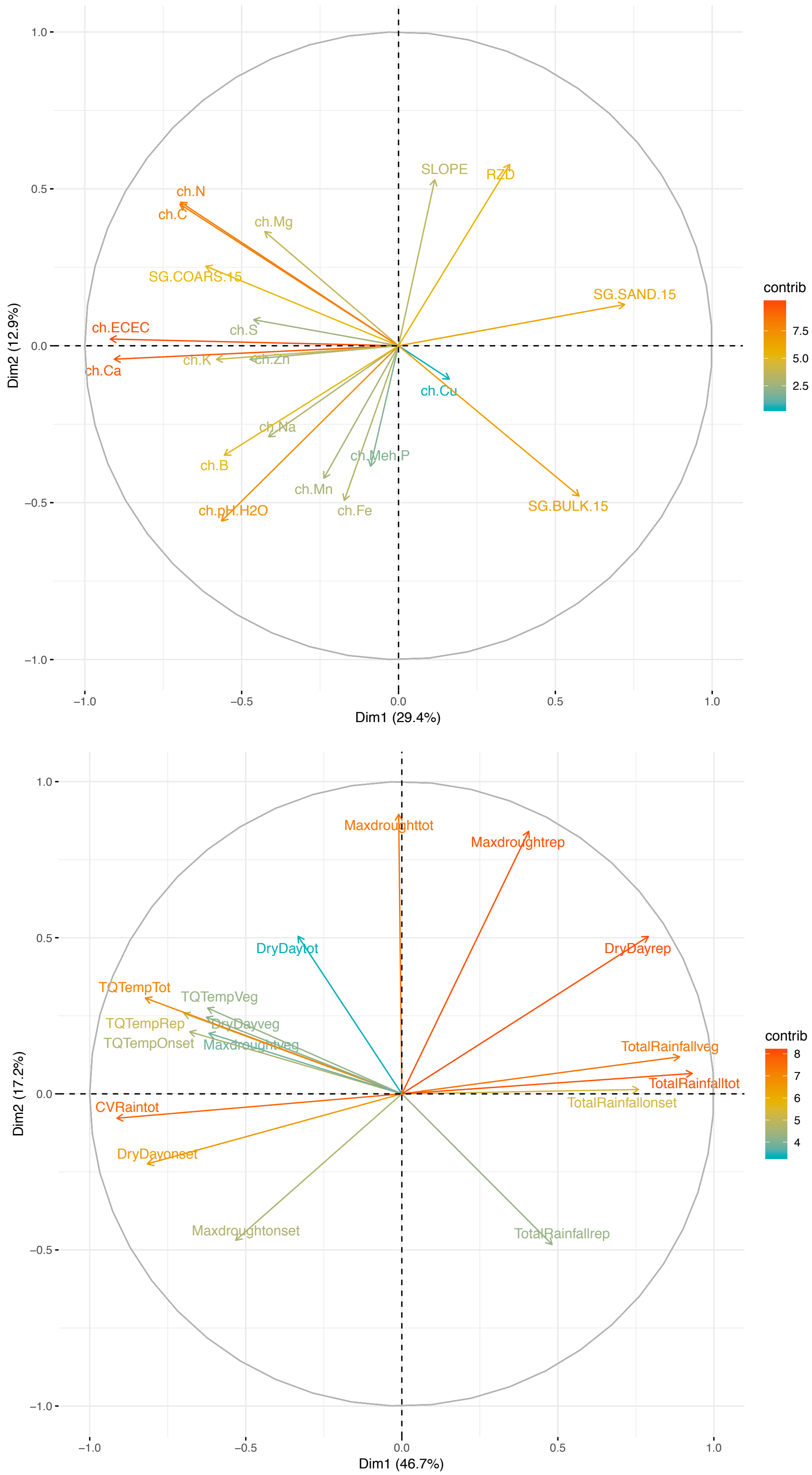

Figure 17. Principal component analysis of (a) soil properties and (b) climatic conditions of the potential predictors. Gradient colour indicates variable quality of representation ($\cos^2$). Suffix at the end of the variable code is “tot” for GP, “onset” for I, “veg” for II and “rep” for III.

The results of the principal component analysis (Fig. 17) helped to detect collinearity among the potential predictors and to make a first selection prior to modelling. For the soil physical and chemical properties (Fig. 17 upper panel) the first axes contained 29.4 % of the information. Soil Ca, ECEC, organic C, N and sand content contributed the most in this axe. Organic C and N content were confounded and inversely correlated with bulk density, as well as Ca and ECEC that were negatively correlated with sand content. Slope percentage were surprisingly positively correlated with root zone depth but negatively with micronutrient as Mn, Fe and B.

Regarding the weather variables, many were confounded, especially characteristics associated with the same growing stages (Fig. 17 lower panel) with drought characteristics of stage I and stage III. Temperatures (third quartiles of average daily temperature) were confounded between the growing periods and with drought characteristics of the stage II. The total amount of rainfalls was also highly correlated.

Table 10. Stepwise selection of predictors on maize yield variation corrected for treatment application (model 3)

	Model expression	p-value
a	Treatment + Soil physical properties (Sand, CF, BD, RZD, SLOPE)	0.607
b	Treatment + Soil chemical properties (pH, ECEC)	**0.005 ***
c	Treatment + Soil Nutrients (Org. C + Mehlich P + K)	**0.007 ***
d	b + Soil Micronutrients (Org. C + Mehlich P + K)	**0.002 ****
e	Treatment + Temperature (GP) + Rainfall (GP)	0.579
f	Treatment + Temperature (II) + Rainfall (II)	0.793
g	Treatment + Temperature (III) + Drought (I)	**0.033 ***
h	d + Temperature (III) + Drought (I)	**0.044 ***

p-value indicate significant difference from the basic model (treatment alone) via likelihood ratio test. When a sub model is added in the model expression e.g. “d”, the model is tested against it and not again the basic one.
Significant levels of 0.1, 0.05, 0.01, and 0.001significances were indicated with . ,*, **, and *** respectively

From the summary of the district characteristics and the principal component analysis, significance of soil and weather variables were tested (Table 10) to obtain a final model composed of pH, ECEC, org. C, Mehlich P, K, 3rd q. of the average daily temperatures during stage III (grain filling and maturation) and drought during stage I (onset of the growing period and plant emergence). Soil N and Ca were not included due to strong collinearity with org. C and ECEC respectively. Mg was not included in the model testing for simplicity as it was assumed to depend on variables that were already included in the model. Micronutrients were not included in the model because no strong deficiency was observed for most of them, and addition of B in model “d” (Table 10) did not improve the model significantly. Variance inflation factors were below 3 for all variables and residuals normality and homoscedasticity were validated by visual assessment.

4.6. Predicting yield nutrient response

4.6.1. Prediction with linear mixed-effects model

Table 11. Summary statistics of the mixed effects models and results of the farm and district cross validation

Variables Y ~ X	Estimates	s.e.	p-value	All	CV Farms	CV Districts
				R^2 (%)		
Control (kg ha^{-1})						
Drought (I)	-4.5	50.2	0.929	0.68	0.46	0.07
Temperature (III)	22.8	75.7	0.764			
ECEC	5.3	52.6	0.920			
Water pH	449.1	221.0	*0.044			
Org. C	333.2	144.0	*0.022			
Meh. P	-2.5	14.4	0.863			
K	-1.1	0.8	0.193			
Response N (kg ha^{-1})						
Drought (I)	-52.3	33.7	0.130	0.36	0.16	0.01
Temperature (III)	-26.0	55.6	0.642			
ECEC	48.9	45.6	0.285			
Water pH	-323.6	184.5	.0.082			
Org. C	-167.8	122.6	0.174			
Meh. P	10.6	12.7	0.407			
K	-0.3	0.7	0.658			
Response P (kg ha^{-1})						
Drought (I)	-47.3	24.3	.0.060	0.37	0.05	0.00
Temperature (III)	-98.6	40.4	*0.018			
ECEC	13.7	33.8	0.685			
Water pH	-67.4	136.1	0.621			
Org. C	-29.1	90.8	0.749			
Meh. P	1.0	9.5	0.912			
K	-0.4	0.5	0.354			
Response K (kg ha^{-1})						
Drought (I)	46.3	15.3	**0.006	0.17	0.01	0.02
Temperature (III)	22.7	26.8	0.403			
ECEC	29.5	25.7	0.254			
Water pH	-82.7	100.6	0.414			
Org. C	-2.1	69.6	0.975			
Meh. P	-12.2	7.5	0.107			
K	-0.1	0.3	0.801			
Response M (kg ha^{-1})						
Drought (I)	-4.6	25.7	0.857	0.01	0.12	0.01
Temperature (III)	2.5	45.3	0.957			
ECEC	-11.8	45.0	0.795			
Water pH	31.0	175.5	0.860			
Org. C	-31.4	123.1	0.799			
Meh. P	-5.9	13.4	0.661			
K	0.6	0.6	0.337			
Response NPKM (kg ha^{-1})						
Drought (I)	-42.4	45.0	0.353	0.34	0.04	0.00
Temperature (III)	-116.9	75.0	0.125			
ECEC	85.6	63.0	0.176			
Water pH	-439.9	253.7	.0.086			
Org. C	-216.6	169.4	0.203			
Meh. P	-1.0	17.7	0.956			
K	-0.5	0.9	0.548			

Estimates, standard errors and p-value of explanatory variables are presented from the mixed models using all data points.
R^2 represent the squared correlation between predicted and observed values for the testing set, where predictors have been obtained for the training set. Data were split for training and testing at the farms and the districts levels.
Significant levels of 0.1, 0.05, 0.01, and 0.001significances were indicated with . ,*, **, and *** respectively
Model 4 was used to predict the nutrient responses.

Mixed models were used to predict each of the nutrient response variable with as factors the maximum length of drought during onset, the third quartile of the temperatures during the grain filling period, effective cation exchange capacity, soil water pH, soil organic carbon, and soil available P and K. This set of variables explained the most variation in control yield with a R^2 of 68% (Table 11). The effect of soil water pH and organic carbon were significant ($P < 0.05$) for this variable. Response to N, P and NPKM had a reasonable part variation explained with a R^2 of 0.36, 0.38 and 0.34 respectively. Among this variable, only the drought (I) and temperature (III) were significant ($P < 0.05$) (Table 11). Cross validation at the farm and district level showed that model performances dropped to much lower values of R^2 compared to the model with all data points. When predicting on farms at neighboring locations i.e. within the same district, 46% of the variation of control yield could still be explained by the set of explanatory variables. For the nutrient responses, best prediction was achieved with a R^2 of 16% for the nitrogen response. In addition, prediction performances for farms in different districts from the training model were close to 0 indicating that relation observed between the covariates and the response were very different between the training and the validation districts.

4.6.2. Prediction with random forest

Table 12. Summary of the farm and district cross validation for the random forests

Variables Y ~ X	All	CV Farms	CV Districts
	R^2 (%)		
Control (kg ha^{-1})			
Water pH	0.36	0.26	0.08
Drought (I)			
ECEC			
Response N (kg ha^{-1})			
Meh. P	0.07	0.03	0.00
Water pH			
Temperature (III)			
Response P (kg ha^{-1})			
Drought (I)	0.20	0.13	0.11
Temperature (III)			
ECEC			
Response K (kg ha^{-1})			
Meh. P	0.06	0.01	0.12
Drought (I)			
Temperature (III)			
Response M (kg ha^{-1})			
Meh. P	0.00	0.00	0.01
Temperature (III)			
Drought (I)			
Response NPKM (kg ha^{-1})			
Temperature (III)	0.06	0.06	0.00
Meh. P			
Water pH			

R^2 represent the squared correlation between predicted and observed values for the testing set, where predictors have been obtained for the training set. Data were split for training and testing at the farms and the districts levels.
The first three most important covariates are indicated below the nutrient response variable.

Output of the random forest showed a similar trend in term of response performances with a low R^2 when testing on new farms (Table 12). Predictions performances were situated between 0 and 7% for N, K, M and NPKM response. Variation of control yield were still reasonably explained by the model with an out-of-bag prediction explaining 36% of the variation (Table 12).

4.6.3. Prediction with spatial autoregressive error model

Table 13. Summary statistics of the spatial error models and results of the farm and district cross validation

Variables Y ~ X	Estimates	s.e.	p-value	All	CV Farms	CV Districts
					R^2 (%)	
Control (kg ha^{-1})						
Drought (I)	-7.8	49.9	0.876	0.59	0.09	0.01
Temperature (III)	58.4	75.3	0.438			
ECEC	24.0	51.1	0.639			
Water pH	359.9	221.8	0.105			
Org. C	227.9	153.6	0.138			
Meh. P	1.7	13.8	0.903			
K	-1.1	0.8	0.158			
Response N (kg ha^{-1})						
Drought (I)	-54.3	32.4	.0.093	0.24	0.05	0
Temperature (III)	-48.5	53.0	0.360			
ECEC	29.5	43.5	0.498			
Water pH	-333.0	179.4	.0.063			
Org. C	-87.2	125.1	0.486			
Meh. P	9.7	12.0	0.421			
K	-0.1	0.6	0.917			
Response P (kg ha^{-1})						
Drought (I)	-51.1	20.6	*0.013	0.20	0.02	0.01
Temperature (III)	-77.7	35.4	*0.028			
ECEC	19.5	32.6	0.551			
Water pH	-132.1	130.2	0.310			
Org. C	-47.9	91.0	0.599			
Meh. P	-2.0	9.4	0.828			
K	-0.3	0.4	0.441			
Response K (kg ha^{-1})						
Drought (I)	46.2	14.2	*0.001	0.13	0	0.02
Temperature (III)	19.8	25.0	0.430			
ECEC	30.9	24.5	0.209			
Water pH	-82.3	96.0	0.391			
Org. C	-0.3	67.3	0.996			
Meh. P	-12.1	7.3	0.095			
K	-0.1	0.3	0.668			
Response M (kg ha^{-1})						
Drought (I)	-6.7	22.7	0.769	0.02	0.12	0
Temperature (III)	1.6	40.6	0.968			
ECEC	-15.3	42.3	0.718			
Water pH	29.7	162.4	0.855			
Org. C	-39.9	114.2	0.727			
Meh. P	-6.6	13.0	0.611			
K	0.7	0.5	0.196			
Response NPKM (kg ha^{-1})						
Drought (I)	-48.4	42.3	0.252	0.22	0	0
Temperature (III)	-111.6	70.5	0.113			
ECEC	81.0	60.2	0.178			
Water pH	-458.1	245.6	.0.062			
Org. C	-190.0	171.3	0.267			
Meh. P	-4.7	16.8	0.779			
K	-0.4	0.9	0.633			

Estimates, standard errors and p-value of explanatory variables are presented from the mixed models using all farms.
R^2 represent the squared correlation between predicted and observed values for the testing set, where predictors have been obtained for the training set. Data were split for training and testing at the farms and the districts levels.
Significant levels of 0.1, 0.05, 0.01, and 0.001significances were indicated with . ,*, **, and *** respectively

Spatial autocorrelation of the linear regression (with fixed effect only) residuals were tested with Moran's I. After accounting for the effect of the environmental covariates, residuals of control yield, response to N and NPKM were still autocorrelated with a Moran' I of 0.51, 0.22 and 0.16 respectively ($P < 0.001$). Regression residuals of P responses were not correlated after accounting for the effect of covariates. Minor differences were observed in the coefficients and standard error values of the models by comparing the mixed model and SAR approaches. For both, cross validation at farm level and district level is very low, even for the control yield in the SAR model (Table 13). Significant effect of the drought during onset was found for response to P and K and this was already observed for K in the mixed model approach.

4.6.4. Prediction with geostatistics

The variograms of the regression residuals are presented in S7. Fitting a variogram of the residuals and implement its parameter to include a spatial correlation structure in the linear models. As we encountered in 4.3.2. fitting a variogram model on the data was difficult and ended with convergence issue, that is, when the parameter value did not stabilize. Control yield residuals did not show any spatial correlation on long (50 km) and short distance (15 km) with a very erratic behavior. Model residuals of N and NPKM responses seemed to show a spatial dependence (between 0 and 15 km) at 50 km. However, a closer look at this relationship using 15 km as threshold distance did not yield more interpretable information. Regarding P and K response residuals, M being very erratic, a stable but poor spatial dependence was observed at a distance of 2-3 km. Due to the difficulty to obtain variogram parameter and in the presence of non-spatial dependency showed by the variogram, it was decided not to pursue the modelling exercise for this approach.

5. Discussion

This section evaluates the main outcomes of the analysis of the TAMASA campaign in Tanzania for the 2015-2016 maize growing season and discusses their potential implications regarding the formulation of fertilization recommendations. The main objective of this study was to characterize the variation of rainfed maize yield responses to mineral fertilizer application: nitrogen, phosphorus, potassium and secondary macro- and micronutrients and understand their sources of variability.

5.1. Response to mineral fertilizers across the study space

The aim of describing and quantifying yields and responses to nutrient variability across the study space was twofold. On one hand, it provided an estimation on how mineral fertilizer application may increase on-farm maize production, on the other hand it indicated to which extent yield response variation can be related to geographic scales. The existence of a spatial yield structure for maize yield was assumed in this study, the on-farm trials being distributed in a large area covering a wide range of growing conditions. The different scales at which yield limiting factors occurs was hypothesized to be correlated with the spatial variation of yield.

5.1.1. Zone and district scale variation of yields

Large differences in yield were found between the two zones with significantly lower control and treatment yield in the Southern Highlands (Fig. 7). Average yields in the Northern zone were three time higher than in the Southern highlands for the control, respectively 3541 and 851 kg ha^{-1}, and two times higher for the NPK treatment yield, respectively 4845 and 2698 kg ha^{-1}. Such a difference between the two zones was surprising for the NPK treatment. Indeed, climatic conditions are more favorable in the Southern Highlands (Magehema et al., 2014) and historical yield indicated higher average maize yields in this zone (Rowhani et al., 2011). Moreover, rainfall conditions reported in this study concorded with historical trends. It is important to note that the proportions of plants harvested in the experimental plots, in comparison to the expected plant density at planting, was significantly higher in the Northern zone (Fig. 5, S3a). The percentage of plants harvested was likely to vary to according the proportion established plants after sowing, depending on the quality of the field preparation and soil physical properties. The delimitation of the harvested area by the enumerator may also impact the number of plants harvested and the final yield. In figure 5, the highest yielding districts (Fig. S4) are also the ones with the highest proportion of harvested plants. However, the correlation between the number harvested plant was, in general not correlated with the final yield (Fig. 6). It is then difficult to draw clear conclusions regarding the consequence of this potential data collection bias.

N application at the zone and district levels (Fig. 7,10) was associated with a clear increase in crop yield compared to the control and the interactive effect of N and P was of importance in the lowest yielding environments mainly (Fig. 10). Decrease in maize yield in low production districts was particularly linked with a decrease in the number of cobs per plant (Fig 6; S3b), a component sensitive to nitrogen and particularly phosphorus fertilization (Selassie, 2016). Results of this analysis tend to converge to the conclusion that yield reduction through the number of cobs per plant was more pronounced in the Southern Highlands when compared to the reduction of cob weight (Fig. 6). The latter could be partly explained by stronger P deficiency in this zone, but it is important to keep in mind that different maize varieties were used. Traits of these varieties were unknown and might influence differences in plant parameters such as the number of cobs per plant.

5.1.2. Variation of nutrient responses and local scale variation

While aggregated values of grain yield at the zone and district level tend to demonstrate a clear effect of nitrogen and phosphorus application on yield, the magnitude of the response varied strongly, and particularly at local scale, i.e. between farms within a district. N application yielded the highest response with a median of 1060 kg ha^{-1} grain yield, this nutrient being recognized as the most limiting for maize in SSA (Kihara et al., 2016; Vanlauwe et al., 2011). However, response variation for this nutrient was also the largest with an interquartile range of almost 1635 kg ha^{-1} for the entire dataset (Fig. 11). Most of the farms achieved a higher yield in the N treated plots compared to the control but only 50% obtained an increase higher than 1000 kg ha^{-1} (Fig. 12). While 50% of the farms in the Southern Highlands observed a response higher than 650 kg ha^{-1} from P application (Fig. 12). P response variation was also highly heterogenous within districts (Fig. S5) with the highest interquartile ranges observed in Mufindi and Mbozi. High variation in the responses to M and K was observed and a benefit from potassium application was unlikely among the farms included in this analysis. On the other hand, while response to secondary macro- and micronutrient was low or null, some districts showed consistent benefits from this application: Nkasi, Ludewa, Namtumbo, Moshi Rural, Mbozi and Kilolo (Fig. S5). This highlights the importance of secondary macronutrients, here Ca, Mg, S and micronutrient Zn, B in specific conditions. Gain from this treatment has been already reported for several crops in SSA (Kihara et al., 2017).

The presence of response patterns across the study area was limited and variation of responses were weakly correlated with the different scales of analysis. Local variation, represented here by within district differences, is the principal component of nutrient response variation (Fig. 14). Geographic variation of responses was observed for N and P application indicating the presence of medium to large scale processes, between districts and between zones respectively. However, it represents only a small part of the total variation associated with the responses. Upscaling the analysis of yield responses at higher aggregation level is indeed uninformative in the present context as it hides most of the variation and results in misleading conclusion about the biophysical components of risk for the farmers (Ronner et al., 2016; Vanlauwe et al., 2016).

5.1.3. Explaining the variation of nutrient responses

The set of covariates used to predict nutrient responses resulted from a selection of significant yield constraints (Table 10). Only a small part of the nutrient response variation was explained by these variables. The best predictions, once cross validated at the farm level, were 16 % for the response to N (linear mixed model, Table 11) and 13 % for the response to P (random forest, Table 12). These values are much lower than reported in similar attempts by Bielders & Gérard, (2015) and Ronner et al., (2016). Indeed, while a reasonable part of the variation was expected from the model fitted on all the farms, especially for N and P response, prediction performances dropped dramatically to 0 % when predicting in new districts. This indicated clear overfitting when modelling with all farms. Furthermore, it highlights the fact that when predicting for unknown conditions (different districts), yield constraints in the training set are likely to be different than the ones in the validation set, resulting in large differences between modelled and measured yield responses.

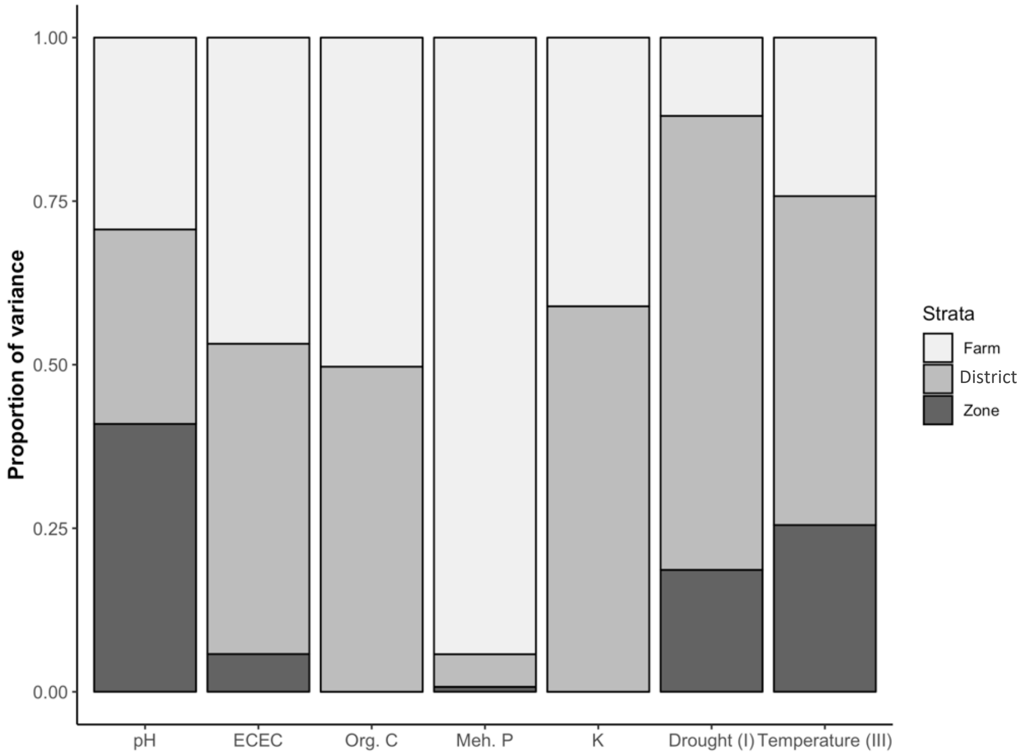

Figure 18. Contribution of the geographic scale (strata) on the variance components of the explanatory variables. Different colours stand for the different strata. Variances were computed from separated models for each variable.

The scales of variation associated with each environmental variable tend to be explained most by between districts and between zones differences for the water pH, ECEC, soil available K, drought during onset and temperature during the grain filling phases (Fig. 18). Variation of soil available P was however explained by local variations. A clear mismatch between the scale of variation of biophysical variables and the variation of nutrient responses was observed. Indeed, the best predictions, for N and P response were related to their scale of variation seen in figure 14. Response to K and M, varying principally at local scale, could not be predicted by these explanatory variables.

Regarding the multiple collected variables, no strong correlations were observed with the nutrient responses. All soils were situated below the bulk density critical threshold of 1600 kg dm^{-3} (Arshad et al., 1996) to limit root development. Moreover, the districts were associated with low and variable organic carbon content, less than 2% for all districts (Musinguzi et al., 2013) with exception of Hai in the Northern Zone where soil organic C was weekly correlated with N response (Fig. 13). Available phosphorus showed large differences between districts, with lowest values in Mbozi, Mbeya Rural, Mufindi, and were associated with medium to strong responses to P fertilization (Fig. S5). Some districts presented a potential soil K deficiency, with a content lower than 125 ppm (van Biljon et al., 2008). However, soil available K was not related to K response, the only district with a positive response to K was Mwanga. This district had a median soil K content of 121 ppm, just below the deficiency threshold. Positive response to this nutrient may be associated with a better tolerance to drought (Grzebisz et al., 2013) as this district was characterized by the lowest amount of rainfall during the vegetative and the early reproductive stages of maize. Considering the threshold value of 9 ppm for soil S (van Biljon et al., 2004), no strong deficiencies were observed. Regarding potential micronutrient deficiency: Cu, Zn, Mn and Fe were well above the deficiency thresholds (1, 0.45-1.77, 2, 4.5 ppm respectively (summarized in Kihara et al., 2016). Boron deficiencies might have been present in some districts with values lower than 0.15-0.5 ppm (Aref, 2011). However, null to positive response to M was observed in these sites.

Soil properties from chemical analysis and weather variables were unable to explain variation in nutrient responses. But associated with climatic variables, they could explain a large part of the control yield with 46 % (Table 11). In this study, control yield is the best indicator of the soil fertility variation that can be observed between farms. Between zone (large scale) variation accounted for more than 50% of the total variance of the control yield. Despite large difference in term of rainfall regimes they were not related with yield. However, soils of the Northern zone appeared to be more fertile. The dominant soil types in the Northern zone are Ferralsols in the region of Arusha, Cambisols and Vertisols in the region of Kilimanjaro and Cambisols and Luvisols in the Region of Manyara. On the other hand, Ferralsols and Acrisols are dominant for most study sites in the Southern highlands (www.soilgrids.org). These soils are often found in sub-tropical areas and are characterized by strong weathering their low capacity to supply and retain nutrients (Bationo et al., 2012). Hence, they require large amount of P fertilizer as phosphate is strongly adsorbed by iron and aluminum oxides (Bationo et al., 2012).

Our ability to explain nutrient responses variation through biophysical factors was, in conclusion, very limited. The variables selected represented mainly the large-scale processes and depicted the heterogeneity between zone and district.

5.2. Implications for the formulation of fertilizer recommendations

5.2.1. Downscaling the analysis to capture local variation

From our understanding of the response variation observed, local variation tends to overrule the total variation. A better understanding of the factors under this level of analysis is required as it is not possible to make predictions with biophysical variables. Hence, as reflected by van Heerwaarden et al., (2017) the lack of correlation between the response and the control yield indicates that productivity constraints, as we observed explaining variation of the control, may not be good indictors for the response itself.

Small-scale variation is likely to vary according to land use, management practices (current and historical) and resource endowment (Vanlauwe et al., 2006). In view of the poor correlation between soil analysis and the yield response to macronutrients (Fig. 13), as reported by Njoroge et al., (2017) in a similar design, these findings support the use of other soil quality indicators (Vanlauwe & Giller, 2006) as experienced in Falconnier et al., (2016).

Moreover, a precise understanding of the local variation was not possible in our case, considering the size of the area and the number of data points per district (about 10 for 1000 km^2). Indeed, assessing local variation would requires a higher sampling density. Although, the district scale in our case, was the best intermediate scale between the farm and the regional scale. The community/village scale has been reported as a relevant for the implementation of agricultural technologies (Giller et al., 2011; Vanlauwe et al., 2015), however, was not assessed in this study.

5.2.2. Potential profitability of fertilizer in the study sites

For the quantity of N fertilizer applied, which was between 100 and 140 kg ha^{-1}, the potential profitability of this application is low. Indeed, the median agronomic efficiency of N was about 8 kg kg^{-1}. This value is comparable with the results of a national survey in Nigeria (Liverpool-Tasie et al., 2017). Under farmer's management, with a mean N application rate of 70 kg ha^{-1}, gain from application was about 7.6 kg kg^{-1}. In contrast, a metanalysis of Vanlauwe et al., (2011), average fertilizer agronomic efficiency was about 19 and 23 kg kg^{-1} for farmer-lead and researcher-led management respectively. In latter study, maximum ANE was about 30 kg kg^{-1} for N application rates higher than 100 kg ha^{-1}. Low agronomic efficiency was found to P (9 kg kg^{-1}) and differences were observed between the two zones. Under researcher-led management, values ranging between 29 and 67 kg kg^{-1} in acidic sandy soils, were reported by Kurwakumire et al., (2014) for an application of 40 kg ha^{-1}. These values are far above the APE of this study, even for the Southern Highlands that show similar environmental conditions, and higher APE.

These values are alarming when considering that plots were under researcher-led management. Moreover, they indicate that such amounts of fertilizer may not be profitable regarding the generally low and variable return from the application. This supports the need of testing alternative fertilization rates in association with additional inputs to increase the agronomic efficiency fertilizer such as lime, manure (Bielders & Gérard, 2015; Kihara et al., 2016) or other organic input (Vanlauwe et al., 2011). However, a further identification of the yield response limiting factors is necessary before testing additional inputs. Testing the influence of historical soil management practices would then be necessary.

5.3. Methodological consideration

5.3.1. On the data collection and processing

From the 296 farms present in the dataset, only 225 had yield data. Even fewer farms had planting dates available and as a result, only 134 data points were used to model yield across the country. Moreover, planting and harvest dates were characterized by strong variation within farms while planting and harvesting of the six plots were supposed to be performed on a single day. However, detailed characteristics of the trial conditions were not available (pest, diseases occurrences and remedial, plot damages, etc.).

In order to correct for the high variation of shelling percentage of the cobs and moisture content at harvest, a farm median aggregated indicator was used. Although, a significant effect of treatment was observed for shelling percentage (Fig. S3d) in the Southern Highlands only. Thus, farm aggregated values could have slightly overestimated of the yield of these plots.

5.3.2. On the model selection and predictions

With the attempt to predict yield and response using different types of models, different outcomes were observed between the different models. Our main purpose was to ensure independence of the model residuals across the study space. Indeed, violation of independence is assumed to create a bias in the model estimates (Zuur et al. 2009). Geostatistical models were dropped from the analysis because of the erratic empirical variogram preventing to fit the data into a variogram model. The highly heterogenous spatial structure of the data points and their relatively low density was likely to cause the erratic variogram obtained (Armstrong, 1984; Yemefack et al., 2005). Moreover, fitting a single variogram for data points situated in very different conditions might be intuitively wrong. It is likely that yields may have a very different spatial structure from one region to another depending on the heterogeneity of the growing conditions within this region. To our knowledge, spatial analysis of yield using geostatistics have been commonly performed at the field scale (Colonna et al., 2004; Kravchenko, 1998; Lambert et al., 2004). It is questionable this method may be used in our context.

Different model estimates and associated standard errors were not statistically tested in this study. Colonna et al., (2004) observed a consistency of the estimates after accounting for spatial dependence. In our study, important change in coefficients were found when comparing linear regression (with fixed effect only) to mixed models and spatial error model. However, we did not observe large differences between mixed models and spatial error model coefficients. Taking the district grouping factors as random effects with a distance criterion of 10 km for the spatial weight matrix, is already accounting for spatial correlation at this scale.

Large differences between model predictions and observed values for the farm level cross validations were observed. The training sample being the same and the estimated coefficient similar (for all data) the different prediction procedure, inherent to the model, were assumed to cause these differences. For the mixed models, in the case of a random intercept case, fitted values for an observation are estimated in the following way: $X_i\beta + b_i$ (where b is the coefficient of the random factors at level i). The fixed value is then corrected by the random effect coefficient. When all random effects are included in the model (all data and farm CV) the fitted values are conditional on all the modes of the random effects (Bates et al., 2015). However, for the district cross validation, values were predicted on unknown random effects. Hence, the predictions were made at the population level where all random effects are set to 0 (Bates et al. 205). Regarding the spatial error model, the predicted value is decomposed into trend, signal and noise (Bivand, 2002). In the case of the error model, the signal being set to 0 the trend (predictive value) takes the form $X\beta +(I- \lambda W)^{-1}\varepsilon$. Finally, one of the main shortcomings of our model testing procedure was replicability. For the fam and district cross validation, only one round of cross validation was performed, implicating a single random splitting of the data. It is reasonable to assumed that different sets of observations for training and testing might have given different estimates and then different prediction performances. It is then likely that out estimates are biased because not representative of the population estimates. Looking at the random forest results, where data are resampled multiple times to make predictions, the proportion of variation explained is already lower compared to the mixed model when testing on the entire dataset. The inherent resampling procedure of this method is assumed to give more stable predictions compared to a model where only a single sample is used to predict.

6. Conclusion

Understanding the scales of fertilizer response variation is a necessary step prior to formulating recommendations. The variability of yield nutrient responses is dominated by small-scale variation. To a lower extent, also large-scale variation of yield responses exists for nitrogen and phosphorus. Yield response variability between farms was very high for every nutrient, implying that mineral fertilization would not be beneficial for a large proportion of the farmers in the trials. Hence, fertilizer as a sole option to increase maize productivity is not sufficient and needs to be associated with better soil management practices.

Variables representative of the biophysical conditions could explain a reasonable part of the variation of yield in the control plot. However, these variables were associated with large-scale factors and explained very little of the variation of the macronutrient responses. The latter are likely to vary according to small-scale factors, such as historical land use and soil fertility management. Model prediction performances, based on these only factors, were very low and therefore deemed not to be good indicators to help the formulation of fertilizer recommendations.

To advise farmers in the use of fertilizers, our study suggests that a better understanding of the yield response constraints is necessary. This requires downscaling the level of analysis, with multiple sites at the landscape/community scale, within similar agroclimatic conditions. The study of the combined effects of small-scale biophysical factors with historical management is needed for the identification of yield response constraints and their scale of influence. Finally, temporal stability of yield response was not assessed in this study but remains an essential component of yield variation, that needs further understanding in the context of maize cultivation in sub-Saharan Africa.

References

Articles

Alley, M. M., & Vanlauwe, B. (2009). *The Role of Fertilizers in Integrated Plant Nutrient Management. IFA- International Fertilizer Industry Association.* .

Aref, F. (2011). Zinc and Boron Content by Maize Leaves from Soil and Foliar Application of Zinc Sulfate and Boric Acid in Zinc and Boron Deficient Soils. *Middle-East Journal of Scientific Research*, *7*(4), 610–618.

Armstrong, M. (1984). Common problems seen in variograms. *Journal of the International Association for Mathematical Geology*, *16*(3), 305–313. https://doi.org/10.1007/BF01032694

Baligar, V. C., Fageria, N. K., & He, Z. L. (2001). Nutrient Use Efficiency in Plants. *Communications in Soil Science and Plant Analysis*, *32*(7–8), 921–950. https://doi.org/10.1081/CSS-100104098

Bationo, A., Hartemink, A., Lungu, O., Naimi, M., Okoth, P., Smaling, E., … Waswa, B. (2012). Knowing the African Soils to Improve Fertilizer Recommendations. https://doi.org/10.1007/978-94-007-2960-5

Beyer, M., Wallner, M., Bahlmann, L., Thiemig, V., Dietrich, J., & Billib, M. (2016). Rainfall characteristics and their implications for rain-fed agriculture: a case study in the Upper Zambezi River Basin. *Hydrological Sciences Journal*, *61*(2), 321–343. https://doi.org/10.1080/02626667.2014.983519

Bielders, C. L., & Gérard, B. (2015). Millet response to microdose fertilization in south-western Niger: Effect of antecedent fertility management and environmental factors. *Field Crops Research*, *171*, 165–175. https://doi.org/10.1016/j.fcr.2014.10.008

Bivand, R. S. (2002). Spatial Econometrics Functions in \proglang{R}: Classes and Methods. *Journal of Geographical System*, *4*(August 2002), 405–421.

Blaes, X., Chomé, G., Lambert, M. J., Traoré, P. S., Schut, A. G. T., & Defourny, P. (2016). Quantifying fertilizer application response variability with VHR satellite NDVI time series in a rainfed smallholder cropping system of Mali. *Remote Sensing*, *8*(6). https://doi.org/10.3390/rs8060531

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brink, A. B., & Eva, H. D. (2009). Monitoring 25 years of land cover change dynamics in Africa: A sample based remote sensing approach. *Applied Geography*, *29*(4), 501–512. https://doi.org/10.1016/j.apgeog.2008.10.004

Buerkert, A., Bationo, A., & Piepho, H. P. (2001). Efficient phosphorus application strategies for increased crop production in sub-Saharan West Africa. *Field Crops Research*, *72*(1), 1–15. https://doi.org/10.1016/S0378-4290(01)00166-6

Bumb, B. L., Johnson, M. E., & Fuentes, P. A. (2011). Policy Options for Improving Regional Fertilizer Markets in West Africa. *IFPRI Discussion Paper*, *No. 01084*(July).

Burke, W. J., Jayne, T. S., & Black, J. R. (2017). Factors explaining the low and variable profitability of fertilizer application to maize in Zambia. *Agricultural Economics (United Kingdom)*, *48*(1), 115–126. https://doi.org/10.1111/agec.12299

Cassman, K. G., Dobermann, A., Walters, D. T., & Yang, H. (2003). Meeting Cereal Demand While Protecting Natural Resources and Improving Environmental Quality. *Annual Review of Environment and Resources*, *28*(1), 315–358. https://doi.org/10.1146/annurev.energy.28.040202.122858

Chianu, J. N., Chianu, J. N., & Mairura, F. (2012). Mineral fertilizers in the farming systems of sub-Saharan Africa. A review. *Agronomy for Sustainable Development*, *32*(2), 545–566. https://doi.org/10.1007/s13593-011-0050-0

Colonna, I., Ruffo, M., Bollero, G., & Bullock, D. (2004). A Comparison of Geostatistical and Spatial Autoregressive Approaches for Dealing with Spatially Correlated Residuals in Regression Analysis for Precision Agriculture Applications. *Annual Conference on Applied Statistics in Agriculture*, 310–326.

Deckers, J. (2002). Management in Soilscapes of Sub-Saharan Africa, (i).

Edmonds, D. E., Abreu, S. L., West, A., Caasi, D. R., Conley, T. O., Daft, M. C., … Raun, W. R. (2009). Cereal nitrogen use efficiency in sub Saharan Africa. *Journal of Plant Nutrition*, *32*(12), 2107–2122. https://doi.org/10.1080/01904160903308184

F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., … Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, *30*(5), 609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

Falconnier, G. N., Descheemaeker, K., Mourik, T. A. V., & Giller, K. E. (2016). Unravelling the causes of variability in crop yields and treatment responses for better tailoring of options for sustainable intensification in southern Mali. *Field Crops Research*, *187*, 113–126. https://doi.org/10.1016/j.fcr.2015.12.015

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., … Michaelsen, J. (2015). The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. *Scientific Data*, *2*, 1–21. https://doi.org/10.1038/sdata.2015.66

Giller, K. E., Tittonell, P., Rufino, M. C., van Wijk, M. T., Zingore, S., Mapfumo, P., … Vanlauwe, B. (2011). Communicating complexity: Integrated assessment of trade-offs concerning soil fertility management within African farming systems to support innovation and development. *Agricultural Systems*, *104*(2), 191–203. https://doi.org/10.1016/j.agsy.2010.07.002

Grassini, P., van Bussel, L. G. J., Van Wart, J., Wolf, J., Claessens, L., Yang, H., … Cassman, K. G. (2015). How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. *Field Crops Research*, *177*, 49–63. https://doi.org/10.1016/j.fcr.2015.03.004

Grzebisz, W., Gransee, A., Szczepaniak, W., & Diatta, J. (2013). The effects of potassium fertilization on water-use efficiency in crop plants. *Journal of Plant Nutrition and Soil Science*, *176*(3), 355–374. https://doi.org/10.1002/jpln.201200287

Haileslassie, A., Priess, J., Veldkamp, E., Teketay, D., & Lesschen, J. P. (2005). Assessment of soil nutrient depletion and its spatial variability on smallholders' mixed farming systems in Ethiopia using partial versus full nutrient balances. *Agriculture, Ecosystems and Environment*, *108*(1), 1–16. https://doi.org/10.1016/j.agee.2004.12.010

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., … Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., … Tondoh, J. E. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS ONE*, *10*(6), 1–26. https://doi.org/10.1371/journal.pone.0125814

Hochman, Z., Gobbett, D., Holzworth, D., McClelland, T., van Rees, H., Marinoni, O., … Horan, H. (2013). Reprint of "Quantifying yield gaps in rainfed cropping systems: A case study of wheat in Australia." *Field Crops Research*, *143*, 65–75. https://doi.org/10.1016/j.fcr.2013.02.001

Jauregui, M. a, & Sain, G. E. (1992). *Continuous economic analysis of crop response to fertilizer in on-farm research*.

Jayne, T. S., & Rashid, S. (2013). Input subsidy programs in sub-Saharan Africa: A synthesis of recent evidence. *Agricultural Economics (United Kingdom)*, *44*(6), 547–562. https://doi.org/10.1111/agec.12073

Kihara, J., Huising, J., Nziguheba, G., Waswa, B. S., Njoroge, S., Kabambe, V., … Coulibaly, A. (2016). Maize response to macronutrients and potential for profitability in sub-Saharan Africa. *Nutrient Cycling in Agroecosystems*, *105*(3), 171–181. https://doi.org/10.1007/s10705-015-9717-2

Kihara, J., Nziguheba, G., Zingore, S., Coulibaly, A., Esilaba, A., Kabambe, V., … Huising, J. (2016). Understanding variability in crop response to fertilizer and amendments in sub-Saharan Africa. *Agriculture, Ecosystems and Environment*, *229*, 1–12. https://doi.org/10.1016/j.agee.2016.05.012

Kihara, J., Sileshi, G. W., Nziguheba, G., Kinyua, M., Zingore, S., & Sommer, R. (2017). Application of secondary nutrients and micronutrients increases crop yields in sub-Saharan Africa. *Agronomy for Sustainable Development*, *37*(4). https://doi.org/10.1007/s13593-017-0431-0

Kravchenko, A. (1998). Spatial variability:Management, Topographical, and Weather Effects on Spatial Variability of Crop Grain Yields. *Environmental Soil Physics*, *47*, 655–675. https://doi.org/10.1016/B978-012348525-0/50026-4

Kurwakumire, N., Chikowo, R., Mtambanengwe, F., Mapfumo, P., Snapp, S., Johnston, A., & Zingore, S. (2014). Maize productivity and nutrient and water use efficiencies across soil fertility domains on smallholder farms in Zimbabwe. *Field Crops Research*, *164*(1), 136–147. https://doi.org/10.1016/j.fcr.2014.05.013

Lambert, D. M., Lowenberg-Deboer, J., & Bongiovanni, R. (2004). A comparison of four spatial regression models for yield monitor data: A case study from Argentina. *Precision Agriculture*, *5*(6), 579–600. https://doi.org/10.1007/s11119-004-6344-3

Landau, S., Mitchell, R. A. C., Barnett, V., Colls, J. J., Craigon, J., & Payne, R. W. (2000). A parsimonious, multiple-regression model of wheat yield response to environment. *Agricultural and Forest Meteorology*, *101*(2–3), 151–166. https://doi.org/10.1016/S0168-1923(99)00166-5

Leenaars, J. G. B., Claessens, L., Heuvelink, G. B. M., Hengl, T., Ruiperez González, M., van Bussel, L. G. J., … Cassman, K. G. (2018). Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa. *Geoderma*, *324*(February), 18–36. https://doi.org/10.1016/j.geoderma.2018.02.046

Liverpool-Tasie, L. S. O., Omonona, B. T., Sanou, A., & Ogunleye, W. O. (2017). Is increasing inorganic fertilizer use for maize production in SSA a profitable proposition? Evidence from Nigeria. *Food Policy*, *67*, 41–51. https://doi.org/10.1016/j.foodpol.2016.09.011

Lobell, D. B., Cassman, K. G., & Field, C. B. (2009). Crop Yield Gaps: Their Importance, Magnitudes, and Causes. *Annual Review of Environment and Resources*, *34*(1), 179–204. https://doi.org/10.1146/annurev.environ.041008.093740

Magehema, A., Chang, L., & Mkoma, S. (2014). Implication of rainfall variability on maize production in Morogoro, Tanzania. *International Journal of Environmental Sciences*, *4*(5), 1077–1086. https://doi.org/10.6088/ijes.2014040404547

Mather, D., Waized, B., Ndyetabula, D., Temu, A., & Minde, I. (2016). The profitability of inorganic fertilizer use in smallholder maize production in Tanzania : Implications for alternative strategies to improve smallholder maize productivity, 1–45.

Minot, N., & Benson, T. (2009). Fertilizer subsidies in Africa: Are vouchers the answer?, (July), IFPRI Issue Brief 60. Retrieved from http://ideas.repec.org/p/fpr/issbrf/60.html

Musinguzi, P., Tenywa, J. S., Ebanyat, P., Tenywa, M. M., Mubiru, D. N., Basamba, T. A., & Leip, A. (2013). Soil Organic Carbon Thresholds and Nitrogen Management in Tropical Agroecosystems: Concepts and Prospects. *Journal of Sustainable Development*, *6*(12). https://doi.org/10.5539/jsd.v6n12p31

Muthoni, F. K., Baijukya, F., Bekunda, M., Sseguya, H., Kimaro, A., Alabi, T., … Hoeschle-Zeledon, I. (2017). Accounting for correlation among environmental covariates improves delineation of extrapolation suitability index for agronomic technological packages. *Geocarto International*, *6049*, 1–23. https://doi.org/10.1080/10106049.2017.1404144

Njoroge, S., Schut, A. G. T., Giller, K. E., & Zingore, S. (2017). Strong spatial-temporal patterns in maize yield response to nutrient additions in African smallholder farms. *Field Crops Research*, *214*(May), 321–330. https://doi.org/10.1016/j.fcr.2017.09.026

Raun, W. R., Barreto, H. J., Westerman, R. L., & Raun W.R., Barreto H.J., W. R. L. (1993). Use of Stability Analysis for long -Term Soil Fertility Experiments. *Agronomy Journal*.

Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C., & Foley, J. A. (2012). Recent patterns of crop yield growth and stagnation. *Nature Communications*, *3*, 1293–1297. https://doi.org/10.1038/ncomms2296

Ronner, E., Franke, A. C., Vanlauwe, B., Dianda, M., Edeh, E., Ukem, B., … Giller, K. E. (2016). Understanding variability in soybean yield and response to P-fertilizer and rhizobium inoculants on farmers' fields in northern Nigeria. *Field Crops Research*, *186*, 133–145. https://doi.org/10.1016/j.fcr.2015.10.023

Rowhani, P., Lobell, D. B., Linderman, M., & Ramankutty, N. (2011). Climate variability and crop production in Tanzania. *Agricultural and Forest Meteorology*, *151*(4), 449–460. https://doi.org/10.1016/j.agrformet.2010.12.002

Sanchez, P. a. (2002). Ecology. Soil fertility and hunger in Africa. *Science (New York, N.Y.)*, *295*(5562), 2019–2020. https://doi.org/10.1126/science.1065256

Selassie, Y. G. (2016). Response and economic feasibility of maize (Zea mays L.) to P fertilization in acidic Alfisols of North-western Ethiopia. *Environmental Systems Research*, *5*(1), 3. https://doi.org/10.1186/s40068-016-0056-3

Sheahan, M., & Barrett, C. B. (2017). Ten striking facts about agricultural input use in Sub-Saharan Africa. *Food Policy*, *67*, 12–25. https://doi.org/10.1016/j.foodpol.2016.09.010

Shehu, B., Merckx, R., Jibrin, J., Kamara, A., & Rurinda, J. (2018). Quantifying Variability in Maize Yield Response to Nutrient Applications in the Northern Nigerian Savanna. *Agronomy*, *8*(2), 18. https://doi.org/10.3390/agronomy8020018

Smaling, E. M. A., Nandwa, S. M., & Janssen, B. H. (1997). Soil fertility in Africa is at sake. *Replenishing Soil Fertility in Africa.*, *SSSA Spec.*(51), 47–62.

Smaling, E. M. A., Stoorvogel, J. J., & Windmeijer, P. N. (1993). Calculating soil nutrient balances in Africa at different scales - II. District scale. *Fertilizer Research*, *35*(3), 237–250. https://doi.org/10.1007/BF00750642

Stocking, M. A. (2003). Tropical Soils and Food Security: The Next 50 Years. *Science*, *302*(5649), 1356–1359. https://doi.org/DOI 10.1126/science.1088579

Tesfaye, K., Gbegbelegbe, S., Cairns, J. E., Shiferaw, B., Prasanna, B. M., Sonder, K., … Robertson, R. (2015). Maize systems under climate change in sub-Saharan Africa. *International Journal of Climate Change Strategies and Management*, *7*(3), 247–271. https://doi.org/10.1108/IJCCSM-01-2014-0005

Tittonell, P., & Giller, K. E. (2013). When yield gaps are poverty traps: The paradigm of ecological intensification in African smallholder agriculture. *Field Crops Research*, *143*, 76–90. https://doi.org/10.1016/j.fcr.2012.10.007

Tittonell, P., Muriuki, A., Klapwijk, C. J., Shepherd, K. D., Coe, R., & Vanlauwe, B. (2013). Soil Heterogeneity and Soil Fertility Gradients in Smallholder Farms of the East African Highlands. *Soil Science Society of America Journal*, *77*(2), 525. https://doi.org/10.2136/sssaj2012.0250

Tittonell, P., Vanlauwe, B., Corbeels, M., & Giller, K. E. (2008). Yield gaps, nutrient use efficiencies and response to fertilisers by maize across heterogeneous smallholder farms of western Kenya. *Plant and Soil*, *313*(1–2), 19–37. https://doi.org/10.1007/s11104-008-9676-3

Tittonell, P., Vanlauwe, B., de Ridder, N., & Giller, K. E. (2007). Heterogeneity of crop productivity and resource use efficiency within smallholder Kenyan farms: Soil fertility gradients or management intensity gradients? *Agricultural Systems*, *94*(2), 376–390. https://doi.org/10.1016/j.agsy.2006.10.012

Tittonell, P., Vanlauwe, B., Leffelaar, P. A., Rowe, E. C., & Giller, K. E. (2005). Exploring diversity in soil fertility management of smallholder farms in western Kenya: I. Heterogeneity at region and farm scale. *Agriculture, Ecosystems and Environment*, *110*(3–4), 149–165. https://doi.org/10.1016/j.agee.2005.04.001

Tittonell, P., Zingore, S., van Wijk, M. T., Corbeels, M., & Giller, K. E. (2007). Nutrient use efficiencies and crop responses to N, P and manure applications in Zimbabwean soils: Exploring management strategies across soil fertility gradients. *Field Crops Research*, *100*(2–3), 348–368. https://doi.org/10.1016/j.fcr.2006.09.003

Tully, K., Sullivan, C., Weil, R., & Sanchez, P. (2015). The State of soil degradation in sub-Saharan Africa: Baselines, trajectories, and solutions. *Sustainability (Switzerland)*, *7*(6), 6523–6552.

https://doi.org/10.3390/su7066523

Vågen, T.-G., Winowiecki, L. A., Tondoh, J. E., Desta, L. T., & Gumbricht, T. (2016). Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. *Geoderma*, *263*, 216–225. https://doi.org/10.1016/j.geoderma.2015.06.023

Vågen, T. G., Winowiecki, L. A., Abegaz, A., & Hadgu, K. M. (2013). Landsat-based approaches for mapping of land degradation prevalence and soil functional properties in Ethiopia. *Remote Sensing of Environment*, *134*, 266–275. https://doi.org/10.1016/j.rse.2013.03.006

van Biljon, J. J., Fouche, D., & Botha, A. D. P. (2004). Threshold values for sulphur in soils of the main maize-producing areas of South Africa. *South African Journal of Plant and Soil*, *21*(3), 152–156. https://doi.org/10.1080/02571862.2004.10635041

van Biljon, J. J., Fouche, D. S., & Botha, A. D. P. (2008). Threshold values and sufficiency levels for potassium in maize producing sandy soils of South Africa. *South African Journal of Plant and Soil*, *25*(2), 65–70. https://doi.org/10.1080/02571862.2008.10639897

van Heerwaarden, J., Baijukya, F., Kyei-Boahen, S., Adjei-Nsiah, S., Ebanyat, P., Kamai, N., … Giller, K. (2017). Soyabean response to rhizobium inoculation across sub-Saharan Africa: Patterns of variation and the role of promiscuity. *Agriculture, Ecosystems & Environment*, (August), 0–1. https://doi.org/10.1016/j.agee.2017.08.016

Van Ittersum, M. K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., & Hochman, Z. (2013). Yield gap analysis with local to global relevance-A review. *Field Crops Research*, *143*, 4–17. https://doi.org/10.1016/j.fcr.2012.09.009

van Ittersum, M. K., van Bussel, L. G. J., Wolf, J., Grassini, P., van Wart, J., Guilpart, N., … Cassman, K. G. (2016). Can sub-Saharan Africa feed itself? *Proceedings of the National Academy of Sciences*, *113*(52), 14964–14969. https://doi.org/10.1073/pnas.1610359113

Vanlauwe, B., Coe, R., & Giller, K. E. (2016). Beyond Averages: New Approaches To Understand Heterogeneity and Risk of Technology Success or Failure in Smallholder Farming. *Experimental Agriculture*, 1–23. https://doi.org/10.1017/S0014479716000193

Vanlauwe, B., Coyne, D., Gockowski, J., Hauser, S., Huising, J., Masso, C., … Van Asten, P. (2014). Sustainable intensification and the African smallholder farmer. *Current Opinion in Environmental Sustainability*, *8*, 15–22. https://doi.org/10.1016/j.cosust.2014.06.001

Vanlauwe, B., Descheemaeker, K., Giller, K. E., Huising, J., Merckx, R., Nziguheba, G., … Zingore, S. (2015). Integrated soil fertility management in sub-Saharan Africa: unravelling local adaptation. *Soil*, *1*(1), 491–508. https://doi.org/10.5194/soil-1-491-2015

Vanlauwe, B., & Giller, K. E. (2006). Popular myths around soil fertility management in sub-Saharan Africa. *Agriculture, Ecosystems and Environment*, *116*(1–2), 34–46. https://doi.org/10.1016/j.agee.2006.03.016

Vanlauwe, B., Kihara, J., Chivenge, P., Pypers, P., Coe, R., & Six, J. (2011). Agronomic use efficiency of N fertilizer in maize-based systems in sub-Saharan Africa within the context of integrated soil fertility management. *Plant and Soil*, *339*(1), 35–50. https://doi.org/10.1007/s11104-010-0462-7

Vanlauwe, B., Tittonell, P., & Mukalama, J. (2006). Within-farm soil fertility gradients affect response of maize to fertiliser application in western Kenya. *Nutrient Cycling in Agroecosystems*, *76*(2–3), 171–182. https://doi.org/10.1007/s10705-005-8314-1

Voortman, R. L., Sonneveld, B. G. J. S., & Keyzer, M. A. (2003). African land ecology: opportunities and constraints for agricultural development. *Ambio*, *32*(5), 367–373. https://doi.org/10.1579/0044-7447-32.5.367

Waddington, S. R., H.K.Murwira, Kumwenda, J. D. T., Hikwa, D., & Tagwira, F. (1998). *Soil fertility research for maize-based farming systems in Malawi and Zimbabwe*. *Proceedings of the Soil Fertility Network Results and Planning Workshop, Mutare, Zimbabwe, 7-11 July 1997*. Retrieved from http://repository.cimmyt.org/xmlui/handle/10883/539#

Waha, K., Huth, N., Carberry, P., & Wang, E. (2015). How model and input uncertainty impact maize yield simulations in West Africa. *Environmental Research Letters*, *10*(2), 024017.

https://doi.org/10.1088/1748-9326/10/2/024017

Xu, Z., Guan, Z., Jayne, T. S., & Black, R. (2009). Factors influencing the profitability of fertilizer use on maize in Zambia. *Agricultural Economics*, *40*(4), 437–446. https://doi.org/10.1111/j.1574-0862.2009.00384.x

Yemefack, M., Rossiter, D. G., & Njomgang, R. (2005). Multi-scale characterization of soil variability within an agricultural landscape mosaic system in southern Cameroon. *Geoderma*, *125*(1–2), 117–143. https://doi.org/10.1016/j.geoderma.2004.07.007

Zingore, S., Murwira, H. K., Delve, R. J., & Giller, K. E. (2007). Soil type, management history and current resource allocation: Three dimensions regulating variability in crop productivity on African smallholder farms. *Field Crops Research*, *101*(3), 296–305. https://doi.org/10.1016/j.fcr.2006.12.006

Books, reports and websites

AGRA, (2014). The Africa Agriculture Status Report 2014: Climate Change and Smallholder Agriculture in Sub-Saharan Africa. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), Copenhagen, Denmark

Anselin & Bera, (1998). Spatial Dependence in linear Regression Models with an Introduction to Spatial Econometrics. In: *Handbook of Applied Economic* A. Ullah and D.E.A. Giles, Eds., Marcel Dekker, NY. (1998), pp. 237-289.

Arshad, M.A., Lowery, B. and Grossman. B. (1996) Physical Tests for Monitoring Soil Quality. In: Doran, J.W. and Jones, A.J., Eds., Methods for Assessing Soil Quality, Soil Science Society of America Special Publication 49, SSSA, Madison, WI, 123-142.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. Journal of Statistical Software, 67(1). doi:10.18637/jss.v067.i01

Chen J., Ban Y., Li S. China: Open access to Earth land-cover map[J]. Nature, 2014, 514(7523): 434-434. DOI:10.1038/514434c.

De Pauw, (1984). Soils, physiography and agro-ecological zones of Tanzania. Crop monitoring and early warning systems project GCS/URT/047. NET. Ministry of Agriculture, Dar Es Salaam. Food and Agriculture Organization of the United Nations, Rome

FAO, (2017). FAO Statistical Databases. FAO, Available at: ⟨http://faostat.fao.org/⟩

Fox, G. A., In Negrete-Yankelevich, S., & In Sosa, V. J. (2015). Ecological statistics: Contemporary theory and application.1st Edition

FURP, (1994). Fertilizer Use Recommendations. Fertilizer Use Recommendation Project. Kenya Agricultural Research Institute, National Agricultural Research Laboratories, Nairobi, Kenya.

IFDC, (2006). Africa Fertilizer Summit Proceedings, Abuja, Nigeria June 9–13 2006, IFDC, Florence, Alabama

Pinheiro, J. C., and D. M. Bates. (2000). Mixed-Effects Models in S and S-PLUS. New York: Springer.

Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2018). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-137, https://CRAN.R-project.org/package=nlme.

Senkoro C., Ley G., Marandu A., Wortmann C., Mzimbiri M., Msaky J., Umbwe R., S D Lyimo S. (2017). Optimizing Fertilizer Use within the Context of Integrated Soil Fertility Management in Tanzania. In: OFRA: Fertilizer Use Optimization in Sub-Saharan Africa. Charles S. Wortmann and Keith Sones (eds). CAB International, Nairobi, Kenya, pp. 176–192.

Wan, Z., Hook, S., Hulley, G. (2015). MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD11A2.006

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., (2009). Mixed Effects Models and Extensions in Ecology with R (Statistics for Biology and Health). Springer.

Supplementary Material

S1. Comparison of soil measurements methods NIR-MIR vs Wet Chemistry

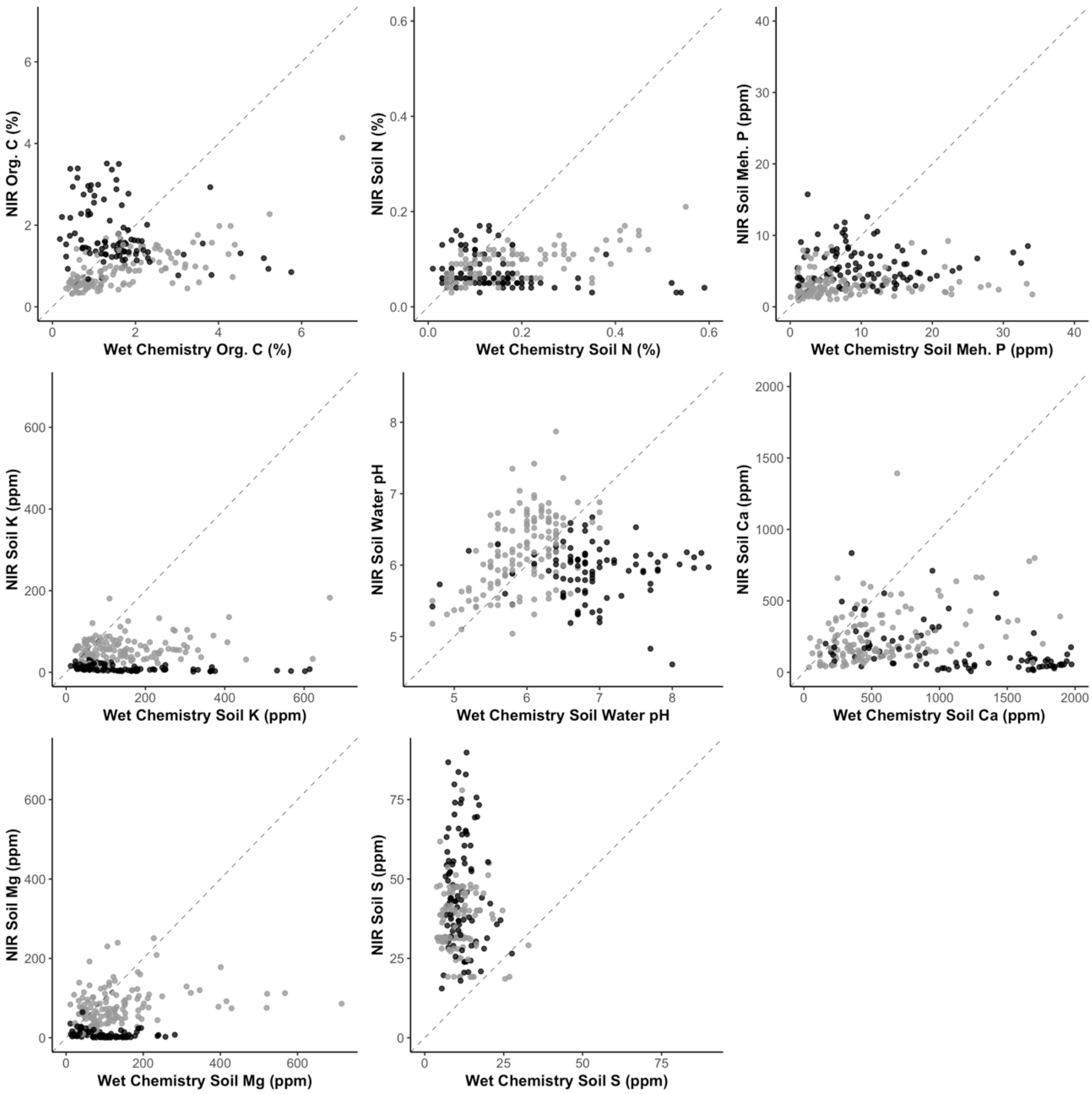

Figure S1.a. Comparison of soil properties (Org. C, N, P, K, Soil Water pH, Ca, Mg, S) measured with wet chemistry and NIR-MIR. Black and grey colours stand for the Northern zone and the Southern highlands zone respectively.

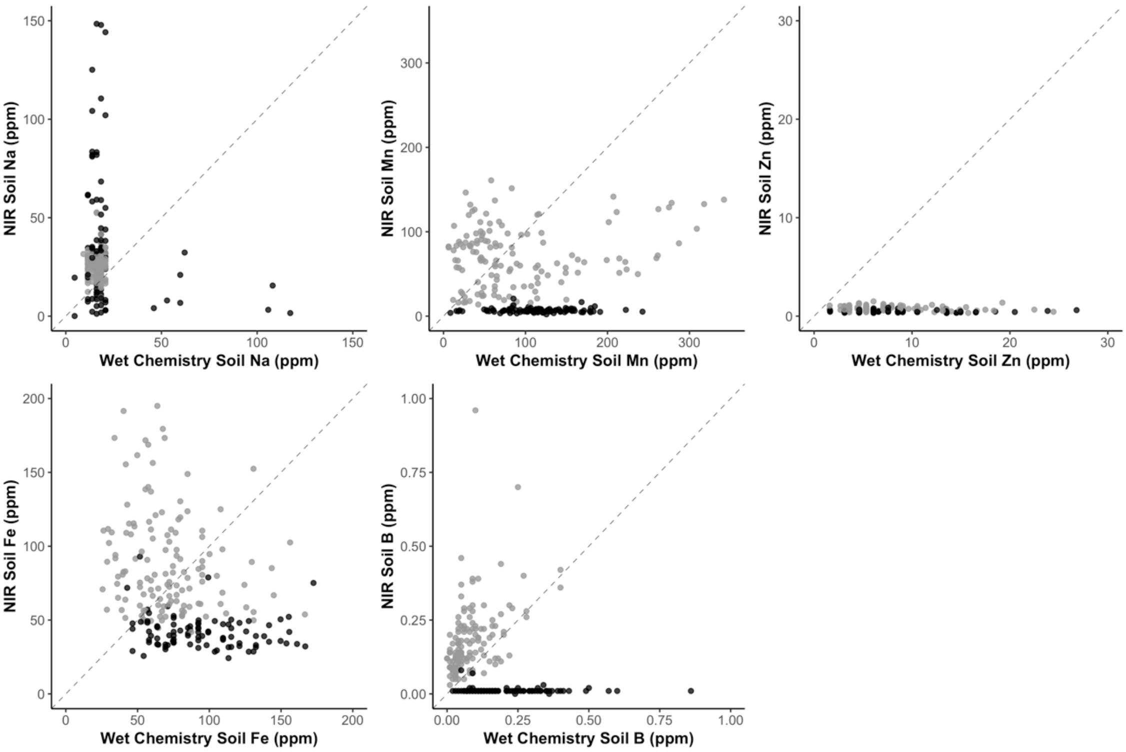

Figure S1.b. Comparison of soil properties (Na, Mn, Zn, Fe, B)) measured with wet chemistry and NIR-MIR. Black and grey colours stand for the Northern zone and the Southern highlands zone respectively.

S2. Homogenization of farm clusters size based on geographic distance

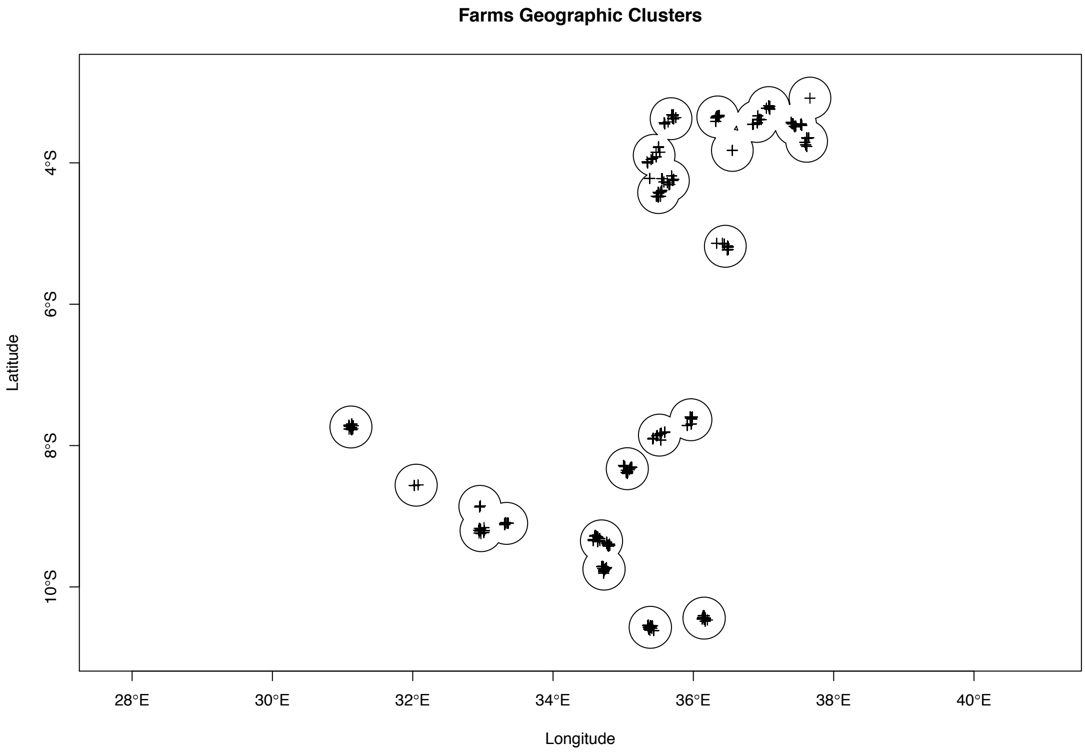

Figure S2. Result of the farm clustering based on geographic distance, each circle represents a cluster of farms referred as districts.

S3. Estimated means and differences for yield components

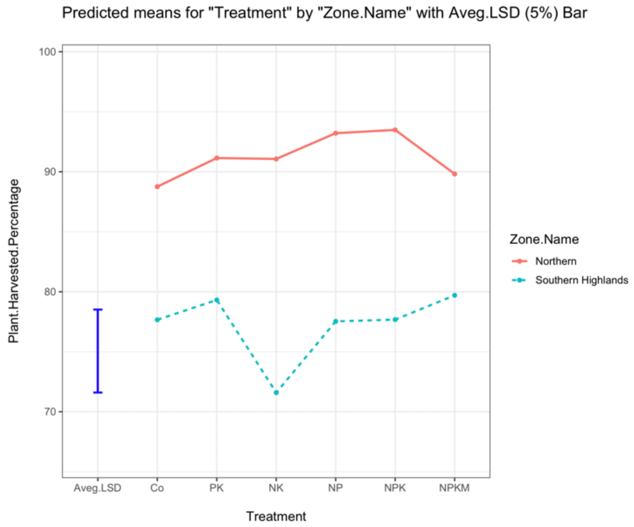

Figure S3.a. Estimated means (model 1) for the proportion of plant harvested per plot affected by zone and fertilization treatment. Avg. LSD = 6.92

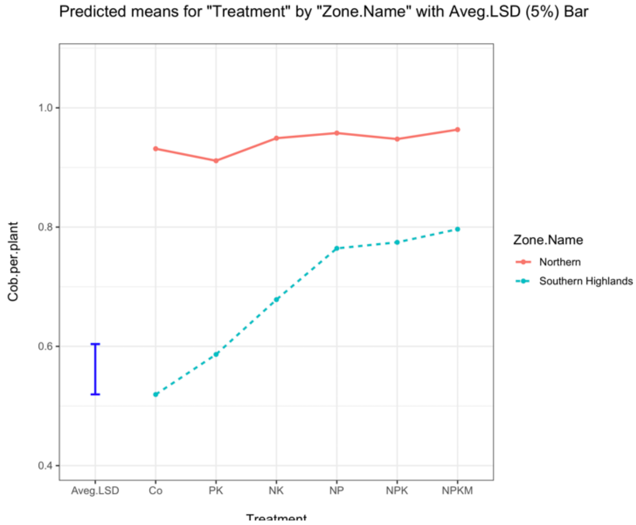

Figure S3.b. Estimated means (model 1) for the number of cobs per plant affected by zone and fertilization treatment. Avg. LSD = 0.08

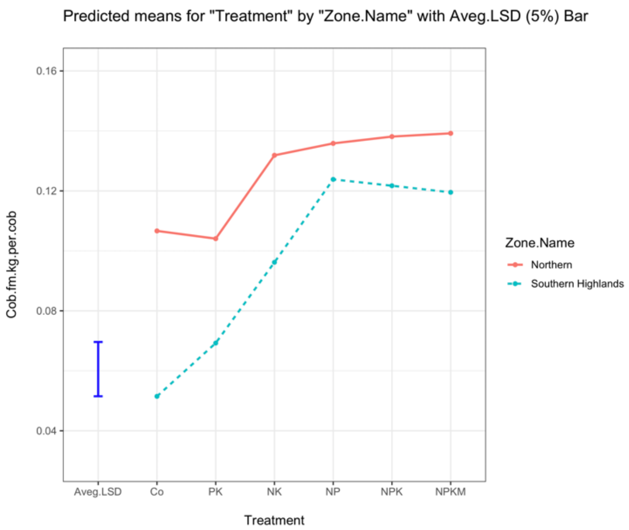

Figure S3.c. Estimated means (model 1) for the cob weight per cob affected by zone and fertilization treatment. Avg. LSD = 0.02

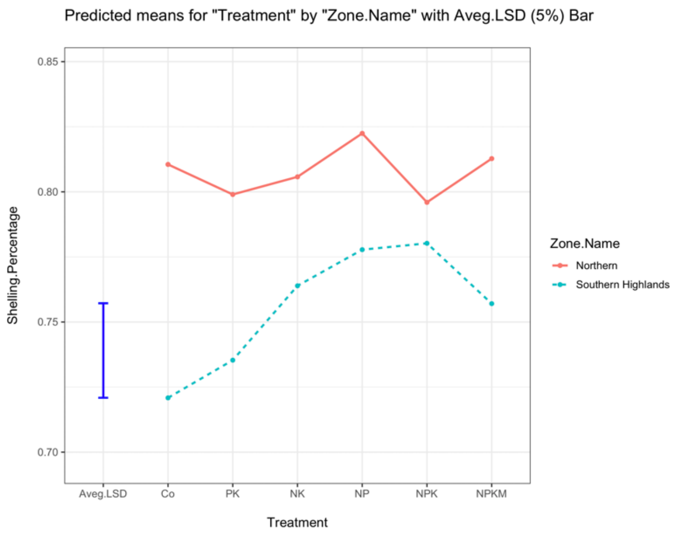

Figure S3.d. Estimated means (model 1) for the percentage of grain weight per cob affected by zone and fertilization treatment. Avg. LSD = 0.04

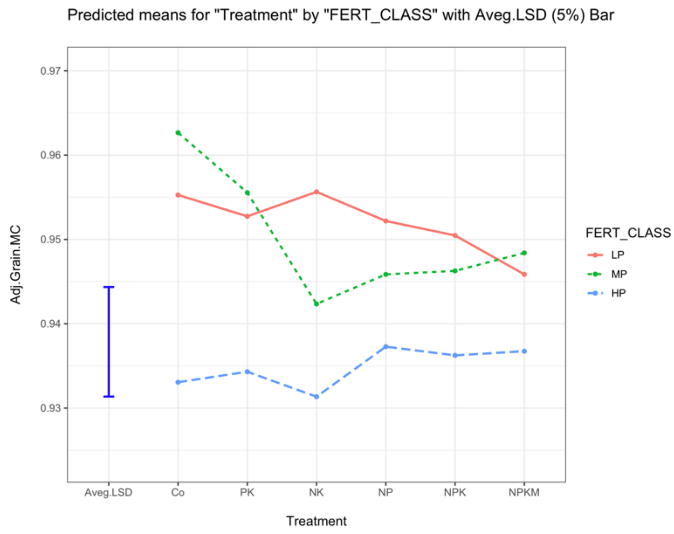

Figure S3.e. Estimated means (model 1) for the adjusted coefficient of moisture content representing the deviation from a standard moisture content of 12.5%, affected by fertilization rate and fertilization treatment. Avg. LSD = 0.04

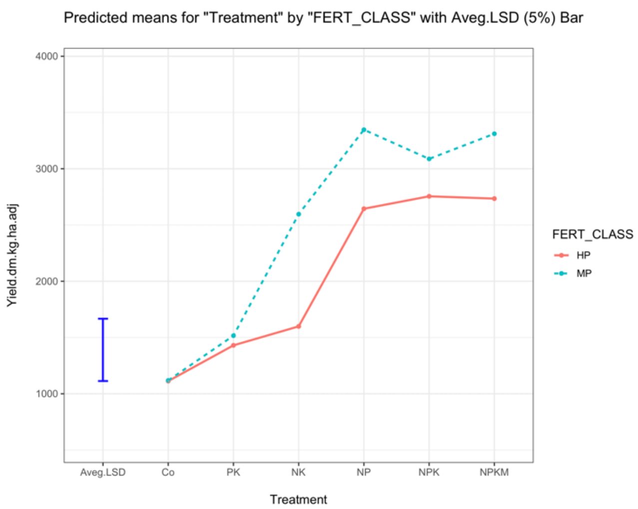

Figure S3.f. Estimated means (model 2b) for maize grain yield (kg ha^{-1}) in the Southern Highlands affected by fertilization rate and fertilization treatment, Avg. LSD = 555.25

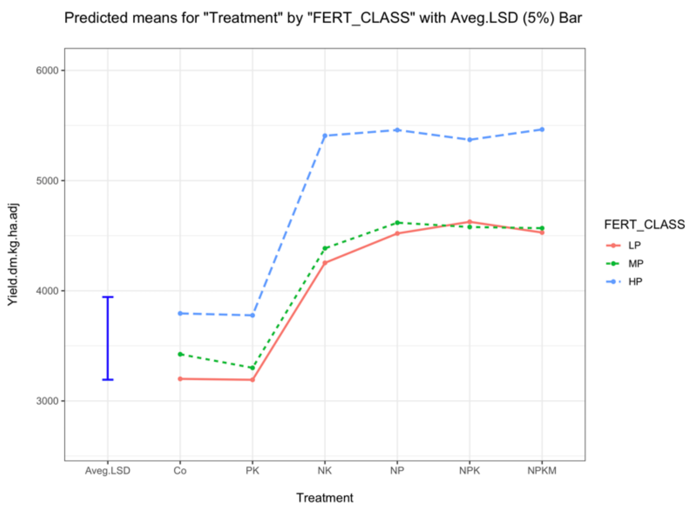

Figure S3.g. Estimated means (model 2a) for maize grain yield (kg ha^{-1}) in the Northern zone affected by fertilization rate and fertilization treatment, Avg. LSD = 555.25

S4. Treatment variations

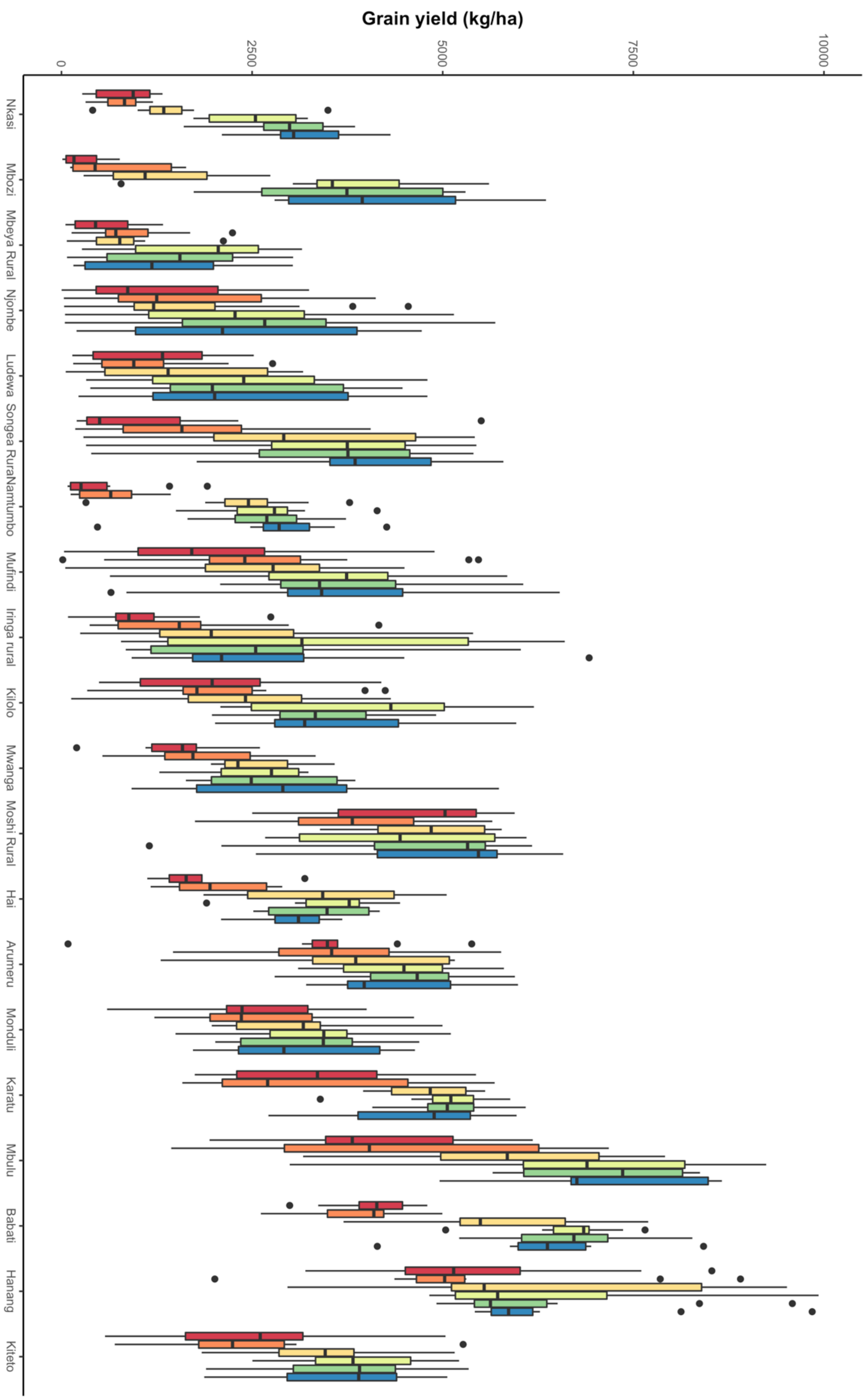

Figure S4. Maize grain yield distribution (kg.ha^{-1}) per district and treatment within district. Different colours indicates fertilization treatment, from the left to the right (Control, PK, NK, NP, NPK, NPKM).

S5. Response variations

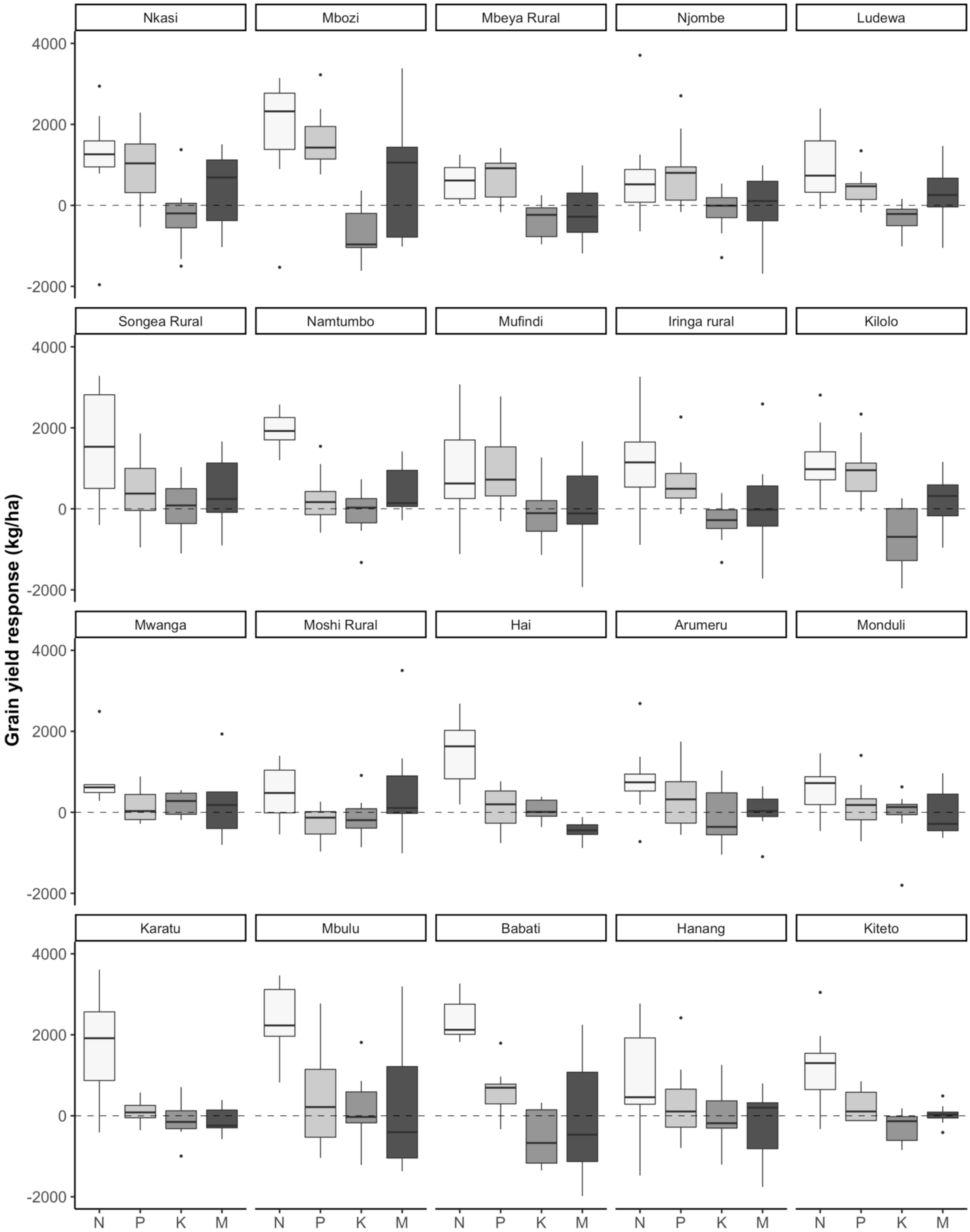

Figure S5. Maize grain yield response distribution (kg.ha^{-1}) per districts and nutrients within districts . Different colours indicate nutrient, from the left to the right (N, P, K and M).

S6. Topographic, land use, and drought characteristics of the districts

Table S6.a. No. of dry days and maximal drought period of the districts for the total growing period (GP) and during maize growing stages

District	Dry GP		Dry I		Dry II		Dry III (days)		Dro. GP		Dro. I		Dro. II.		Dro. III	
S. Highlands	1591.2	164.0	2.4	1.6	9.0	5.9	55.5	63.8	1591.2	164.0	2.4	1.6	9.0	5.9	55.5	63.8
Mbozi	**123**	1	**6**	0	**45**	4	70	1	68	1	**2**	1	**6**	1	68	1
Mbeya Rural	142	1	12	0	58	1	72	1	71	1	5	0	12	0	**71**	1
Mufindi	139	1	12	1	55	1	**72**	1	38	1	6	0	13	0	38	1
Iringa rural	135	0	13	4	58	3	65	3	65	0	9	1	13	1	64	2
Kilolo	146	0	12	1	68	0	66	0	64	0	6	0	17	0	64	0
Northern																
Mwanga	125	1	21	0	62	4	**42**	0	**26**	1	**21**	1	15	2	23	1
Moshi Rural	148	12	**24**	1	73	7	51	3	51	12	14	4	26	2	27	0
Hai	145	1	20	0	70	1	55	0	77	1	8	0	22	1	55	0
Arumeru	146	4	21	0	72	4	53	1	61	4	9	0	15	1	53	1
Monduli	148	1	20	0	77	1	52	1	**85**	1	8	0	**34**	1	52	1
Karatu	147	5	22	1	73	1	54	1	64	5	8	0	14	3	54	1
Mbulu	142	1	18	1	69	4	52	6	18	1	8	3	17	0	**18**	3
Babati	**160**	0	15	1	83	1	62	0	53	0	7	0	11	1	53	0
Hanang	157	0	12	0	**86**	1	59	0	57	0	3	0	18	0	57	0
Kiteto	139	1	11	1	78	1	50	1	34	1	5	1	**34**	4	34	0
CD (%)	4.2		27.3		12.4		12.8		28.2		20.0		28.8		29.7	

Values presented are median (left) and median absolute deviation (right) per district of each soil properties. Max and min values are highlighted in bold
I: Onset of the growing period, from 10 days before sowing to 15 days after sowing
II.: Vegetative, early reproductive and anthesis growing stages, from end of onset to two third of the total growing period
III.: Grain filling and maturation growing stages, from end of II to harvest

Table S6.b. Topographic and land use characteristics of the districts

District	Elevation (masl)		Slope (%)		Tree Cover (%)		Cropland cover (%)	
S. Highlands	1591.2	164.0	2.4	1.6	9.0	5.9	55.5	63.8
Mbozi	1604.5	84.7	4.0	3.1	9.0	4.4	**0.0**	0.0
Mbeya Rural	1855.1	11.9	2.7	1.4	**16.0**	0.0	**100.0**	0.0
Mufindi	1723.2	157.7	2.2	1.3	**16.0**	3.0	51.5	52.6
Iringa rural	1529.3	92.5	2.9	1.4	6.0	7.4	59.0	50.4
Kilolo	1380.1	42.1	2.2	1.6	3.0	3.0	78.0	32.6
Northern	1372.0	278.8	1.5	1.1	5.0	5.9	100.0	0.0
Mwanga	1058.5	155.3	**4.8**	4.2	5.0	7.4	**100.0**	0.0
Moshi Rural	**707.4**	12.3	**0.3**	0.1	5.0	3.0	**100.0**	0.0
Hai	1198.7	42.3	1.5	0.6	9.0	0.7	**100.0**	0.0
Arumeru	1025.4	23.9	1.2	0.4	5.0	4.4	**100.0**	0.0
Monduli	1420.5	36.0	2.1	1.0	5.0	1.5	**100.0**	0.0
Karatu	1440.0	38.8	2.7	2.8	4.0	2.2	**100.0**	0.0
Mbulu	**1883.1**	102.0	0.9	0.5	**0.0**	0.0	**100.0**	0.0
Babati	1369.1	40.5	2.1	1.3	6.0	0.7	**100.0**	0.0
Hanang	1560.8	7.6	1.1	0.4	1.0	1.5	**100.0**	0.0
Kiteto	1421.5	84.0	1.8	0.4	**0.0**	0.0	**100.0**	0.0
CD (%)	9.1		50.0		79.5		30.7	

Values presented are median (left) and median absolute deviation (right) per district of each soil properties. Max and min values are highlighted in bold

S7. Residual variogram for control yield and input responses

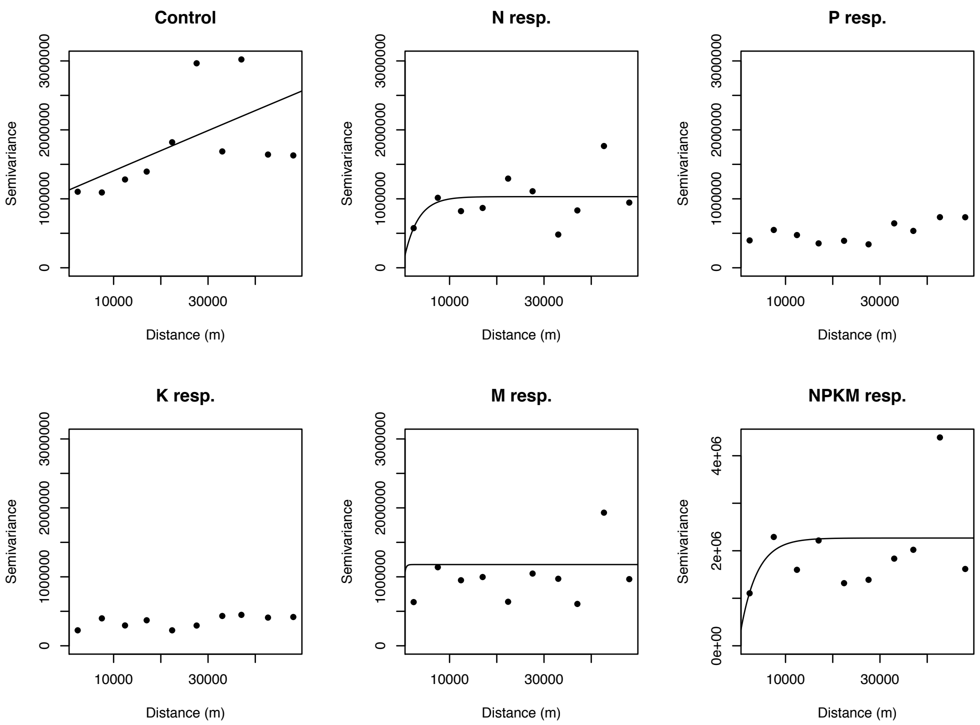

Figure S7a. Empirical and fitted model variogram of the nutrient responses on 50 km distance

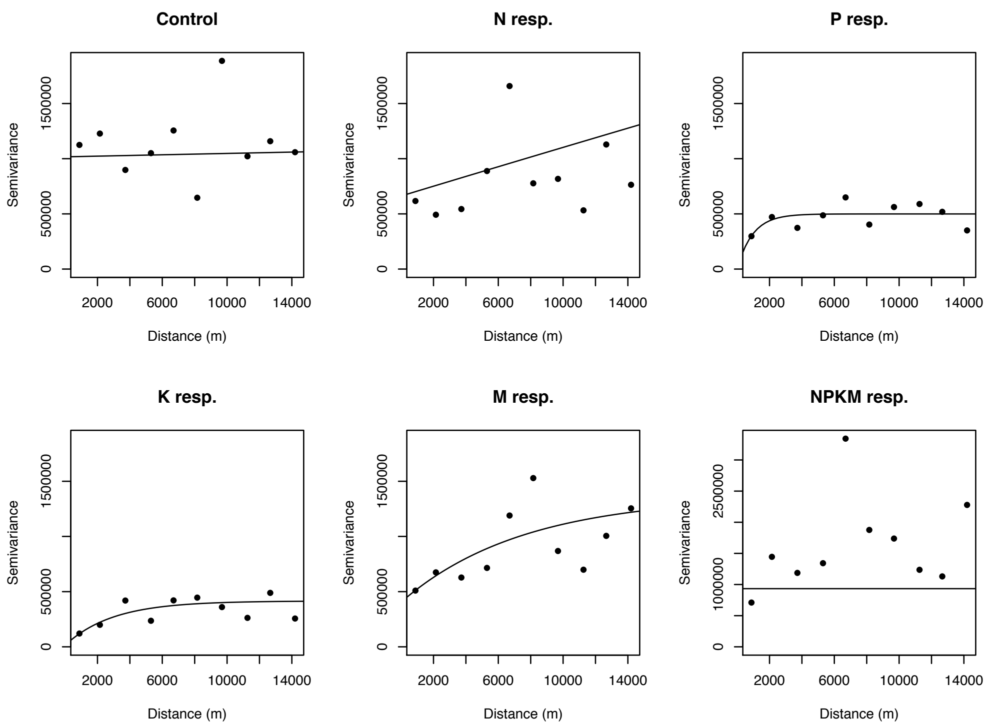

Figure S7b. Empirical and fitted model variogram of the nutrient responses on 15 km distance

S8. Input response predicted vs observed values

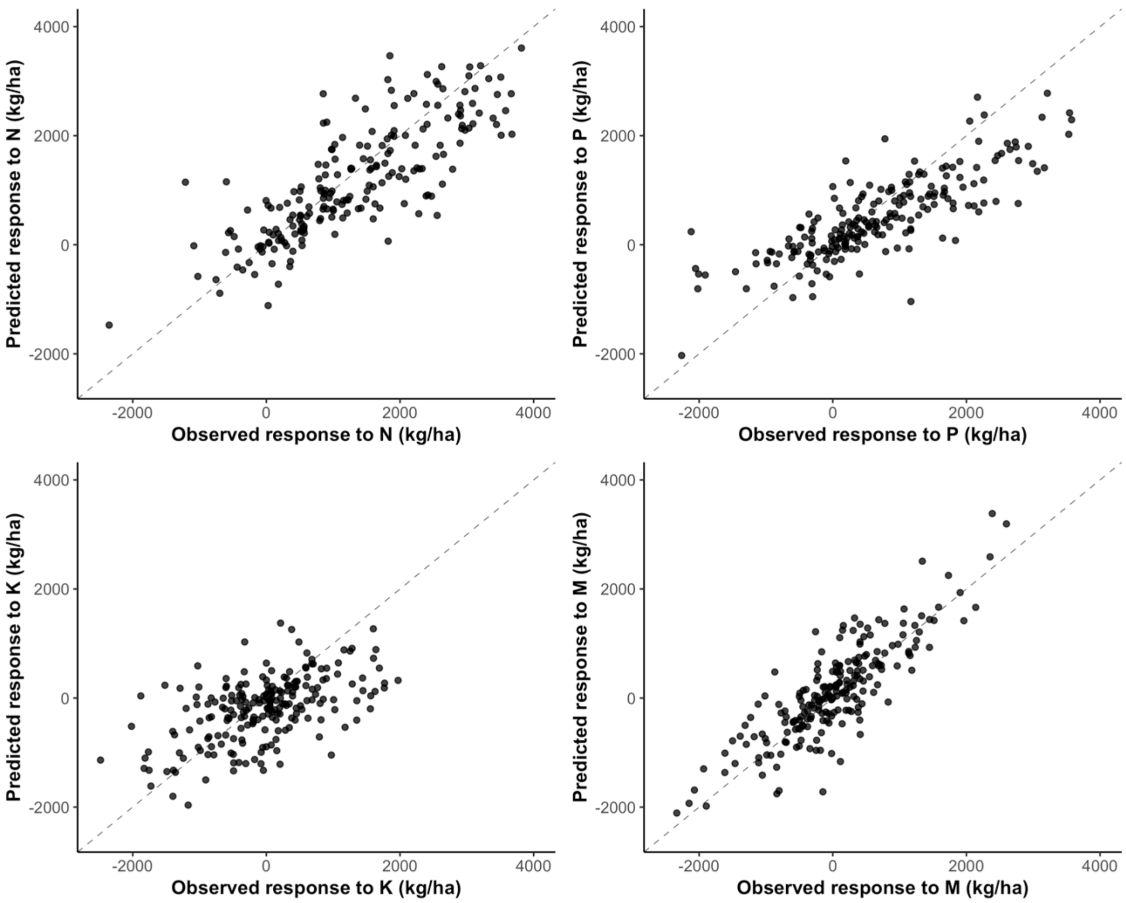

Figure S8. Model predicted (BLUE) against observed nutrient (Y NPK plot – Y Nutrient omitted plot) response