Whither Broad or Spatially Specific Fertilizer Recommendations?

A PLAN B PAPER

SUBMITTED TO THE FACULTY OF GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Maxwell Gibson Mkondiwa

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

Advisor: Philip Pardey, Co-advisor: Terrance Hurley,

Committee Member: Jeffrey Coulter

September 2015

Acknowledgements

I acknowledge the intellectual support from my advisors and committee members;

- Philip Pardey, Department of Applied Economics;
- Terrance Hurley, Department of Applied Economics;
- Jeffrey Coulter, Department of Agronomy and Plant Genetics; and
- Jason Beddow, Department of Applied Economics.

For the full financial support throughout my studies, I appreciate the generous sponsorship from the McKnight Foundation. I particularly wish to express my heartfelt gratitude to many individuals who assisted me in acquiring the funding. Some of them are listed here:

- Rebecca Nelson, Cornell University
- Richard (Ric) Coe, University of Reading and World Agroforestry Centre
- Charlie Riches, Natural Resources Institute (NRI)
- Jane Maland Cady, McKnight Foundation
- Tim Chancellor, Natural Resources Institute (NRI)

I would like also to acknowledge Todd Benson, International Food Policy Research Institute (IFPRI) for sharing the data used in this study.

Finally, I offer my sincere gratitude to my teachers, friends (both at University of Minnesota and elsewhere) and family.

Abstract

Are spatially specific agricultural input use recommendations more profitable to smallholder farmers than broad recommendations? This paper provides a theoretical and empirical modeling procedure for determining the optimal spatial scale at which agricultural researchers can make soil fertility recommendations. Theoretically, the use of Bayesian decision theory in the spatial economic optimization model allows the complete characterization of the posterior distribution functions of profits thereby taking into account spatial heterogeneity and uncertainty in the decision making process. By applying first order spatial scale stochastic dominance and Jensen's inequality; theoretically and empirically, this paper makes the case that spatially specific agricultural input use recommendations will always stochastically dominate broad recommendations for all non-decreasing profit functions ignoring the quasi-fixed cost differentials in the decision itself.

These findings are consistent with many economic studies that find precision agriculture technologies to be more profitable than conventional fertilizer (regional or national recommendations based) application approaches. The modeling approach used in this study however provides an elegant theoretical justification for such results. In addition, seasonal heterogeneity in maize responses was evident in our results. This demonstrates that broad recommendations may not only be wrong spatially but also seasonally. Further research on the empirical aspects of spatio-temporal instability of crop responses to fertilizer application using multi-location and multi-season data is needed to fully address the question posed initially. The decision making theory developed here can however be extended to incorporate spatio-temporal heterogeneity and alternative risk preferences.

Keywords: Bayesian decision theory, spatial scale stochastic dominance, spatial heterogeneity, Malawi.

Table of Contents

List of Tables

List of Figures

List of Selected Acronyms and Abbreviations

ADD	Agricultural Development Division
BLADD	Blantyre ADD
EPA	Extension Planning Area
FOSSSD	First Order Spatial Scale Stochastic Dominance
KADD	Kasungu ADD
KRADD	Karonga ADD
LADD	Lilongwe ADD
MADD	Machinga ADD
MZADD	Mzuzu ADD
N	Nitrogen
Nsq	Nitrogen Squared
RDP	Rural Development Programme
SLADD	Salima ADD
SUR	Seemingly Unrelated Regression
SVADD	Shire Valley ADD

1.0 Introduction

Soil fertility is regarded as the most limiting factor to crop productivity in Africa. In terms of inorganic fertilizer application for example, sub-Saharan Africa (SSA) lags behind with average application rates at 13 kilograms per hectare (ha) against the 94 kilograms per ha in other developing countries (Sheahan, Black and Jayne 2012). In addition, soils in SSA are losing nutrients at an alarming rate. Studies in the 1990s estimated an annual average loss of 22 kilograms of Nitrogen (N), 2.5 kilograms of Phosphorus (P) and 15 kilograms of Potassium (K) per hectare for the whole SSA. Malawi experiences similar challenges. In Malawi, it is reported that N is the most deficient nutrient in nearly all soils and P is the second most limiting nutrient especially in light soils. A nutrient depletion status study indicated that Malawi soils lose nutrients at annual rates of not less than 40 kg N per ha, 6.6 kg P per ha, and 33.2 kg K per ha; these rates are higher than average for Sub-Saharan Africa (Makumba 2003).

These twin challenges coupled with weak markets, poor infrastructure and other challenges have led to perpetual low levels of crop productivity. Given the strong linkage between soil fertility, crop productivity and food security; addressing the decline in soil fertility remains an important challenge for policy makers Ojayi *et al*.(2011). Initiatives to improve soil fertility have been at the pinnacle of all Malawi's political and economic discourses for more than half–century. At any particular time during Malawi's post-independence era (i.e. 1964 to date), there have been various large scale agricultural research efforts to improve soil fertility. These include country wide area-specific fertilizer experimental trials (1995-98) and legume trials (1998-99). In the last decade, integrated soil fertility management technologies have been touted as the solution to the soil fertility problem. In addition to inorganic fertilizers and legume integration; agricultural researchers have also recommended agro-forestry and conservation agriculture. All these recommendations however have proved futile in increasing fertilizer adoption.

According to Zingore *et al*. (2007), soil fertility varies considerably at the farm and landscape levels in many smallholder farming systems, leading to variability in crop productivity and crop response to additions of fertilizer and organic nutrient resources. Suri (2011), using observational data in Kenya, argued that heterogeneity is the reason for the empirical puzzle of low adoption of notably profitable technologies like hybrid seeds and organic soil fertility technologies. The major finding of Suri's study was that farmers with higher returns to the technology adopt it while those with a lower return do not. Another related argument by Duflo, Kremer and Robinson (2008) is that while fertilizer can be very profitable when used correctly, farmers may not use fertilizer and hybrid seeds because official recommendations are not adapted to many farmers in that particular region or country.

This study deals with this spatial aspect of heterogeneity in profits using experimental data from Malawi. The general hypothesis in this study, as other scholars (e.g. Giller *et al*., 2010; Smale, Byerlee & Jayne, 2011) have also argued, is that broad or blanket recommendations fail to account for the significant spatial variation and heterogeneity in the environmental and economic circumstances faced by farmers that do (and should) influence their crop production decisions. There is consensus among researchers that practices must be adapted to local conditions in order to account for agro climatic circumstances, population pressure, labor availability and the stage of infrastructural and institutional development (Smale, Byerlee and Jayne 2011). While this conjecture remains strong, evidence in many countries suggests that single and nationally based recommendations are still common.

Until 1995 for example, there was a single fertilizer recommendation for all hybrid maize grown in Malawi:

> *"The farmer is to apply 87 kg of Diammonium phosphate (DAP) per hectare by dolloping 10 cm away from the planting station and 10 cm deep soon after emergence. This is to be followed two to three weeks later by an application of 175 kg of urea per hectare using the same method. The nutrients applied using these amounts of these fertilizers amount to 96 kg of nitrogen and 40 kg of phosphate per hectare."*

This recommendation was used throughout the whole country by field assistants prior to the initiation of area-specific fertilizer verification trials during the 1995/6 growing season. The trials were initiated to juxtapose the use of this single-blanket recommendation for mainly two reasons. Firstly, it was realized that the geographical, climatic and soil conditions were extremely heterogeneous within the country and that any simplistic recommendation across the board did not provide higher maize yields across all areas. Secondly, it was noted that the dynamics of fertilizer and maize prices in the country made the single-blanket recommendation inefficient as the price differentials during the time the recommendation was made changed dramatically as of 1995/96 growing season. Area specific recommendations based on the analysis of these trials were developed in 1999 and documented in the guide to agricultural production in Malawi (Government of Malawi 2012). However, Malawi's policies rooted in the broad recommendation ideology remain prevalent. One such example is the farm input subsidy program that provides the same amount of fertilizer across the whole country regardless of the spatial economic circumstances of the poor smallholder farmers. Assuming recommendations are adopted prima facie and that policies to support them can be developed; this study addresses the question of whether spatially specific recommendations are more profitable to smallholder farmers and society.

This study makes three main contributions. Firstly, it provides an alternative and simple theoretic framework as to why specific recommendations and precision agriculture are mostly found to be profitable when quasi-fixed cost differentials are ignored. This framework relies on spatial scale stochastic dominance and Jensen's inequality. Secondly, this paper is among a few (e.g. Brorsen 2013) that casts the determination of optimal nitrogen recommendations into a Bayesian decision theoretic framework that takes into account parameter uncertainty. Finally, we suggest an empirical strategy that naturally comes out from the theory by using spatially varying coefficient models so as to determine spatially optimal nitrogen recommendations. To the best of our knowledge, this model has never been used in agricultural economics literature before. The rest of the paper is organized as follows. Chapter 2 summarizes the scope and objectives of the study. Chapter 3 summarizes the data issues and proposes the theoretical and empirical models that are used. Chapter 4 provides the results and discussion. Finally, conclusions are offered in chapter 5.

2.0 Objectives

The main agricultural research and economic problem addressed in this study is that of developing fertilizer recommendations to be used by farmers based on a sample of plots. This problem has statistical, agronomic and economic dimensions. In a statistical sense, it concerns the uncertainty and heterogeneity of parameters in model estimation across space and thus the optimal decisions recommended to smallholder farmers. It is mostly forgotten in the presentation of agricultural research results and development of recommendations that point estimates are not known with certainty. In terms of agronomy, it is a well-known fact that responses to fertilizer application vary by topography, soil type and so on. This is why precision agriculture has been gaining ground in developed countries. Unfortunately, agricultural experimental research in Malawi and other African countries is still one aimed at finding "silver bullets" e.g. best variety for all places or least cost legume combinations for all seasons.

These two dimensions of the problem (statistical and agronomic) result in a fundamental economic decision making challenge as any deviance due to these problems results in suboptimal decision rules. This study therefore provides a conceptual and theoretical model based on decision-theoretic tools particularly Bayesian decision theory that helps in developing the optimal decisions that researchers can use to provide relevant research recommendations to smallholder farmers given the uncertainty in their soil fertility research results. However, the study does not attempt to provide prescriptive recommendations for smallholder farmers to use in Malawi as this would require a complex process beyond the scope of the current study. The general objective of this study is to analyze the economically optimal spatial scales for targeting soil fertility recommendations in Maize based farming systems in Malawi. The specific research objectives are:

i. To analyze the regional and temporal heterogeneity in maize response to fertilizer application in Malawi;
ii. To analyze the regionally disaggregated, economically optimal nitrogen rates for Malawi; and
iii. To determine the differences between using broad recommendations and spatially specific recommendations.

3.0 Methods

3.1 Data

The study uses area specific fertilizer verification trials data for the 1995/96 and 1997/8 growing seasons. The experimental protocol was documented in the manual for extension workers developed to guide the implementation of the trials (Government of Malawi 1995). The trials were carried out under the guidance of a Maize Productivity Task Force (MPTF) consisting of national and international experts. After the successful implementation of the 1995/96 verification trials, area specific recommendations were developed for each of the Extension Planning Areas (EPAs) in the country. The variations were based on soil structure/texture and the objectives of the farmer (home consumption or for market). The Fertilizer Verification Trials were modified in 1997/98 to improve the collection of data. The modifications were compiled in a manual that was developed to guide the implementation of the trials (Government of Malawi 1997).

3.1.1 Design of the Trial

In 1995/6, 1,676 of 1,919 agronomic trials rolled out across the country were successfully implemented to evaluate six different inorganic fertilizer packages for hybrid maize grown by smallholders. The distribution of successful trials was unbalanced across the sites. The map (Figure 1) below shows the distribution of trials across the extension planning areas in the country. The maximum number of trials per Extension Planning Area was 25 while the lowest was 1. The map also illustrates the differences in the number of trials planted for each of the seasons. In terms of treatments, six treatments were tested in 1995/96 trials while four treatments were tested in the 1997/98 as shown in Table 1.

Table 1: Six fertilizer treatments were tested in 1995/96 and four in 1997/98

Treatment	**Nutrients**			**Fertilizer**	
	Nitrogen (kg per ha)	Phosphate (Kg per ha)	Sulphur (Kg per hectare)	Basal (50kg bags per ha)	Top dressing (50kg bags per ha)
			1995/96 Growing Season		
1	96	40	0	1.75DAP	3.5Urea
2	0	0	0	0	0
3	35	0	0	0	1.5Urea
4	35	10	2	1 (23:21:0+4S)	1 Urea
5	69	21	4	2 (23:21:0+4S)	2 Urea
6	92	21	4	2 (23:21:0+4S)	3 Urea
			1997/98 Growing Season		
2	0	0	0	0	0
4	35	10	2	1	1
5	69	21	4	2	2
6	92	21	4	2	3

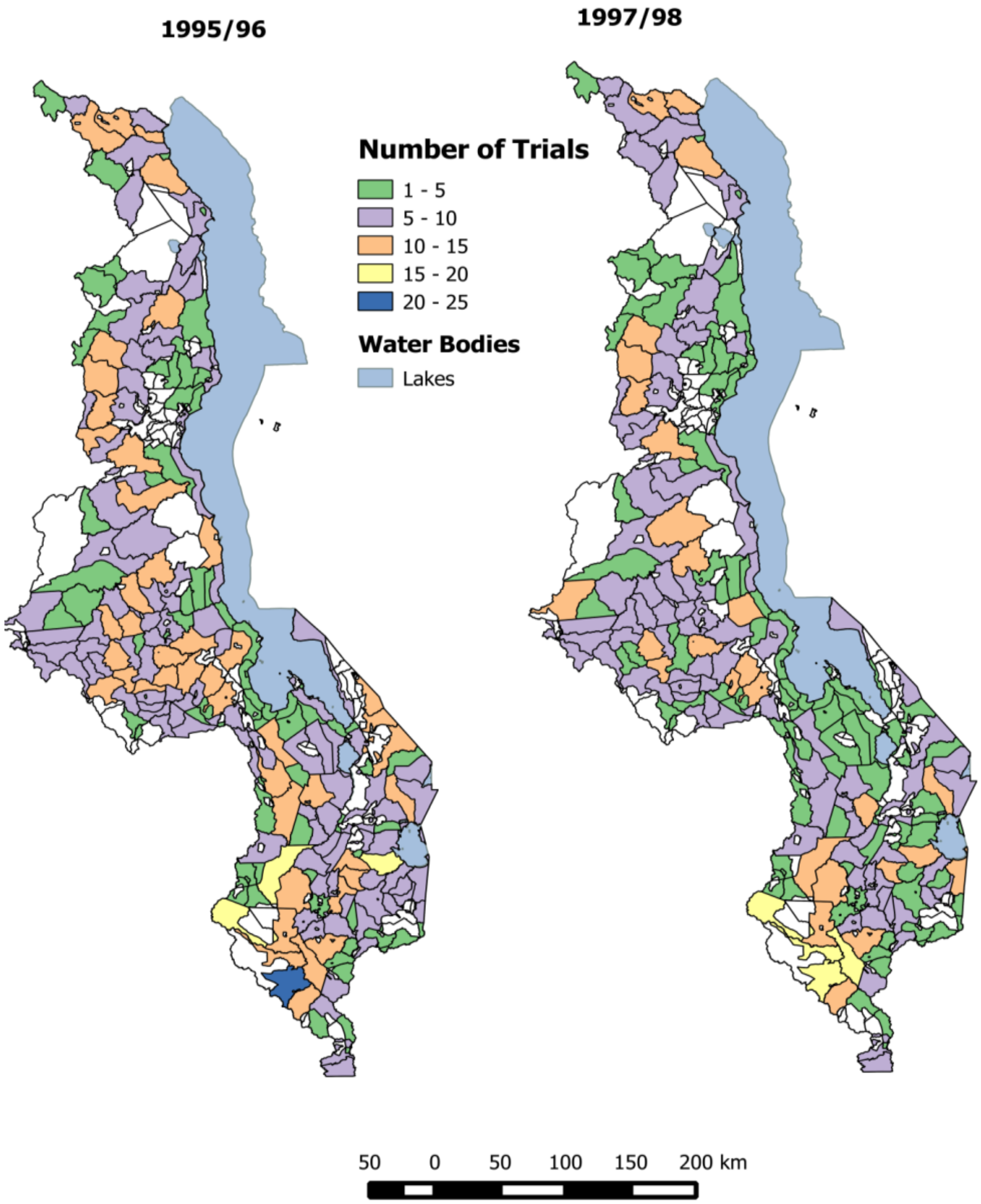


Figure 1: Distribution of fertilizer trials across the country.

In each of the two seasons, two hybrid maize varieties were planted; MH17 was planted in upland sites with historically good rainfall conditions and MH18 was supplied for trials in lowland areas and at those upland sites in rain-shadow areas. A few sites also tested composite varieties.

3.2 Conceptual Framework

The spatial scale economic decision problem, that of identifying the economically optimal spatial scale for targeting agricultural recommendations, requires a conceptual framework that takes into account the spatial nature of the problem, the biological and agro-ecological processes associated with agricultural systems, and the associated socio-economic factors. Figure 2 illustrates the different levels of analyzing the problem. Broad recommendations and precision agriculture are considered as the two polar ends. The former have been the common characteristic of recommendation guidelines from agricultural research. For instance, it is common to find a soil fertility improvement research program directed towards finding that superior legume integration system that increases yields in all regions.

At the other end of the spectrum, there have been studies directed towards maximizing the potential of targeting technologies or inputs to their most suited finer points within the field using geographical information systems (GIS) and precision agriculture technology. With precision agriculture it is well known that costs of information, extension and research and development for fine-tuning the recommendations are higher. Thus a decision maker or researcher faces a tradeoff between the potential yield benefits due to better specific recommendations and the higher costs expected to be incurred to generate the recommendations.

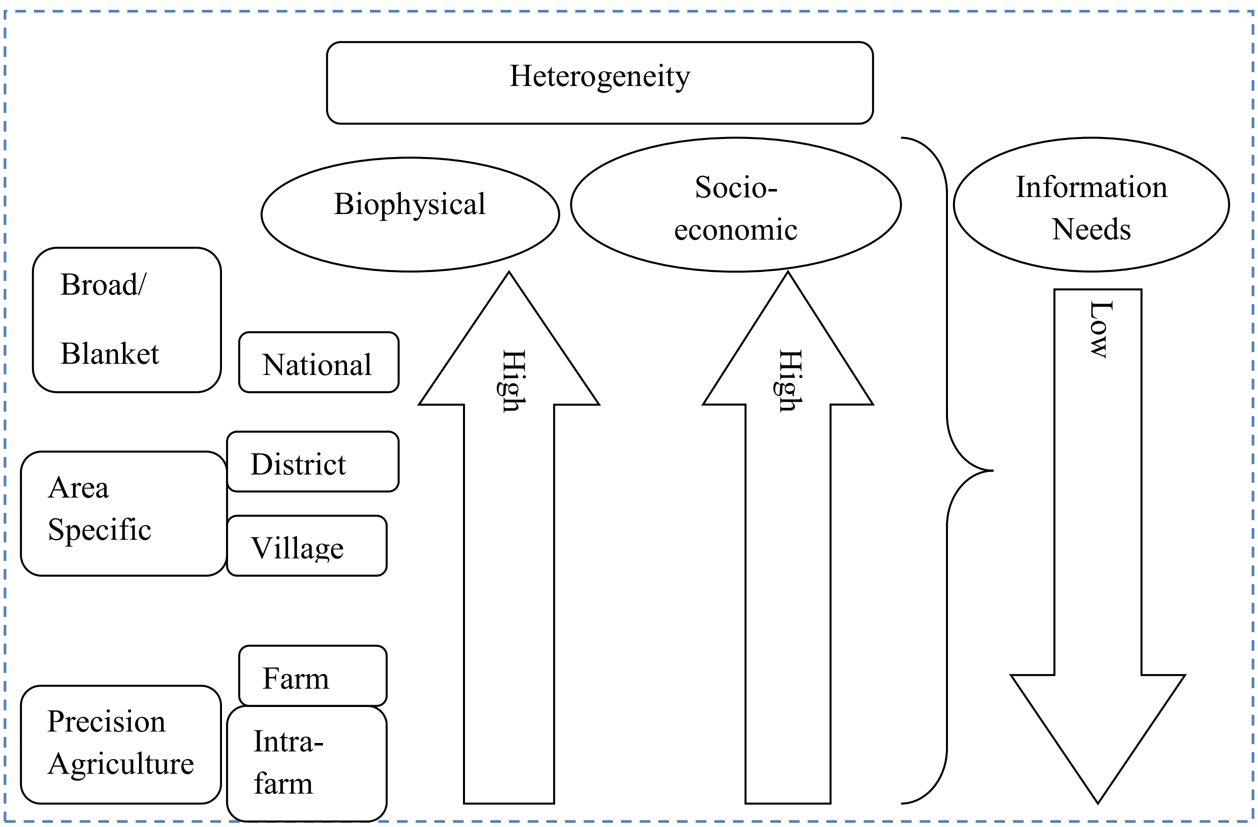


Figure 2: Levels of spatial scaling from precision agriculture to blanket recommendations

This conceptual framework is consistent with the concept behind the development and testing of recommendation domains by different agencies of the CGIAR (Consultative Group of International Agriculture Research Institutes) especially CIMMYT (International Maize and Wheat Improvement Center). A recommendation domain is defined as a group of farmers facing a set of agronomic circumstances similar enough for the same recommendation to be appropriate for all the farmers in the group (Jauregul and Sain 1992). In order to develop a recommendation domain, one faces the problem of aggregation bias. Can the aggregate bio-economic relations reflect individual bio-economic relations? The problem of aggregation has been and is still the most elusive in economics and statistics. King *et al*. (1993) argued that without appropriate aggregation of site specific responses, serious biases can be introduced into a policy analysis.

3.3 Theory

3.3.1 Spatial Bio-Economic Optimization model

A spatial bio-economic optimization framework is the appropriate integrated strategy for modeling heterogeneous agricultural processes. Several scholars (Liu, Swinton, and Miller 2006; Bullock, Lowenberg-DeBoer, and Swinton 2002; and Oishi 2006) have presented the spatial profit maximization problem for spatially specific input application. It can be conceptualized as optimization problem for each individual farm plot **i** in each spatial unit **j** (e.g. particular country, District, Village) such that each plot within a respective spatial configuration can be indexed as ij. This formulation is consistent with the use of plot specific response functions to develop recommendation domains (Jauregul and Sain 1992). Equation (1) shows the general spatial input management optimization model:

$$\text{Max}_{X}\ E(\pi) = \sum_{j=1}^{M}\sum_{i=1}^{N}(P_y Y_{ij} - P_x X_{ij}) - G - FC \ \ \textit{subject to}\ Y_{ij} \leq f(X_{ij}, Z, C). \quad \textbf{(1)}$$

In Equation (**1**), P_y is the output price, P_x is the vector of input prices, Y is the yield and X_{ij} is a vector of variable inputs (e.g. inorganic fertilizer use, organic fertilizer use, seeds) . For the farmer, it is assumed that profit is increasing in yield and that crop yield is a function of managed variable inputs X_{ij} (e.g. fertilizer, seeds, pesticides), unmanageable stochastic inputs, Z (e.g. rainfall) and unmanageable, but non-stochastic site characteristics (e.g. topography), C (Bullock, Lowenberg-DeBoer, and Swinton 2002). The last two terms in the objective function are additional costs to the variable inputs: G is quasi-fixed costs of collecting and analyzing downscaled spatial information, and FC are other fixed costs.

The key assumptions for using the biological response relation $Y_{ij} = f(X_{ij}, Z, C)$ in equation (**1**) are: continuous smooth causal relation between the inputs and yield, diminishing returns prevail and decreasing returns to scale (Dillon and Anderson 1990). The classic economic rule says that the optimal X^* is obtained when the marginal cost is equal to the marginal revenue (Hurley, Malzer and Kilian 2004):

$$P_Y \frac{\partial Y(X^*, Z, C)}{\partial X} - P_x \leq 0 \text{ with equality for } X^* > 0. \quad \textbf{(2)}$$

One important aspect in the use of this response relation concerns the functional form. The common examples include quadratic, Cobb-Douglas, Translog, linear response plateau and Spillman-Mitscherlich functional forms. Among these, the quadratic functional form is the most extensively used form because it's simple and consistent with the agronomic theory. For a quadratic functional form, the economically optimal input rate X^* for a single input (N fertilizer in this case) and single output is $X^* = \frac{R-\beta_1}{2\beta_2}$ where $R = \frac{P_X}{P_Y}$.

The optimal amount of the variable input depends on the output price, the input price and the fixed inputs through implicit function theorem (Hurley, Malzer and Kilian 2004):

$$\frac{\partial X^*}{\partial Z} = -\frac{\partial^2 Y(X^*, Z, C)}{\partial X \, \partial Z} \frac{\partial X^2}{\partial^2 Y(X^*, Z, Z)} \neq 0 \quad \text{if} \quad \frac{\partial^2 Y(X^*,Z,C)}{\partial X \, \partial Z} \neq 0. \tag{3}$$

If however, $\frac{\partial^2 Y(X^*,Z,C)}{\partial X \, \partial Z} = 0$, it implies that there is no interaction between the variable and fixed input as the optimal amount of variable input does not change with the amount of fixed input such that there is no value in varying the variable input. This condition summarizes the precision agriculture hypothesis which asserts that, *"farmers or the environment can benefit from varying management within or between fields."* Thus, confirming the precision agriculture hypothesis without knowledge of important fixed inputs is useful because it indicates whether searching for such inputs is worth an effort. If the precision agriculture hypothesis cannot be confirmed generally or the value of discovering which fixed inputs are important is small, then it makes sense to devote research effort elsewhere (Hurley, Malzer and Kilian 2004).

According to Byerlee and Anderson (1969), in the absence of perfect knowledge, the decision maker or producer is unable to choose the levels of the variable input which will necessarily be optimal ex-post and the decision maker may incur a loss i.e. a cost of uncertainty. Thus, the interaction between controlled and un-controlled factors is the necessary condition for additional information on the uncertain factors to have economic value. For this study, the economic value of going with specific recommendations will be evident if and only fertilizer response varies across locations.

In addition to the validation of the necessary condition for site-specific crop response functions or site-specific crop economic return functions, it is also important to understand that the management activities are usually related to the farmer specific individual characteristics some of which may be observable while others are not. Precision agriculture for smallholder farmers therefore has to incorporate both the management activities; factors not managed by the farmer which are typically site-specific; and farmer specific random factors. In this framework, there are therefore a myriad of factors to consider. It has been argued in the recent literature that point estimates of the response rates and thus the economically optimal input (nitrogen) rates have uncertainty which is normally ignored (Hernandez and Mulla 2008). Our framework addresses this concern by considering the whole distribution of parameters.

3.3.2 Aggregation Theory and Optimal Spatial Scaling

In this section, we review the theory of aggregation of economic relations and thus justification for or against the use of broad recommendations. This theory is based on the seminal work by Henry Theil in the 1954 book *"Linear Aggregation of Economic Relations."* While much of the theory in the book was on linear relations, Theil also provided an introduction to generalizations in the case of nonlinear economic relations e.g. the quadratic relation.

In an ideal case, one would want a response function estimated at a finer scale as much as possible from as many replicates at each of those points. But the reality is that only a coarser scale than the naturally occurring soil microclimatic processes can be used. Assume that such ideal equations or micro level equations are a set of M quadratic micro equations represented by an index j where $j = 1, \dots, M$ (disregard the seasonal dimension $t = 1, \dots, T$ and the other controls for now):

$$Y_j = \beta_{0j} + \beta_{1j}X_j + \beta_{2j}X_j^2 + \varepsilon_j \text{ for } j=1,\dots, M \quad \textbf{(4)}$$

where Y_j is shorthand for Y_{ij} and ε_j is the conventional error term.

Then consider an aggregate model represented by a single quadratic macro-equation:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon. \quad \textbf{(5)}$$

The macro-variables being $Y = \sum_{j=1}^{M} Y_j$ and $X = \sum_{j=1}^{M} X_j$. Note that the macro-variables can also be expressed as expectations or quantities per hectare. According to Theil, we can use the rule of perfect aggregation to determine if the macro-equation perfectly aggregates the micro-equations. The rule of perfection is that

> *" There is no contradiction between the macro-equation and the micro-equations corresponding to it, for whatever values and changes assumed by the micro-variables and at whatever point or period of time"* (Theil 1954, p. 140).

In order to determine conditions under which the aggregate model or the broad model perfectly corresponds to the micro-equation results, let's assume that the micro-variables increase by certain infinitesimally small amounts represented by ΔX_j so that X increases by $\Delta X = \sum_{j=1}^{M} \Delta X_j$. The increase in Y due to the change in **X** is;

$$\frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X. \tag{6}$$

And for the micro-equations:

$$\sum_{j=1}^{m} \frac{\partial Y_j}{\partial X_j} = \sum_{j=1}^{m} (\beta_{1j} + 2\beta_{2j} X_j) \tag{7}$$

where again $\beta_1 = \sum_{j=1}^{m} \beta_{1j}$ and $\beta_2 = \sum_{j=1}^{m} \beta_{2j}$.

Thus, Theil's rule of perfection implies that the expression $\frac{\partial Y}{\partial X} = \sum_{j=1}^{m} \frac{\partial Y_j}{\partial X_j}$ must be identical for whatever values of the micro-changes and hence of the macro-variables as soon as $\Delta X = \sum_{j=1}^{M} \Delta X_j$. In order to investigate the possibility of this perfection, let's assume that all the changes in X are zero for all other locations except one location j such that $\Delta X = \Delta X_j$. Then the equality $\frac{\partial Y}{\partial X} = \sum_{j=1}^{m} \frac{\partial Y_j}{\partial X_j}$ simplifies to

$$\beta_1 + 2\beta_2 X = \beta_{1j} + 2\beta_{2j} X_j. \tag{8}$$

This is achieved for all values of ΔX_j if the corresponding parameters are equal. The following proposition can thus stand as both necessary and sufficient for the possibility of a perfect aggregate/broad quadratic response function:

Proposition I-Perfect aggregation of a quadratic response function: Perfect aggregate quadratic response function is achieved if and only if; $\beta_1 = \beta_{1j}$ and $\beta_2 = \beta_{2j}$ for all locations j (Theil 1954, p. 142).

The proof for the more general polynomial case which also includes the quadratic case is provided in (Theil 1954). The main conclusion derived from the condition is that the perfect aggregate quadratic response function can be achieved under very stringent circumstances. One can consider multiple weighted moments to achieve a perfect representation of the micro-equations. This result thus justifies the consideration of precision agriculture and the quest for conditions under which specific recommendations based on the micro-equations are better than the broad recommendations based on the macro-equation.

Since proposition I entails an obvious deviation between broad recommendations based on the macro-equation and the specific recommendations based on the micro-equations, we can consider the difference as a potential aggregation bias. When we attach prices of inputs and outputs to this bias, we get economic benefits or losses of broad recommendations. How does this knowledge help in determining the optimal spatial scale of making N recommendations? Here, let's first use pooled/aggregate estimated quadratic response function and by first order conditions get the solution, $X^* = \frac{R-\beta_1}{2\beta_2}$ where $R = \frac{P_X}{P_Y}$.

Inserting this optimal nitrogen rate into the yield response equation and hence the profit equation, gives the maximum profits that can be attained at that particular price ratio for all real-values of X. This result is consistent with Jensen's inequality of expectations which in modified form states that:

Proposition II-Jensen's inequality of a quadratic profit function: "if $\pi: \mathbb{R} \rightarrow \mathbb{R}$ *is a concave function defined on* $\mathbb{R}$ *and* $E[|Y|] < \infty$*, then;* $\pi[E(Y|X)] \geq E[\pi(Y)|X]$ *for all X in a set that has probability equal to 1* (Wooldridge 2010, p.31) *."*

Let's consider two distinct locations 1 and 2 for expository purposes, then by Jensen's inequality it should be the case that

$$\pi_1[E(Y_1|X_1)] \geq E[\pi_1(Y_1)|X_1] \text{ and} \quad \textbf{(9)}$$

$$\pi_2[E(Y_2|X_2)] \geq E[\pi_2(Y_2)|X_2]. \quad \textbf{(10)}$$

We can thus write the aggregate or broad profit equation as:

$$\pi[E(Y_1 + Y_2)|(X_1 + X_2)] \geq E[\pi(Y_1 + Y_2)|(X_1 + X_2)] \quad \textbf{(11)}$$

The challenge of determining the spatial scale of making recommendations is therefore that of determining whether;

$$\pi[E(Y_1 + Y_2|X_1 + X_2)] >/=/< E[\pi_1[E(Y_1|X_1)] + \pi_2[E(Y_2|X_2)]] \quad \textbf{(12)}$$

Inserting the optimal levels $X^*, X_1{}^*, X_2{}^*$ into this profit expression we get;

$$\begin{aligned} P_Y\,[\beta_0 + \beta_1 X^* + \beta_2 X^{*2}] - P_X[X^*] &>/=/ \\ &< E[(P_Y(\beta_{01} + \beta_{11} X_1{}^* + \beta_{21} X_1^{*2}) - P_X[X_1{}^*]) \\ &+ (P_Y\,(\beta_{02} + \beta_{12} X_2{}^* + \beta_{22} X_2^{*2}) - P_X[X_2{}^*])] \end{aligned} \quad \textbf{(13)}$$

or

$$P_Y\left[\beta_0+\beta_1\left\{\frac{R-\beta_1}{2\beta_2}\right\}+\beta_2\left\{\frac{R-\beta_1}{2\beta_2}\right\}^{*2}\right]-P_X\left[\left\{\frac{R-\beta_1}{2\beta_2}\right\}\right]$$
$$>/=/< \ E[(P_Y\left(\beta_{01}+\beta_{11}\left\{\frac{R-\beta_{11}}{2\beta_{21}}\right\}+\beta_{21}\left\{\frac{R-\beta_{11}}{2\beta_{21}}\right\}^{*2}\right)-P_X\left[\left\{\frac{R-\beta_{11}}{2\beta_{21}}\right\}\right]\Bigg)$$
$$+\left(P_Y\left(\beta_{02}+\beta_{12}\left\{\frac{R-\beta_{12}}{2\beta_{22}}\right\}+\beta_{22}\left\{\frac{R-\beta_{12}}{2\beta_{22}}\right\}^{*2}\right)-P_X\left[\frac{R-\beta_{12}}{2\beta_{22}}\right]\right)] \quad \textbf{(14)}$$

where P_Y and P_X are assumed to be constant scalars such that the form of the whole expression depends on values of $\beta_0, \beta_1, \beta_2, \beta_{01}, \beta_{11}, \beta_{21}, \beta_{02}, \beta_{12}$ and β_{22}. Instead of using the expectation operator, one may also assume any other convex combination of the two objective functions. It is apparent from the expression that a quest for an analytical solution to this problem can be daunting. Most importantly, each of the fitted parameters $\beta_0, \beta_1, \beta_2, \beta_{01}, \beta_{11}, \beta_{21}, \beta_{02}, \beta_{12}$ and β_{22} have some estimation uncertainty due to a lack of fit of the quadratic function to the data, making the optimal X^*, X_1^* and X_2^* ratios of two normally distributed functions whose means and variances were determined from the regression estimations (Jaynes 2011). Although optimal levels are defined by a ratio of two normal distributions, the distribution of the economically optimal nitrogen is not typically normal, but depends on the coefficients of variation and the correlation coefficient of the parameters (Shanmugalingam 1982). If we however assume normality of the expected profits, we can characterize the form of the analytical solution to this problem. Firstly, we assume a well-known property of normal distributions as stated in proposition III.

Proposition III-Normality: If π is normally distributed, $\pi \sim N(\mu, \sigma)$ with μ as the mean and σ as the variance; then a linear combination such as, $\pi_\alpha = \alpha\pi_1 + (1-\alpha)\pi_2$, where π_2 is a constant and for all $\alpha > 0$, will also be normally distributed: $\pi_\alpha \sim N(\alpha\pi_1 + (1-\alpha)\pi_2), \alpha\sigma)$. (adapted from Levy 2006, p.199).

This proposition then allows us to compare the two profit distributions. It is important to note that Proposition III has received mixed debates on its appropriateness especially when the functions are not independent (Rosenberg 1965). In our framework however, we assume that profit functions in two different locations are independent. Assuming a riskless decision problem, decision makers/researchers will choose using Theorem 1.

Theorem 1: First Order Spatial Scale Stochastic Dominance - Let π^* and π' be the profits attained when two distinct recommendations are made with the former being site specific while the latter being broad. The cumulative distribution functions are Φ^* and Φ' respectively. Assume that π^* and π', the two random variables, are normally distributed with the following parameters: $\pi^* = N(\mu_1, \sigma_1)$, $\pi' = N(\mu_2, \sigma_2)$. Then Φ^* will dominate Φ' by First Order Spatial

Scale Stochastic Dominance, if and only if the following holds: (a) $\mu_1 > \mu_2$ *and* (b) $\sigma_1 = \sigma_2$ (adapted from Levy 2006, p. 200).

Proof. The informal proof for this theorem is straightforward. One important property of normal distributions is that whenever $\sigma_1 = \sigma_2$ for two different distributions, the cumulative distribution functions do not intercept. In addition, the condition $\mu_1 > \mu_2$ implies that $\Phi^*(\pi) < \Phi'(\pi)$ for all π hence Φ^* stochastically dominates Φ'. The other side of the proof is also similar as Φ^* stochastically dominates Φ' implies that Φ^* and Φ' will not intercept. This completes the proof (adapted from Levy 2006). □

The normality assumption on the distribution of profits may be questionable thereby warranting distribution free decision rules. Theorem II provides the general criterion for first order spatial scale stochastic dominance.

Theorem II: General First Order Spatial Scale Stochastic Dominance- Let F(.) and G(.) be two cumulative distributions of profits defined based on different spatial scales at which they were optimized. The distribution of profits F(.) will first order stochastically dominate the distribution of profits G(.) if and only if F(π) is less than or equal to G(π) for every π and there is at least one π for which a strong inequality holds (adapted from Levy 2006, p. 56).

This theorem essentially means that for every amount of profit we get by applying the optimal amount of fertilizer; the probability of getting at least a certain level of profits is higher under the distribution of profits for the maximal spatial scale than any other spatial scale. The proof for this theorem can be found in (Mas-Colell, Whinston and Green 1995). Some of the important properties of first order spatial scale stochastic dominance include (i) it does not imply that every possible location's profit of the superior spatial scale is larger than every possible profit of the inferior spatial scale of making recommendations and (ii) although it implies that the mean under F(.) is greater than the mean under G(.), a ranking of the means of the two distributions does not imply one stochastically dominates the other, rather the entire distribution matters. Figure 3 shows three hypothetical cumulative distribution functions: F(.), G(.) and Q(.). In the figure, F(.) first order stochastically dominates G(.) since F(.) < G(.). However, other higher order stochastic dominance criteria are needed to compare F(.) and Q(.) or G(.) and Q(.). We only consider first order stochastic dominance in this paper because the analysis of second order stochastic dominance requires further assumptions on risk preferences, a complexity that is beyond the scope of this study. If one makes other risk assumptions, then higher order spatial scale stochastic dominance decision rules may apply.

Figure 3: First order "spatial scale" stochastic dominance

Proposition II and Theorem II lead to Proposition IV which is the heart of the study and makes conclusive arguments about the nature of the relationship between broad recommendations and specific recommendations.

Proposition IV: If a profit function $\pi(\theta, X)$ is a concave function of X for all states of nature/uncertain parameters θ, then the optimized levels of profits $\pi(\theta, X^*)$ from using specific recommendations X^* will always be the same or stochastically dominate optimized levels of profits $\pi(\theta, X')$ from using broad recommendations X' assuming zero cost to the decision itself[1].

Proof. The proof for this proposition relies on the first order stochastic dominance theorem and Jensen's Inequality stated previously. Let F(.) be the cumulative distribution function of profits from using specific recommendations and let G(.) be the cumulative distribution function (cdf) of profits from using a broad recommendation. Without loss of generality, assume F(.) dominates G(.) as stated in the conclusion of the proposition IV. Then by definition, $F(.) \leq G(.)$ *for every* π. Since we are considering different spatial scales, the profit function from using specific recommendations is essentially a convex combination of the location specific profits. Therefore, $F(\pi)$ is the cdf for $\pi(\theta, X^*)$. The profits from a broad recommendation are a direct characterization of the profits and can be represented as $\pi(\theta, X')$ with a cdf, $G(\pi)$. Let $X^* = PX'$ where P is a sufficient statistic for θ (i.e. the conditional distribution of X is independent of θ) By Jensen's Inequality, if π is a concave function of X; then

$$\pi(\theta, X^*) = \pi(\theta, PX') \geq P[\pi(\theta, X')]$$

[1] This proposition is similar to the Blackwell's theorem as stated and proved in Berger (1985). The reader is refered to pages 35 to 41 of the book for details.

which is also an implication of the definition of first order spatial scale stochastic dominance $F(.) \leq G(.)$. One can also use complex and compelling results on preservation of log-concavity of cumulative distribution functions as reported by Bagnoli and Bergstrom (2004) in proving this proposition □

It is imperative to note that the assumption of zero cost to the decision itself can be relaxed so that a lower bound cost can be defined and compared to the differences in the two distributions. How can these theoretical results be used in practical empirical settings given uncertainty over the parameter space and the distribution of the expected profits? This is where we turn to numerical methods particularly the Monte Carlo approach. Precisely, we develop a Bayesian method for calculating the optimal nitrogen rate which can then be directly compared to make a judgment as to which spatial scale is most profitable.

3.3.3 Determining the Optimal Nitrogen Rate using Bayesian Decision Theory

The literature on Bayesian estimation of yield response to nitrogen is quite recent and relatively rare in being used to determine optimal allocation of limiting resources in production economics (Ouedraogo and Brorsen 2014). However, Bayesian estimates via Markov Chain Monte Carlo (MCMC) have more intuitive and correct interpretation in spatial studies since the draws from MCMC sampling can be used to produce posterior distributions for functions of the parameters that are of interest thereby making the testing of complicated parameter relationships like the one we are concerned with in this study quite easy (Lesage and Pace 2009; Gelfand *et al*., 1990).

It is imperative to note that Bayesian decision theory is not just an alternative to frequentist statistics as is bayesian statistics[2]; it is rather an alternative to the plug in approach showed in Section 3.3.1. According to Dorfman(1997)**,** Bayesian decision theory differs from the plug in approach in that in the latter the researcher takes estimates of unknown parameters as if they were certain and ignores the effect that uncertainty may have on choosing the optimal set of the control variables or independent variables. The former however is regarded as decision making under estimation risk in that it works with uncertain parameter estimates in the solution of an optimal decision problem. When do these two approaches give the same result? These two approaches will give the same answer when the certainty equivalence principle holds. The certainty equivalence principle states that

> *"Ignorance of parameter uncertainty as assumed by the plug in approach will lead to the same solution as a bayesian approach that fully accounts for the parameter uncertainty if and only if the posterior distribution of the parameters is normal and the objective function is linear-quadratic in the unknown parameters"*- (Dorfman 1997).

The general procedure in Bayesian decision theory is that one chooses the values of the control variables (nitrogen in this case) that minimize the expected loss of that decision, where the

[2] For a complete treatment of spatial Bayesian statistics and/or econometrics employed in this study, interested readers are encouraged to consult (Banerjee, Carlin and Gelfand 2015).

expectation is taken with respect to the posterior distribution of the unknown parameters (Berger 1985). We cast the Bayesian decision theory as a decision problem of determining the optimal level of nitrogen and then which spatial scale (broad or specific) is the most profitable. Let the unknown parameters involved in the decision problem be θ and assume they have a posterior distribution defined by $p(\theta|Y,X)$ where X is the vector of independent variables and Y is the dependent random variable with a probability density given x defined as $p(Y|X,\theta)$. The optimal level of nitrogen or any control variable is determined by the loss equation

$$X^* = \operatorname{argmin} \int_{\theta}\int_{Y} F(Y,X|\theta)\ p(Y|X,\theta)p(\theta|Y,X)dYd\theta \tag{15}$$

The expected loss of choosing a particular value of the control variable X is evaluated by integrating out uncertainty concerning the unknown parameters θ and any residual uncertainty about the endogenous variables Y (Dorfman 1997). Since our aim is to maximize profits, we can define the loss function $F(Y,X|\theta)$ as;

$$F(Y,X|\ \theta\ ) = \pi_{max} - \pi(Y|\theta) \tag{16}$$

where $\pi(Y|\theta)$ is the profit earned from selling Y at a particular price ratio conditional on parameters θ .The value of π_{max} is however unknown and can be assumed to be a constant such that the loss function reduces to $F^*(Y,X|\theta) = -\pi(Y|\theta)$ which is to be minimized by the same value of Y as the original loss function $F(Y,X|\theta)$. Minimizing a negative of profit is equivalent to maximizing profit such that we choose the optimal nitrogen (X^*) by

$$X^* = \operatorname{argmax} \int_{\theta}\int_{Y} F'(Y,X|\theta)\ p(Y|X,\theta)p(\theta|Y,X)dYd\theta \tag{17}$$

where $F'(Y,X|\theta) = \pi(Y|\theta)$. In this case, the integrals are taken to eliminate the agent's uncertainty regarding the true but unknown vector of parameters θ. In our analysis, we assume the value of Y is deterministic given X and θ such that $p(Y|X,\theta)$ is degenerate and thus will not affect the calculation of the expected profits. If $F^*(Y,X|\theta)$ is bounded above which is trivially true by definition of the functional form, Fubini's theorem can be invoked to interchange the order of integration and use Bayes rule to get:

$$X^* = \operatorname{argmax} \int_{Y} [\int_{\theta} F'(Y,X|\theta)\ p(\theta|Y,X)d\theta]p(Y)dY \tag{18}$$

Since Y is degenerate given X and θ, we can drop Y in the subsequent derivations. With this derivation, we get a similar result to that reported by (Brorsen 2013; Byerlee and Anderson 1969) in which using Bayesian decision theory, the optimal level of nitrogen is determined by maximizing expected profits as follows:

$$\mathrm{Max}_{X\geq 0}\int E[F'(X|\theta)]p(\theta|X)d\theta$$
$$\mathrm{Max}_{X\geq 0}\int E\pi(X|\theta)p(\theta|X)d\theta \qquad \textbf{(19)}$$

where $E\pi(X|\theta)$ is the expected profit, θ is the vector of relevant parameters including the coefficient vector β for the variable input of interest, and $p(\theta|X)$ is the posterior distribution for θ. Muus, Scheer & Wansbeek (2002) argued that the plug in decision rule in section 3.3.1 is simply a half space approximate of the Bayesian decision region and may lead to suboptimal level of profits. The Bayesian method has the advantage that it uses the whole distribution of the parameters other than just point estimates. This method explicitly takes into account the estimation uncertainty (Muus, Scheer and Wansbeek 2002) when the decision maker or researcher does not know the parameter vector**.** If the integral of the cumulative distribution function over the posterior density of the parameters can be evaluated, then a closed form and analytical solution can be derived. The normality assumption is usually imposed to get an analytical solution. The general result from the derivation of an analytical solution is that a Bayesian decision rule (i.e. that takes into account the uncertainty) leads to higher profits whenever there is greater uncertainty about the parameters. This is the case because the expected profit from the plug-in approach is simply a subset of the set of optimal rates determined by the Bayesian decision rule. Instead of this abstract analytical solution one can use a Bayesian statistical approach through Monte-Carlo integration. Define $\tilde{\theta}_1, \ldots, \tilde{\theta}_n$ as the random vectors of parameters drawn from the posterior distribution using the Metropolis-Hastings algorithm. The optimum is obtained by approximating the integral with Monte Carlo method to get the expectation form (equation 20) that converges to equation 19 when *n* goes to $+\infty$ according to the law of large numbers

$$\mathrm{Max}_{X\geq 0}(\tfrac{1}{n})\textstyle\sum_{i=1}^{n} E\pi\left(X|\tilde{\theta}_i\right). \qquad \textbf{(20)}$$

This can then be solved by conventional nonlinear optimization. The advantage of the functional form used (a quadratic response function) is that whether with stochastic or non-stochastic parameters one gets the same analytic solution because the random parameters enter the profit function linearly (Tumusiine, et al. 2011). Thus, having determined the spatially specific optimal levels X^* across the whole distribution of $\tilde{\theta}$, this can be compared to the broad optimal X' given the whole distribution of $\tilde{\theta}$:

$$\mathrm{Max}_{X^*\geq 0}(\frac{1}{n})\sum_{i=1}^{n} E\pi\left(X^*|\tilde{\theta}_i\right)$$
$$\textbf{vs.} \qquad \textbf{(21)}$$
$$\mathrm{Max}_{X'\geq 0}(\frac{1}{n})\sum_{i=1}^{n} E\pi\left(X'|\tilde{\theta}_i\right)$$

The resultant expected profits are then compared using the first order spatial scale stochastic dominance rule. Based on the theoretical results, the first expression will stochastically dominate the last expression. The main reason for this is Jensen's inequality with a concave profit function. This conclusion is similar to what Bullock *et al*. (2009) asserted: "assuming information and technology are free, given any information structure the farmer cannot lose from solving a less constrained problem. " In our case, the first expression results in a less constrained informational problem. In fact, establishing that the first expression first order stochastically dominates the last expression implies that the former is more informative than the latter. Our approach nevertheless has several advantages. Firstly, it combines theory with empirics on incorporating parameter uncertainty. Secondly, our approach does not require a lot of assumptions regarding preferences. By using profit maximization which is well known among economists and agronomists, our approach allows cross-disciplinary understanding of the value of taking into account parameter uncertainty in developing N recommendations.

3.3.4 Summary of Theoretical Results

The main conclusion from the theory above is that for the concave profit functions normally used in literature, specific recommendations and hence precision agriculture always stochastically dominates broad recommendations. This has been established in Proposition IV using Jensen's Inequality and first order spatial scale stochastic dominance. The application of this proposition is possible by following a Bayesian decision theoretic approach which accounts for parameter uncertainty and through MCMC allows a direct characterization of cumulative distribution functions for making stochastic dominance comparisons. This novel way of establishing the theoretical advantages of specific recommendations allows us to easily fit econometric models to directly test the theory. These theoretical results are however consistent with advocates of precision agriculture. The results are also consistent with theoretical results reported by Bullock *et al*. (2009) that used an information theoretic approach to determine utility and profit maximizing levels of nitrogen under variable rate technology and uniform rate technology. The main finding was that given any information structure, the farmer cannot lose (or in general will gain) from variable rate technology as compared to uniform rate technology because the problem with variable rate technology is less constrained than that with uniform rate technology. Before presenting the empirical strategy, it has to be pointed out that the question of whether to rely on a single estimate or confidence bounds or whole distribution in making optimal N recommendations has been under constant debate in the literature since the 1960s.[3] This debate can be interpreted as that of using Bayesian decision theory or plug-in-approach in economic optimization and the theoretical results in this paper illustrate that there is value to using the whole distribution of parameters.

[3] For more details, read the American Journal of Agricultural Economics (AJAE) commentaries in Anderson and Dillon (1968, 1970) and Seagraves (1970).

3.4 Review of Econometric Models and Specification

Table 2 summarizes the various econometric models that have been used in the literature as candidates for determining the optimal input recommendations. These models are categorized into single equation and system estimation models. The difference between the two is in the way the crop response function is estimated where for single equation models you have a single crop response averaged over all sample points considered while system estimation has site specific crop responses within a single model. There has been extensive use of the single equation models in the economics of crop response literature and less on system estimation. We provide a review of these models in the sections that follow.

Table 2: Specification of Empirical models in response function estimation

Category	Model[4]	Advantage	Weakness
Single N response coefficient models	*Pooled models with spatial fixed effects*	It's parsimonious	ignores individual within equations and correlation across equations heterogeneity, the spatial correlation
	Pooled models with spatial random effects	It's parsimonious	Ignores individual heterogeneity
	Geo-statistical Models)	In comparison with spatial error and lag models, the continuous covariance is arguably more suitable for variables such as crop yield whose variation in space is clearly continuous (Pringle, et al. 2010)	Computationally inefficient
	Spatial error and lag models	- computationally efficient - concerned with areal spatial data	Ignores individual heterogeneity within areal spatial units.
Heterogeneous N response coefficient models (system estimation)	*Quasi-spatial SUR*	Captures cross equation correlation Allows cross equation hypothesis tests	Ignores individual heterogeneity and spatial correlation within equation
	Spatial SUR	Incorporates spatial correlation within equation and error correlation between equations.	Ignores individual heterogeneity
	Spatial Cluster/Regionally Varying Coefficient Models:	Captures heterogeneity across the spatial clusters	Assumes homogeneity in response function within a spatial cluster
	Spatially Varying Coefficient Models	Incorporates all complex correlation structures and individual heterogeneity	Computationally inefficient but there are alternatives

The literature on the economics of precision agriculture has concentrated on the use of single equation ordinary least squares models, generalized least squares models, geo-statistical (Bullock

[4] Categorization into single N and heterogeneous N mostly reflects specification choices. For heterogeneous N models however, the natural specification of the model considers heterogeneity apriori.

and Lowenburg-DeBoer, 2006; Hurley, Malzer and Kilian, 2004) and spatial econometric models (Anselin, Bongiovanni and Lowenberg-DeBoer,2004). A comparative review of spatial econometric and geostatistical models for response function estimation can be found in Hurley, Malzer and Kilian (2004). In this paper, we focus on the system models.

3.4.1 Quasi-Spatial Seemingly Unrelated Regression Models

The Seemingly Unrelated Regression (SUR) estimation approach is proposed for the research questions posed in this study on the grounds that the error terms might be correlated across the locations i.e. the equations due to omission of variables. This is Zellner's original justification for considering a SUR framework. SUR is able to provide estimates of how relationships can potentially vary over the data dimensions as well as providing a convenient vehicle for testing hypotheses about these relationships (Fiebig 2003). According to Srivastava and Giles (1987), a common situation which may suggest a SUR specification is where regression equations explaining a certain economic activity in different geographical locations are to be estimated. The consideration for using a quasi-spatial seemingly unrelated regression model and a full spatial seemingly unrelated regression unlike a pooled model is that the pooled model does not account for spatial and temporal heterogeneity (Elhorst 2014). Spatial units are likely to differ in their background variables which are usually space specific, time invariant variables that do affect the dependent variables but which are difficult to measure. In the case of crop and soil fertility management recommendations, the spatial units may differ in rainfall, topography and other biophysical characteristics.

The application of SUR models explaining economic activities in different geographical locations have been rather sparse. However, typical examples include Munnell (1990). The important question pursued by Munnell was whether it is valid to assume that the coefficient vector is the same for all states (individuals) in the sample. Greene (2012) named a SUR based model that has data matrices that are group specific datasets on the same set of variables as a *"multivariate regression model."* In this paper, we use the term quasi-spatial SUR because it reflects that the specific equations represent spatial units though within spatial unit correlation is not explicitly analyzed. The quasi-spatial seemingly unrelated regression model consists of **M** equations which can be written as:

$$\begin{gathered} Y_1 = X_1 B_1 + \varepsilon_1 \\ X_2 = X_2 B_2 + \varepsilon_2 \\ X_3 = X_3 B_3 + \varepsilon_3 \\ \cdots \\ Y_M = X_M B_M + \varepsilon_M . \end{gathered} \tag{22}$$

This is a quasi-spatial model because M equations represent M spatial units (i.e. distinct and mutually exclusive districts or regions). In each spatial unit, there are N trials representing individual observations. The number of observations in each spatial unit does not necessary have to be equal. However, the computational requirements for the seemingly unrelated regression models with unequal number of observations were considered beyond the scope of this study. If we assume same number of observations and thus match the individual trials in each location, we can represent the quasi-spatial seemingly unrelated regression model as:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ 0 & 0 & \dots & X_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix} \quad \textbf{(23)}$$

$$Y_j = X_j B_j + \varepsilon_j\,, \quad j = 1, \dots, M$$

We assume strict exogeneity of X_j such that $E[\varepsilon|X_1, X_2, \dots, X_M] = 0$. Also, assume that the disturbances are uncorrelated across observations but correlated across equations, $E\left[\varepsilon_{ji}\varepsilon'_{ij'}\middle|X_1, X_2, \dots, X_M\right] = \sigma_{jj'}$, if t = s and 0 otherwise for all spatial clusters j and j'. This un-correlation assumption is unrealistic in the case of farm production as such spatial SUR is considered in the next section. However, proceeding with the SUR derivation, the disturbance formulation (i.e., non-spherical variance-covariance matrix under homoscedasticity) can therefore be represented as:

$$E\left[\varepsilon_j\varepsilon'_{j'}\middle|X_1, X_2, \dots, X_M\right] = \sigma_{jj'} I_T$$

$$E[\varepsilon\varepsilon'|X_1, X_2, \dots, X_M] = \Omega = \begin{bmatrix} \sigma_{11}I & \sigma_{12}I & \dots & \sigma_{1M}I \\ \sigma_{21}I & \sigma_{22}I & \dots & \sigma_{2M}I \\ \vdots & \vdots & & \vdots \\ \sigma_{M1}I & \sigma_{M2}I \dots & & \sigma_{MM}I \end{bmatrix}$$

$$\Omega = \begin{bmatrix} \sigma_{11} & \sigma_{12} \dots & & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2M} \\ \vdots & \vdots & & \vdots \\ \sigma_{M1} & \sigma_{M2} \dots & & \sigma_{MM} \end{bmatrix} \otimes I \quad \textbf{(24)}$$

$$\Omega = \Sigma \otimes I$$

$$\Omega^{-1} = \Sigma^{-1} \otimes I$$

According to Zellner (1962), the generalized least squares estimator is the most efficient estimator. The GLS estimator can be written as:

$$\dot{\beta} = [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}Y = [X'(\Sigma^{-1} \otimes I)X]^{-1}X'(\Sigma^{-1} \otimes I)Y \quad \textbf{(25)}$$

Since Ω is unknown, it is common to use its estimate $\dot{\Omega}$, the feasible generalized least squares estimator (FGLS).

Scaling Tests

According to Anselin (1988), there are two aspects of this SUR specification that need particular attention in a scaling problem. These are regional homogeneity and spatial aggregation and testing spatial dependence of an unspecified form. The hypothesis of particular interest in the SUR framework is the homogeneity restriction of equal coefficient vectors (Greene 2012). The regional homogeneity problem deals with coefficient stability across regions (e.g., to assess the extent to which all regions/districts in a spatial system respond to a certain application of fertilizer). Thus in a SUR framework the response function of each location can be modeled as a separate equation, related to the rest of the system by error covariance. A test of regional coefficient homogeneity can then be carried out as a hypothesis test on equality of parameters in the SUR model. If the joint null hypothesis of equality for all parameters in the model cannot be rejected, the data for all regions can reasonably be pooled.

The homogeneity restriction is that $\beta_i = \beta_m$ for i = 1,…,M - 1. The hypothesis would be tested as

$$RB = \begin{bmatrix} I & 0 & \dots & 0 & -I \\ 0 & I & \dots & 0 & -I \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & I & -I \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_M \end{bmatrix} = \begin{bmatrix} \beta_1 - \beta_M \\ \beta_2 - \beta_M \\ \dots \\ \beta_{M-1} - \beta_M \end{bmatrix} = 0. \quad \textbf{(26)}$$

This form of restriction states that the micro-units (individual smaller spatial units) are homogeneous insofar as their regression coefficient vectors are concerned. Further if valid, there is no aggregation bias such that the expectation of the macro/aggregated estimator (broad) will be equal to the micro (specific) parameter vector (Zellner, 1962). Finally, when the different equations in a SUR system pertain to regions, a test on spatial autocorrelation of an unspecified form is equivalent to a test on the diagonality of the inter-equation error covariance matrix.

3.4.2 Spatial Seemingly Unrelated Regression Models

Anselin (1988) developed and coined the term "spatial SUR" to a SUR specification containing regression equations that pertain to cross sections at different times or cross sections of cross sections. Most of the illustrations however concerned a spatial SUR in which the response coefficients varied over time rather than across space. In the context of this study, the spatial SUR is meant to exploit the spatial dependence of crop responses across spatial clusters. Thus, this would be considered a fully spatial SUR in that it considers both within and between spatial unit variations as well as having two spatial dimensions (i.e., cross sections of cross sections). In addition to the spatial dependences estimated via a spatial lag component and the weight matrix based spatial error component, one would study explicitly the error terms for each spatial cluster to determine if they are correlated across equations, implying a seemingly unrelated regression (SUR) structure. This kind of a framework was also used by Zhou and Kockelman (2009) in an employment study. Following Zhou and Kockelman (2009), a SUR model specification, with spatial lag and spatial error processes is a multi-equation extension of the equation:

$$y_m = \rho_m W_1 y_m + X_m B_m + \varepsilon_m; \varepsilon_m = \lambda_m W_2 \varepsilon_m + \mu_m. \quad \textbf{(27)}$$

According to Anselin(1988), when spatial units are grouped and estimation is carried out on a cross section in each group, plus the errors between the elements of the two groups show dependence, then a particular form of spatial SUR results. Thus, if the errors within each cluster equation show spatial dependence, a certain form of nested spatial effects is present, in the sense of within and between cluster dependence. The elements in each spatial cluster needed to be matched to those in the other, such that each is related to only one in the other cluster due to the assumption that $E\left[\varepsilon_{ri}\varepsilon_{sj}\right] = 0$ for $i \neq j$ for all clusters s and j. This is indeed a restrictive structure as Anselin (1988) noted and has very few applications.

In our framework, the coefficients can be constant across extension planning areas (EPAs) but vary over Agricultural Development Divisions (ADDs) for the first set of models and over districts for the other set of models. Since estimation is based on aggregate spatial units with arbitrary boundaries, spatial error autocorrelation could be present within each cross section. If the clusters are of unequal size, the simplifying Kronecker product results of the SUR model no longer apply, although the situation can still be considered as a case on non-spherical error variance (Anselin 1988).

If we assume a general case in which both intra- as well as inter-spatial cluster dependence is present, the model becomes a special case of a non-spherical error covariance matrix models (Anselin 1988). For instance, consider two spatial clusters/equations. A nested spatial error SUR can have error dependence with the following characteristics:

$$\begin{gathered}\varepsilon_m = \lambda_m W_2 \varepsilon_m + \mu_m \text{ for } m = 1,2 \\ E[\mu_{1s}\mu_{2t}] = \sigma_{12} \text{ for all } s, t.\end{gathered} \qquad \textbf{(28)}$$

According to Anselin(1988), this implies a constant variance between the errors of two spatial clusters for each pair of spatial units **s,** t.

The error covariance is therefore

$$\Omega = [I - (\lambda \otimes W)]^{-1} U [I - (\lambda \otimes W')]^{-1} \qquad \textbf{(29)}$$

with

$$U = \begin{bmatrix} \sigma_{11}.\, I & \sigma_{12}.\, E \\ \sigma_{21}.\, E & \sigma_{22}.\, I \end{bmatrix}$$

where E is T × T matrix of ones and I is an identity matrix of order T.

It was noted that using quasi-spatial seemingly unrelated regression model and the full spatial seemingly unrelated regressions can be satisfactory in the development of recommendations if unequal number of observations can be allowed for each spatial unit. The computational demands of this kind of analysis were too demanding for the study period. Thus, an alternative that is computationally lighter and allows unequal number of observations was considered. It can be established that the quasi-spatial SUR is equivalent to a GLS multivariate multi-level model while spatial SUR is equivalent to a GLS multivariate model with spatial correlation.

3.5 Alternatives to SUR and Spatial SUR

3.5.1 Region and Season Varying Coefficient Process

In many applications, the objective is to build regression models to explain a response variable over a region of interest under the assumption that the responses are spatially correlated. In nearly all of this work, the regression coefficients are assumed to be constant over the region. However, in some applications, coefficients are expected to vary at the local or sub-regional level (Gelfand *et al*., 2003). One important application is the determination of the appropriate scale of soil fertility recommendations since it is expected that input coefficients to maize production in each sub-region or sub-plot are different. For these types of problems, we can use a modeling approach called mixed effects modeling or hierarchical modeling. The justification for using this class of models is that a policy maker or agricultural researcher is not interested in a single plot but rather in a region or more generally population of plots, and thus the wider statistical problem is to optimize fertilizer amount for the population while also including the possibility of recommending different amounts for different plots in the population, as a function of some plot characteristics (Wallach 1995). These models are all estimated using the standard ordinary least squares (OLS) as such we presume all OLS assumptions (as discussed in Greene 2012) apply.

A region specific, varying coefficient model allows the parameter matrix to be region specific or spatially indexed. Consider a region specific, varying coefficient model

$$\mathrm{Y_{im}} = \beta_0 + \mathrm{B_{0m}} + \mathrm{B_1X}_{im} + \mathrm{B_2X_{im}^2} + \cdots + \epsilon_{\mathrm{im}} \tag{30}$$

where $\epsilon_{im} \sim N(0, \Lambda)$, m = 1,... M, Λ is the variance covariance matrix. This is a model where the regression coefficients depend on the location of the observation and thus can be seen as an application of the varying-coefficient models. The variation of the regression coefficients B through space enables greater adaptation by the model to changes unaccounted for by the covariates used in the model. It also means that there are many more parameters in the model. This poses a modeling challenge and additional elements are required to relate them (Gamerman and Moreira 2004).

By using different levels of spatial clustering to define the regions (i.e., Extension Planning Areas, districts/RDPs or Agricultural Development Divisions) we make comparisons of the models and ascertain whether further lower level spatial clusters are needed or whether broad regions are enough to explain the variation in the response functions.

Similarly a season specific coefficient model allows the parameter matrix to be season specific just as the region specific model above:

$$\mathrm{Y_t = \beta_0 + B_{0t} + B_1X_{1t} + B_1X_{2t}^2 + \cdots + \epsilon_t} \tag{31}$$

where $\epsilon_t \sim N(0, \Lambda)$, t = 1,... T.

Finally, we can consider a case where the coefficients vary by region and season:

$$\mathrm{Y_{rt} = \beta_0 + B_{0rt} + B_1X_{1rt} + B_1X_{2rt}^2 + \cdots + \epsilon_{rt}} \tag{32}$$

where $\epsilon_{rt} \sim N(0, \Lambda)$, r = 1,..., R and t = 1,...,T. This final model was not estimated due to data limitations during the implementation of the analysis.

3.5.2 Spatially Varying Coefficient Models via the Bayesian Modeling Approach

The modification to the generic geostatistical model is the incorporation of spatial random effects which can be set at different spatial clusters (e.g., Districts, Rural Development Programme or Agricultural Development Division) including the individual points. When we use spatial clusters or regions, as in the regionally varying coefficient models, we get into the same concerns about the arbitrariness of the scale of resolution, the lack of smoothness of the surface, and the inability to interpolate the value of the surface to individual locations (Banerjee, Carlin and Gelfand 2015). In a spatially varying coefficient model process, we aim to allow coefficients to vary by each point at which the trial was done thereby creating a spatial surface of response rates.

Before delving into the spatially varying coefficient model which is normally estimated using Spatial Bayesian Statistics, it is appropriate to explain why a Bayesian modelling framework is the most appropriate in the context of this study. For a start, the output we get from a Bayesian analysis is different from the point estimates that we seek to get with the "classical or frequentist" approaches. Instead of producing a point estimate of the parameters, a Bayesian analysis produces as its prime piece of output a density function for the parameters called "posterior" (Kennedy 1992). This has the advantage that we can be able to incorporate parameter uncertainty in the economic decision making by using the whole distribution of the parameters which is impossible with the frequentist methods. This is essentially the thrust of Bayesian decision theory which has been discussed before in Section 3.3.3.

The spatially varying coefficient model is a special class of spatial models that use Bayesian techniques for computational and numerical benefits. The spatially varying coefficient model can thus be written as (Cressie 1991; Gelfand *et al.*, 2003):

$$\mathrm{Y(s) = u(s) + W(s) + \delta(s), \delta(s) \sim G(0, \Sigma(\theta))} \tag{33}$$

where u(s) is the mean process, $u(s) = x(s)^T\beta$. The residual comprises a spatial process W(s), capturing spatial association and independent white noise process $\delta(s)$ which is often called the *nugget.* The white noise process has these assumptions: $E(\delta(s)) = 0$, $\mathrm{var}(\delta(s)) = \tau^2$ and $\mathrm{cov}(\delta(s), \delta(s')) = 0$. W(s) is a second order stationary mean 0 process independent of the white noise process: $E(W(s)) = 0$, $\mathrm{var}(W(s)) = \sigma^2$, $\mathrm{cov}(W(s), W(s')) = \sigma^2\, \rho(s, s'; \varphi)$, where ρ is a valid two-dimensional correlation function and φ is the decay parameter.

$W(s)$ captures spatial random effects and implicitly defines a hierarchical model providing local adjustment (with structured dependence) to the mean, interpreted as capturing the effect of unmeasured or unobserved covariates with spatial pattern. Letting $Y(s) = \beta_0 + \beta_1 x(s) + \beta_2 x(s)^2 + \beta_3 T + \beta_4 ST + \beta_5 V + \delta(s)$, write $W(s) = \beta_0(s)$ and define $\tilde{\beta}_0(s) = \beta_0 + \beta_0(s)$. Here, $x(s)$ is amount of Nitrogen, T is season, ST is soil type and V is variety. Then $\beta_0(s)$ can be interpreted as a random spatial adjustment at location s to the overall intercept β_0. Equivalently, $\tilde{\beta}_0(s)$ can be viewed as the random intercept process.

Following Banerjee, Carlin and Gelfand (2015) for an observed set of locations $s_1, s_2, \dots, s_n$, given $\beta_0, \beta_{1,\dots,}, \{\beta_0(s_i)\}$ and τ^2, $Y(s_i)$ for i=1,…, n, are conditionally independent. During the implementation of this modeling framework, it however turned out that the data were too large for this process to produce results. We thus considered an alternative model that preserves the advantages of the spatially varying coefficient process but can efficiently reduce the computational burden. Banerjee, Carlin and Gelfand (2015) proposed the predictive process models in which the *N* spatial random effects $W(s)$ where *n* is the total number of observations are replaced with *M* spatial random effects $\overline{W}(s)$ with $m < n$ as a way of dealing with large spatial datasets. The locations used for reduced model are called knots. The predictive process model was empirically implemented instead of the full spatial varying coefficient process model. The model then becomes

$$Y(s) = u(s) + \overline{W}(s) + \delta(s).$$

The analysis was conducted using spBayes package (Finley & Banerjee, 2013) in R programming environment. We estimated the model using Markov Chain Monte Carlo in which Gibbs sampling updates were used to estimate the regression equation parameters β. The remaining parameters were updated using blocked random walk Metropolis steps using multivariate normal proposals. This process was made possible by first obtaining ordinary least squares (OLS) residuals which were then used to estimate a Gaussian empirical semi-variogram. We obtained estimates of the parameters (partial sill, σ^2; nugget, τ^2; and decay parameter, φ) from this process to be used as starting values. We used K means approach to knot selection and selected 200 knots. The results were invariant to other knot selection algorithms and numbers of knots.

The modified predictive process model was then estimated by specifying (i) the knots, (ii) model type: we use the spatially varying univariate model and bias adjusted predictive process,

(iii) starting values, (iv) tuning values, (v) number of iterations, and (vi) priors. The analysis in this paper uses a flat or a non-informative prior which is already embedded in the spBayes R Package for the coefficient parameters β. Therefore, we specified the priors for the hyper parameters only (nugget and sill). The priors included a uniform distribution (3/1000,3/50) for the φ, inverse gamma (2,0.08) for the σ^2 and inverse gamma (2,0.02) for the τ^2. These hyper priors have been better at allowing faster convergence in most of previous analyses using this model (e.g. Banerjee *et al*., 2015; Finley *et al*., 2011). Trace plots were used to assess convergence. Two MCMC chains were run for 1000 iterations each at different initial values and each started mixing at 200'th iteration. We then discarded the initial 250 as burn in and recovered the remaining posterior samples in the subsequent analysis. The standard Deviance Information Criterion was used to choose the best model.

Computational Issues

While some of the computational challenges are explained for each respective model, this section provides a summary of the challenges that were met. The simple OLS and GLS models were easily implemented in R. Implementing the SUR and Spatial SUR failed because of two issues. Firstly, models that involve unequal number of observations are not yet fully implementable in the conventional statistical packages. The geostatistical models via the frequentist approaches were not converging in large sparse datasets. We thus considered Bayesian geostatistical approaches since the computations were more efficient. In addition, this modeling approach provided a direct link with Bayesian decision theory which is ideal for decision making under parameter uncertainty.

4.0 Results and Discussion

4.1 Descriptive Statistics and Yield Maps

This section provides the results for the exploratory analysis of yield variations against the different covariates that were considered in the experiments. Figure 4 shows the distribution of experimental maize yields for each of the treatments. It is apparent from the density plots that the zero N treatment had the lowest mean yield which was expected considering that nitrogen fertilization is considered yield increasing. The structuring of treatments may suggest that Nitrogen and Phosphorus effects may be confounding. However, since Nitrogen is regarded as the most limiting macro-nutrient, all the interpretation is on the differing levels of Nitrogen.

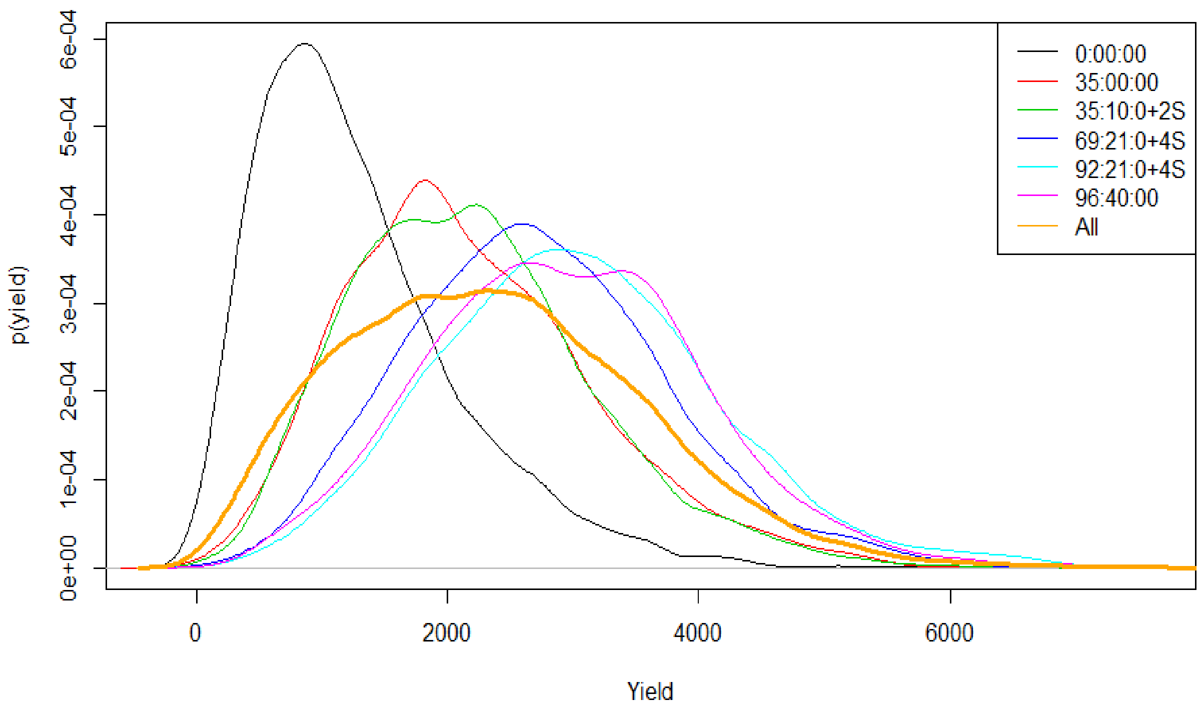


Figure 4 : Maize yields density plots by for each treatment

The important aspect in the density plot is however the differences in the variation and kurtosis of yields across the different treatments. The variability observed can be attributed to many factors including location, weather, and topography. The focus for this study is on spatial aspects of the variation. In order to explore the inherent heterogeneity of the soils across the country; maize yields with the no fertilizer treatment were mapped. It was expected that with zero fertilizer, the yields would not vary much across the country. Figure 5 shows that there was variation in yield even without fertilizer application signifying the inherent spatial heterogeneity. The control treatment (no fertilizer) was the least performing treatment in almost all locations. It is also evident that while there are wide differences between 35N treatment and 69N treatment; there are few differences between 69N, 92N and 96N treatments (Figure 6).

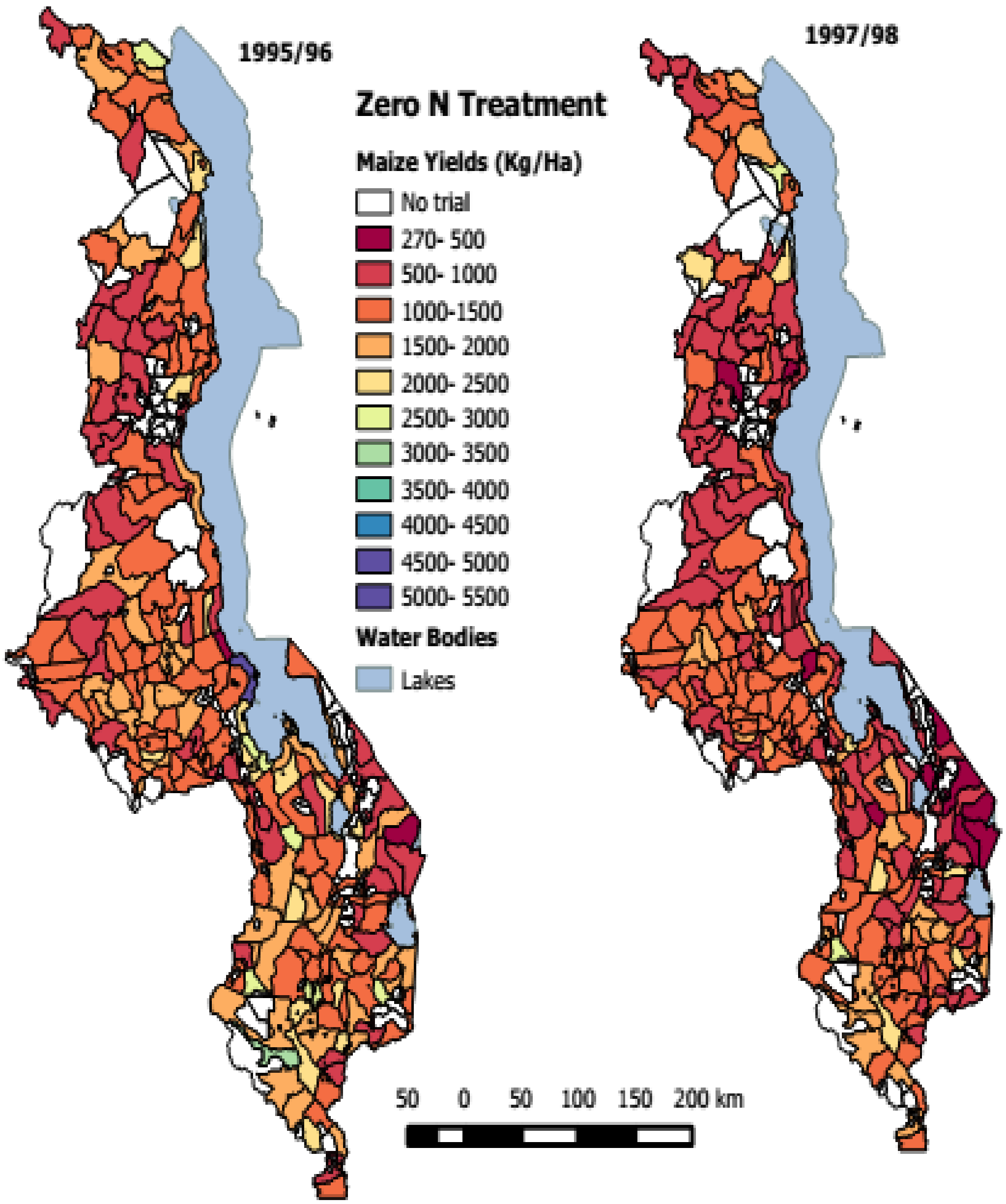


Figure 5: Zero N treatment Yields across the country

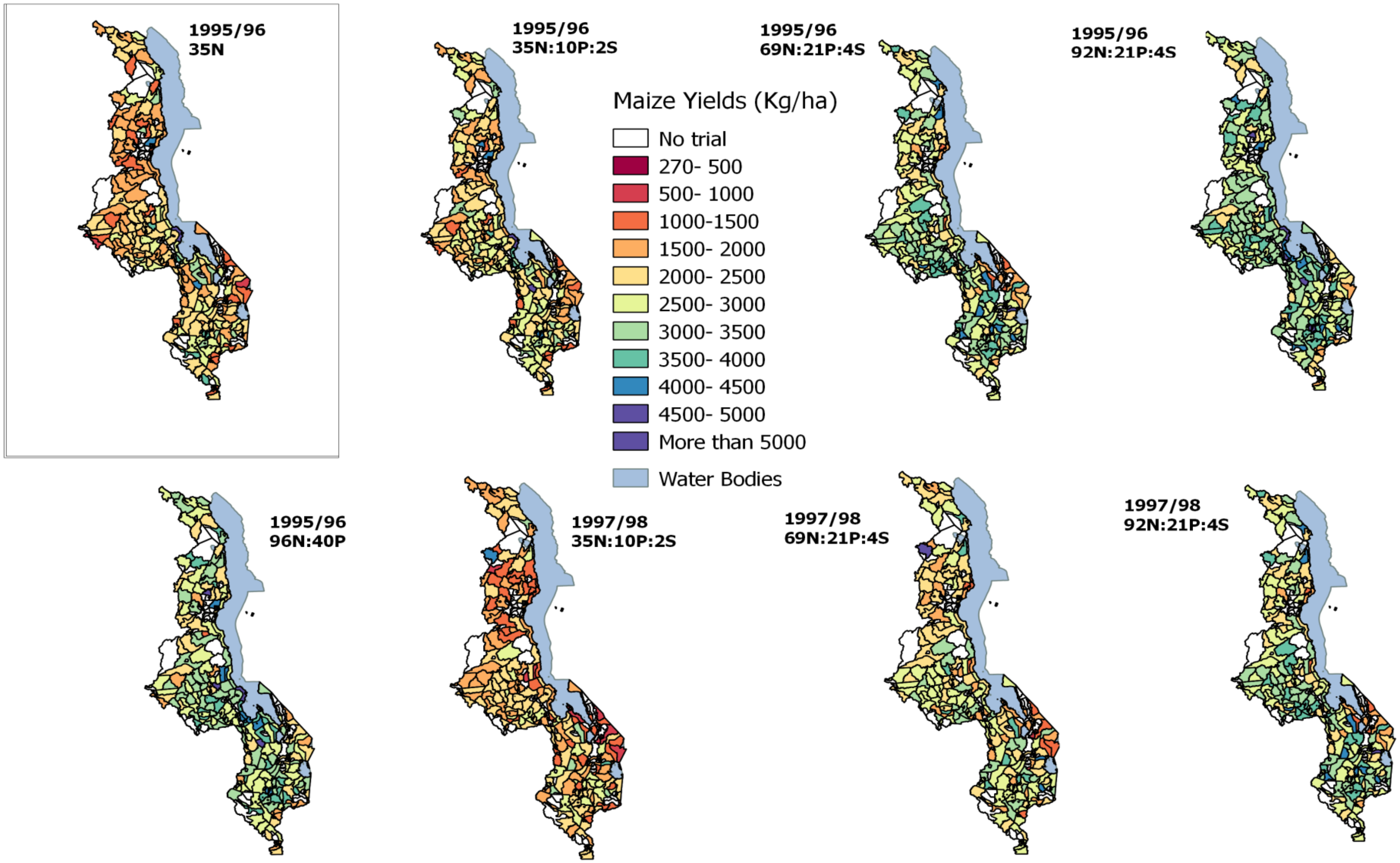


Figure 6: Yield maps for all others treatments except 0N

4.2 Fixed N Response Coefficient Models

4.2.1 ADD Fixed and Random Effects Models

It is evident from the results in Table 3 that location as defined by ADD, season (1 = 1997/8), maize variety (MH17, MH18 and Composite), soil texture (medium and light), nitrogen (N) and nitrogen squared (NSquared) are important factors affecting the yield of maize. These account for about 31 percent of total variation.

Table 3: Fixed N Response Coefficient Models

	OLS	GLS with FE	GLS with RE
(Intercept)	1236.00***	1236.00***	1147.72***
	(118.68)	(118.68)	(125.00)
add KADD	-84.05**	-84.05**	
	(32.61)	(32.61)	
addKRADD	-71.84	-71.84	
	(36.96)	(36.96)	
addLADD	29.64	29.64	
	(30.38)	(30.38)	
addMADD	-268.66***	-268.66***	
	(28.59)	(28.59)	
addMZADD	-286.25***	-286.25***	
	(31.34)	(31.34)	
addSLADD	37.86	37.86	
	(35.19)	(35.19)	
addSVADD	-55.51	-55.51	
	(33.70)	(33.70)	
Season (1997/98=1)	-255.67***	-255.67***	-255.64***
	(16.41)	(16.41)	(16.41)
Variety-MH17	50.38	50.38	49.49
	(116.87)	(116.87)	(116.85)
Variety-MH18	245.16*	245.16*	245.78*
	(116.28)	(116.28)	(116.27)
Soil texture(Medium=1)	152.35***	152.35***	152.86***
	(15.87)	(15.87)	(15.86)
N	**29.40***	**29.40***	**29.40***
	(0.84)	(0.84)	(0.84)
N Squared	**-0.12***	**-0.12***	**-0.12***
	(0.01)	(0.01)	(0.01)
R^2	0.31		
Num. obs.	15690	15690	15690
AIC		260609.91	260687.59
BIC		260724.81	260756.53
Log Likelihood		-130289.95	-130334.79
Num. groups			8

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

4.2.2 RDP Fixed Effects Models

When the results are disaggregated by RDPs or Districts, it is evident that location is still an important factor to yield but variety is not. This is so because of the nature of the trial in which the varieties were not randomized but rather targeted to specific locations. The table in appendix A shows the parameter estimates of a RDP fixed effects model. When the fixed effects and random effects models were compared, the fixed effects model was performing better in terms of having smaller Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values.

4.3 Regionally and Seasonally Varying Slope and Intercept Models

In this section, a specific class of models that consider varying the N response coefficient across locations and seasons are presented and discussed. These models are of importance in the determination of the optimal scale of making research recommendations because it is easier to make crop location or spatial level comparisons.

4.3.1 ADD Varying Slope and Intercept

Figure 7 provides ADD-specific OLS models. It is apparent that Shire Valley ADD (SVADD) provides quite a different N coefficient than all the other ADDs. Though the other N coefficients are not exactly the same, they are within overlapping confidence intervals. This result is counterintuitive to the agro-ecological interpretation of the delineation of the ADDs. It is however a logical result in the sense that ADDs are huge and there is no evidence that they were developed on the basis of agro-ecologies. Each of the ADDs has varying topology and climatic conditions; such that one would expect an averaging out of high and low responsive sites within each of the ADDs. This will be investigated by evaluating the response rates at a lower level: the Rural Development Programme (RDP) level. It is however important to note that the intercepts show marked differences across the ADDs illustrating that while the ADDs are not significantly different from each other in N response, they are still inherently different.

Table 11 in the appendices provides a more detailed ADD varying intercept and slope coefficient model. The model presentation in the table allows us to make formal tests of hypothesis which are impossible using the separate regressions presented in Figure 7.

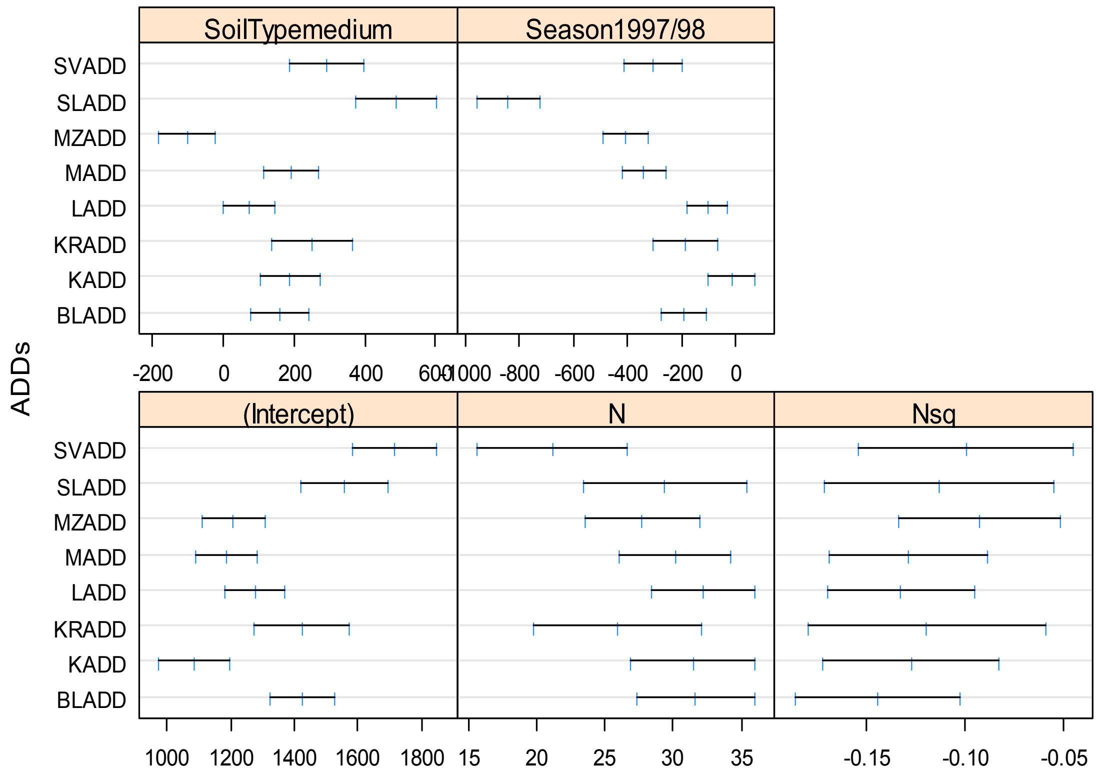


Figure 7: Separate ADD Specific Crop Response Functions

Linear Hypothesis Tests

When a single response model is disaggregated across the ADDs, it is possible to test linear hypotheses regarding the equality of N response across ADDs which substantially helps determine whether having a single N response relationship is optimal. Linear hypothesis tests of equality of each of the agricultural development N coefficients to the N coefficient for the pooled model showed that only Shire Valley ADD has a significantly different N coefficient. This therefore implies that using this level of spatial clustering in comparison to the pooled model, one would expect that only a single region would be given a markedly wrong recommendation in statistical sense. However, this would be under the assumption that crop response functions within each of the ADDs are homogenous. In order to substantiate this assumption, we consider a lower level of spatial aggregation, i.e. Rural Development Programs (RDP) or alternatively Districts.

4.3.2 RDP Intercept and Slope Varying Coefficient Model

When we consider an agronomic policy defined at RDP or district level, subtle aspects of the response functions start to emerge. Figure 8 shows the model estimates and confidence intervals for separate RDP equation models. When RDP is considered the spatial level of making the agronomic recommendations, all the RDPs are responsive to N application just as it was found in the case of ADDs. Obviously, the RDPs in Shire Valley ADD have N response functions significantly different from those of the other RDPs (Table 12). It is also apparent that the quadratic N parameter shows no significance in some of the RDPs unlike in the ADD intercept and slope varying coefficient model in Table 11. As indicated for the ADD varying model, the confidence intervals for the intercepts of the RDP models are not overlapping for most locations implying that while the fertilizer response may be different; these areas are markedly different in terms of their yield potential.

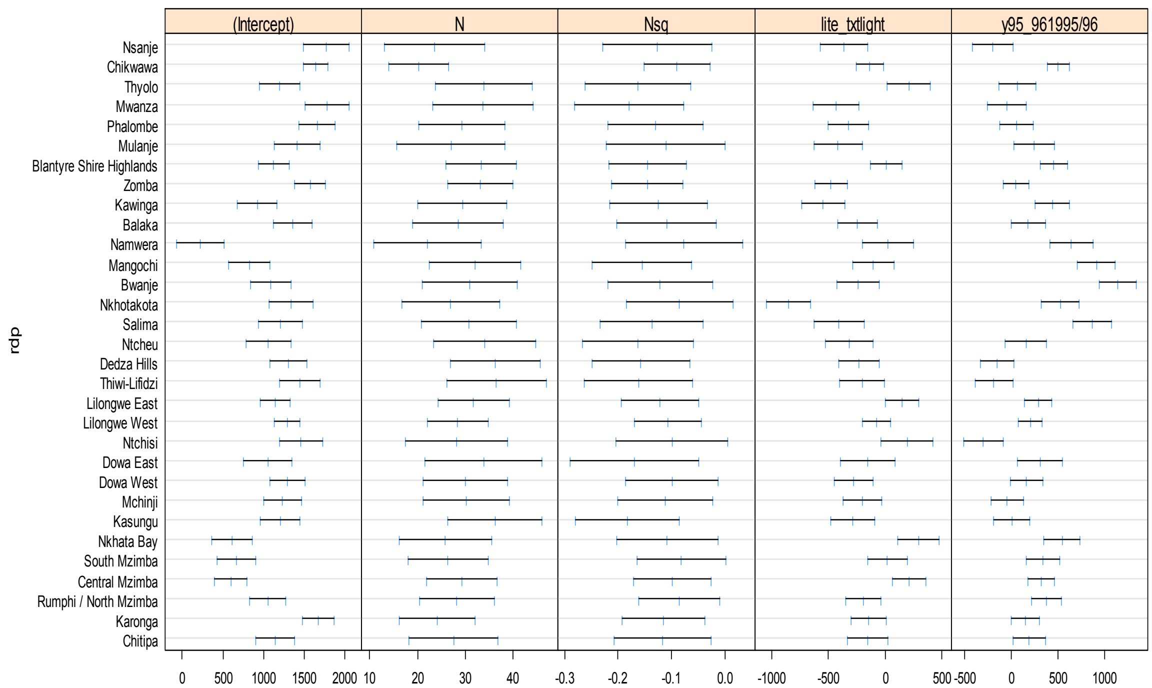


Figure 8: Separate RDP Specific Crop Response Functions

One other important feature when comparing Figure 7 and Figure 8 is the width of the confidence intervals i.e. standard errors. By modeling at lower spatial levels, we are doing this at the expense of sample size and precision.

4.3.3 EPA Varying Intercept and Slope Coefficient Model

We further considered the case in which the N fertilizer recommendations are made at the EPA level. There were 151 EPAs in the data. Because of space, this table is not appended to this paper. However, various descriptive statistics were computed to determine how making recommendations at this spatial level would differ from making recommendations at RDP and ADD levels. Table 4 shows the summaries computed from the model results after fitting an EPA varying intercept and slope coefficient model.

Table 4: Summary Statistics for EPA Intercept and Slope Coefficient Model

		Significance at 5% level		Total
Variable	**Statistic**	**Insignificant**	**Significant**	
N	Percentage number of EPAs	11.26	88.74	100.00
	Average N coefficient	18.81	31.39	29.97
	Standard Deviation of the N coefficients	7.00	6.40	7.58
	Average Standard Error	11.93	9.97	10.19
N Squared	Percentage number of EPAs	88.00	12.00	100.00
	Average N coefficient	-0.11	-0.22	-0.12
	Standard Deviation of the N coefficients	0.05	0.05	0.06
	Average Standard Error	0.10	0.09	0.10

If making broad recommendations is not different from making recommendations at any other level below it, it would be expected that the descriptive statistics in

Table 4 would reflect the results in the models presented previously. However, it is evident that the average standard error is much higher when the response functions are evaluated at EPA level. This basically means that there is still a lot of variation within the individual trials in an EPA. Of course, this reflects the sample size problem. Nonetheless, it also implies that while N may be responsive when the data is pooled across the country, this may not be the case when each individual EPA is considered separately. This has research and experimental design implications particularly since locations are normally used as replicates in most studies. Is it better to have multiple replications of the experiment within one EPA (or other spatial unit)?

4.3.4 Seasonally Varying Intercept and Slope Process

Since the trials were done in two seasons, it is also of importance to determine whether N response functions are different across seasons. Table 5 illustrates the N response functions for the two seasons. The results for the two seasons are different with 1995/96 which was a good year with a higher N response. A linear hypothesis test of equality between N response in 1995/96 (N96) and N response in 1997 (N97) was rejected with a p-value of 0.0096. The seasonality of N response is an important finding as it affects the nature of decisions that can be made from static or single period agronomic trials. These findings are consistent with Liu, Swinton, and Miller (2006) who also found that the maize yield functions were not consistent across time thereby concluding that while crop yield response to applied fertilizer may be site specific in some years, it is neither consistently site specific across years nor is the effect of individual variables consistent across years.

Table 5: Seasonally Varying Coefficient Models[5]

	OLS	GLS with FE	GLS with RE
(Intercept)	1370.11***	1370.11***	1353.30***
	(39.66)	(39.66)	(52.70)
MZADD	-214.40***	-214.40***	
	(35.90)	(35.90)	
KADD	-12.20	-12.20	
	(36.89)	(36.89)	
LADD	101.48**	101.48**	
	(34.89)	(34.89)	
SLADD	109.70**	109.70**	
	(41.91)	(41.91)	
MADD	-196.82***	-196.82***	
	(35.93)	(35.93)	
BLADD	71.84	71.84	
	(36.94)	(36.94)	
SVADD	16.34	16.34	
	(40.66)	(40.66)	
Variety-MH18	194.78***	194.78***	196.28***
	(20.63)	(20.63)	(20.46)
Variety-Composite	-50.38	-50.38	-49.50
	(116.80)	(116.80)	(116.79)
Soil Texture (Light=1)	-152.35***	-152.35***	-152.86***
	(15.86)	(15.86)	(15.85)
Year (1997=1)	-261.00***	-261.00***	-260.97***
	(34.78)	(34.78)	(34.78)
N_1997	**26.09***	**26.09***	**26.09***
	(1.37)	(1.37)	(1.37)
Nsq_1997	**-0.08***	**-0.08***	**-0.08***
	(0.01)	(0.01)	(0.01)
N_1996	**30.59***	**30.59***	**30.59***
	(1.09)	(1.09)	(1.09)
Nsq_1996	**-0.14***	**-0.14***	**-0.14***
	(0.01)	(0.01)	(0.01)
R^2	0.31		
Adj. R^2	0.31		
Num. obs.	15690	15690	15690
AIC		260599.40	260677.07
BIC		260729.61	260761.34
Log Likelihood		-130282.70	-130327.54
Num. groups			8

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

[5] GLS with FE=Generalized Least Squares with Fixed Effects, GLS with RE=Generalized Least Squares with Random Effects.

4.4 Spatial Predictive Process Model

While the models above demonstrate how broad agronomic recommendations can be substantially different from "statistically optimal" lower spatial level recommendations, they implicitly still assume homogeneity of N response within the spatial units at that level. This assumption is potentially wrong as literature shows that soil fertility may vary in very small distances. In order to determine whether this hypothesis holds, we develop spatially varying coefficient models that consider each independent trial as a realization of a spatial process.

In the results that follow, we considered a flat prior or non-informative prior by using priors that have large variances in order to appeal to those who hold frequentist ideals to statistical estimation while also getting a better interpretation of the results. The posterior estimates are therefore in principle similar to the conventional spatially varying geo-statistical model. However, the goal of the analysis is to obtain estimates with associated measure of uncertainty as a continuous surface over the domain (Banerjee, Carlin & Gelfand 2015).

It is important to note that we are now not only dealing with individual trials but rather the whole spatial process within Malawi therefore we consider a surface plot other than just a yield map. It is evident from the Figure 9 that there is spatial variation in the yields across country with yields ranging from about 1000kg/ha to 4000kg/ha.

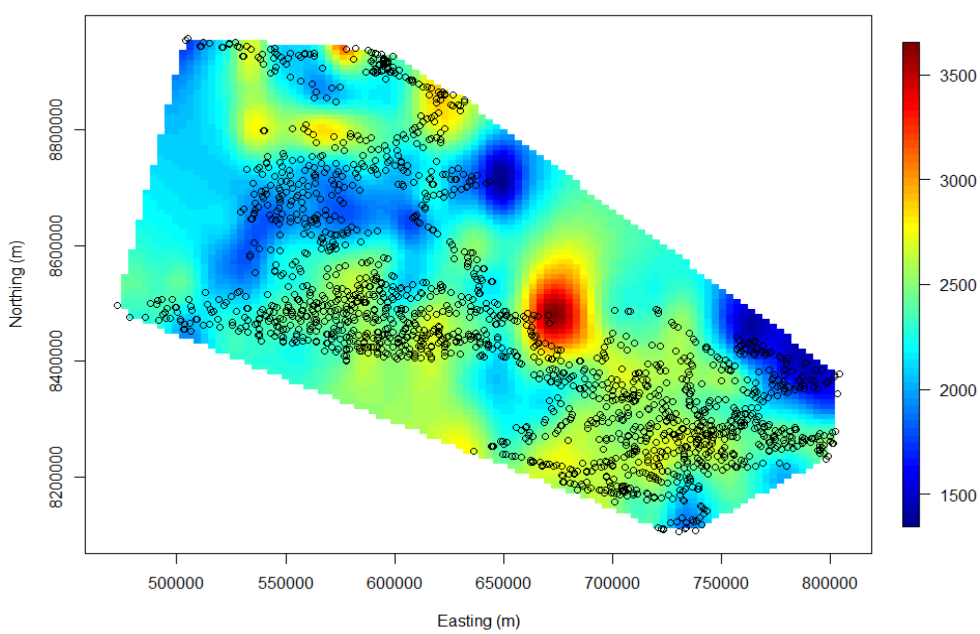


Figure 9: Surface plot of Maize Yields

The predictive spatial process model requires setting initial and tuning parameter values that are obtained from empirical semi-variogram. Figure 10 shows the semi-variogram of the yields. It is apparent in the figure that yields vary quite a lot across locations. It is also apparent that maximum variation is reached at about 50km though the semi-variance is as high as 1,000,000 in the limit of distance to the positive side of zero. Figure 11 shows the knot selection for the case with 200 knots using the different algorithms for choosing the location of the nodes. It is apparent from the figure that the k- means approach is over the entire space of the country thereby being favorable for the subsequent analyses.

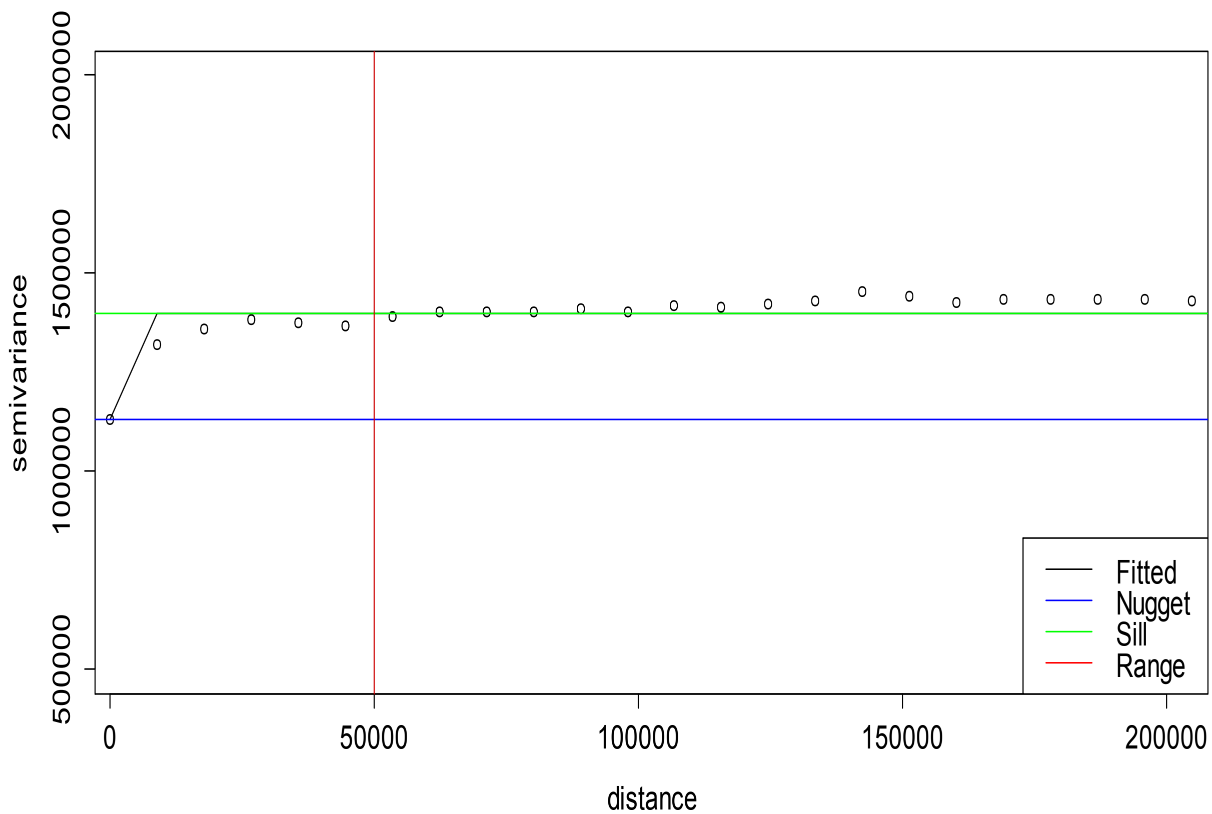


Figure 10:The semi-variogram plot of experimental maize yields

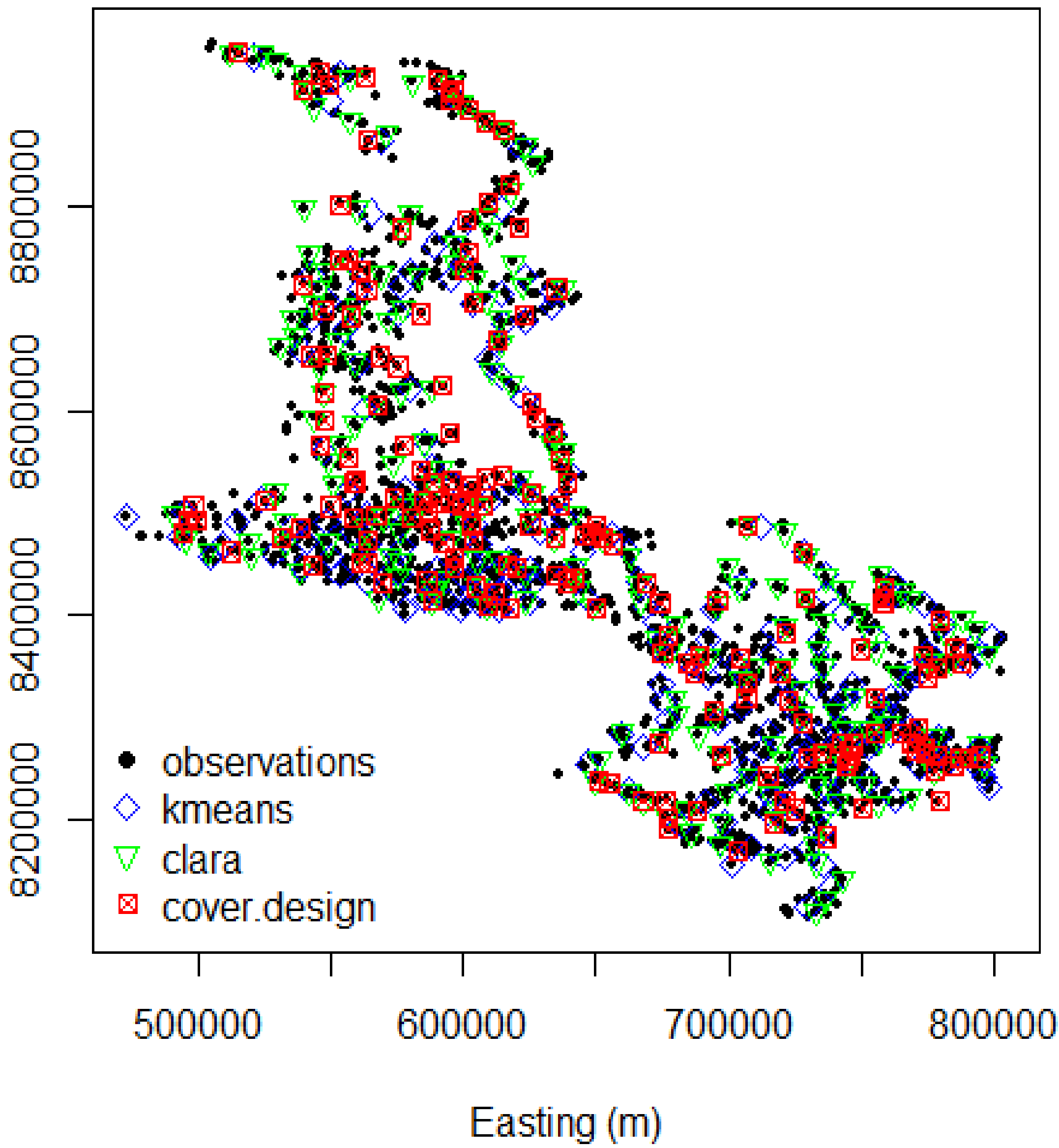


Figure 11: Map of Malawi showing the Knot selection algorithms and the points selected

4.4.1 Knots Sensitivity Analysis and Convergence

The Figure 12 illustrates the posterior predictive distributions for the different numbers of knots. It is apparent that the yield distributions are almost identical which indicates that for this dataset using knots does not affect the results that much.

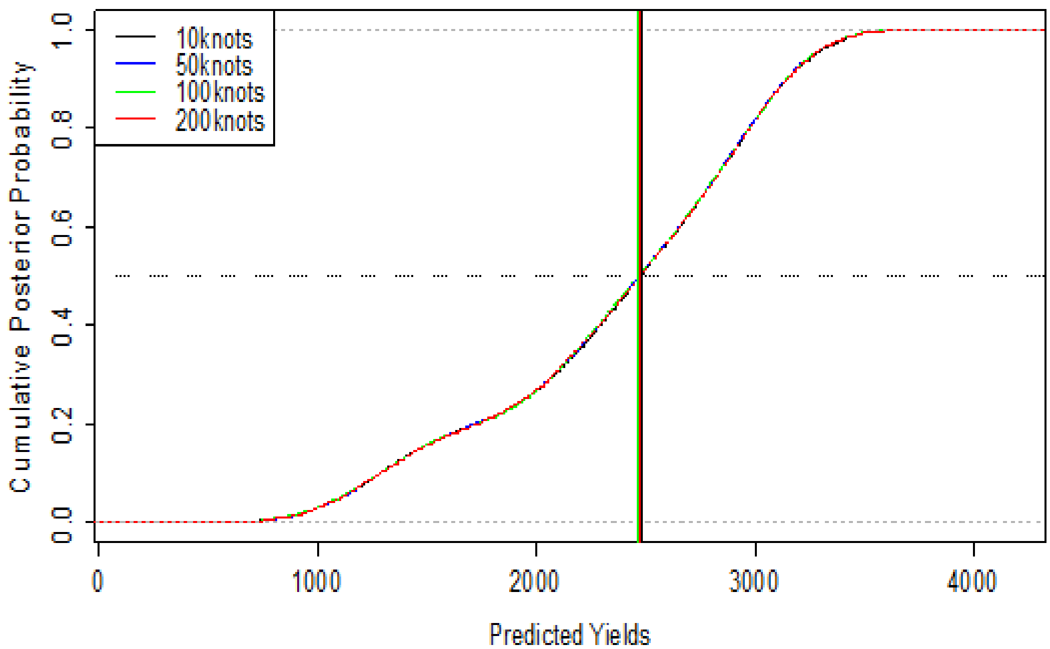


Figure 12: Cumulative posterior distributions of predicted yields at different numbers of knots

In terms of convergence, the general and most conventional approach to checking convergence is to see how well the chains with different starting values are mixing or moving around the parameter space. We can visually see this for each parameter using trace plots. The appendices show the trace plots for three of the parameters using two chains with different starting values. The two chains started to mix after 200 iterations such that 250 of the samples was burn in and the remaining 750 were used in subsequent inferences.

4.4.2 Posterior parameter estimates

Table 6 and Table 7 show posterior summaries and the quintiles of the posterior estimates of each of the parameters respectively. The results are quite similar to the results presented for other frequentist based models presented in preceding sections. It is apparent that the spatial variance (sigma.sq) and non-spatial variance (tau.sq) equally affect the results of the model.

Table 6: Posterior summaries at 200 knots

	Mean	SD	Naive SE
(Intercept)	1123.00	121.30	3.13
N	29.36	0.87	0.02
N squared	-0.12	0.01	0.00
Variety-MH17	42.30	120.30	3.10
Variety-MH18	253.20	120.70	3.12
Soil texture (Medium=1)	178.90	16.22	0.42
Season (1997/98=1)	-253.10	16.80	0.43
sigma.sq	489200.00	489400.00	12630.00
tau.sq	488500.00	488700.00	12610.00
Phi	0.04	0.02	0.00

SD=Standard Deviation, SE= Standard Error

Table 7: Posterior quintiles at 200 knots

	2.50%	25%	50%	75%	97.50%
(Intercept)	892.80	1040.37	1121.00	1206.00	1369.00
N	27.72	28.76	29.33	29.96	31.05
N squared	-0.14	-0.13	-0.12	-0.12	-0.11
Variety-MH17	-202.60	-38.69	43.96	125.60	273.50
Variety-MH18	4.50	172.69	256.30	337.30	484.00
Soil texture (Medium=1)	147.30	168.55	179.00	189.60	211.80
Season (1997/98=1)	-285.80	-263.94	-253.90	-241.80	-220.40
sigma.sq	0.17	0.94	471500.00	975900.00	999800.00
tau.sq	0.01	0.09	474400.00	975700.00	992200.00
Phi	0.01	0.02	0.03	0.05	0.06

Table 8 shows N coefficients varying by regions. The credibility intervals for N coefficient for SVADD (15.12 to 26.48) do not overlap with those for BLADD (27.45 to 35.90) and LADD (28.24 to 35.72). This implies that using the same recommendation would result in suboptimal decision rules. The model in Table 8 was the best model using the Deviance Information Criterion (DIC).

Table 8: Hierarchical Model with 200 knots

	2.50%	25%	50%	75%	97.50%
(Intercept)	1350.00	1429.00	1474.00	1515.00	1597.00
Season (1995/96)	222.80	242.80	253.90	264.60	285.50
MH18 Variety	157.80	182.60	196.30	209.70	235.40
Composite Variety	-286.60	-124.10	-47.94	27.35	156.50
Soil texture (Light=1)	-182.80	-162.30	-151.30	-140.40	-120.80
addidKRADD	-360.50	-242.60	-185.90	-126.00	-12.01
NKRADD	19.95	23.86	25.84	28.06	31.93
NsqKRADD	-0.18	-0.14	-0.12	-0.10	-0.06
addidMZADD	-755.00	-659.30	-607.20	-556.50	-468.30
NMZADD	23.61	26.43	27.85	29.36	32.16
NsqMZADD	-0.13	-0.11	-0.09	-0.08	-0.05
addidKADD	-594.00	-495.80	-440.70	-394.70	-298.90
NKADD	26.23	29.53	31.03	32.52	35.28
NsqKADD	-0.17	-0.14	-0.13	-0.11	-0.08
addidLADD	-497.10	-404.50	-354.80	-307.10	-212.20
NLADD	28.24	30.69	32.06	33.36	35.72
NsqLADD	-0.17	-0.14	-0.13	-0.12	-0.10
addidSLADD	-492.00	-377.40	-318.00	-254.90	-146.60
NSLADD	24.25	27.93	30.01	32.25	36.44
NsqSLADD	-0.17	-0.14	-0.11	-0.09	-0.06
addidMADD	-724.80	-631.50	-578.80	-528.40	-434.20
NMADD	26.30	29.00	30.43	31.73	34.52
NsqMADD	-0.17	-0.14	-0.13	-0.12	-0.09
addidBLADD	-464.70	-367.80	-315.30	-263.20	-164.50
NBLADD	27.45	30.17	31.57	32.99	35.90
NsqBLADD	-0.19	-0.16	-0.14	-0.13	-0.10
NSVADD	15.12	19.21	21.22	23.01	26.48
NsqSVADD	-0.15	-0.12	-0.10	-0.08	-0.04
sigma.sq	939200.00	949100.00	955200.00	962500.00	973400.00
tau.sq	0.69	1.48	2.94	5.39	10.35
Phi (decay parameter)	0.00	0.00	0.01	0.02	0.04

4.5 Economic Optimization and Spatially Specific Recommendations

The economic question we ask is similar to that which economists concerned with precision agriculture ask. Whether the benefits of accuracy in the response functions are huge enough to make a profit difference with the conventional broad estimations? By incorporating price ratio information into the response function estimates, we are able to decipher the economic optimum for N application. In addition, by changing the objective function to be that of yield maximization we can find the optimal agronomic yields. We find the aggregation bias and thus the (dis)advantage of broad recommendations by comparing the optimal results from using a broad optimal level to a spatially disaggregated model. In carrying out the analysis, a range of prices, as reported in different studies in Malawi from 1970s to 2010, were considered (Table 9). We used a 2.35 fertilizer to maize price ratio in all the subsequent analyses unless stated otherwise.

Table 9: Average, minimum and maximum prices of maize and fertilizer

Year	Variable	Mean	Standard Deviation	Min	Max
2009/10 (Kwacha/Kg)	Maize Prices	43.39	6.68	32.67	63.28
(Integrated Household Survey, 2009)	DAP Prices	106.19	10.89	90.00	153.00
	NPK Prices	102.32	8.45	80.00	120.00
	Urea Prices	97.52	5.28	89.47	110.00
	Average Fertilizer Prices	102.01	8.21	86.49	127.67
	Average N prices	34.20	2.19	29.78	39.10
	Fertilizer: Maize Price Ratio	**2.35**	**1.23**	**2.65**	**2.02**
	N:Maize Price Ratio	**0.79**	**0.33**	**0.91**	**0.62**
1997 Analysis	Urea Prices	6.84			
	Maize Producer Price	1.94			
	Maize Consumer Price	3.12			
	Urea: Maize Producer Price Ratio	3.53			
	Urea: Maize Consumer Price Ratio	2.19			
Heisey and Mwangi (1997)	Nitrogen: Maize Price Ratio 1977-87 1988-1994	Median 10.7 7.7			
Meertens 2005	Value-Cost Ratio				
	Early 1980s	7.4			
	Mid 1990s	3.3			
	Early 2000s	1.3			

4.5.1 Spatial Profit Maximization with the Plug-in Approach

The strategy used in the study was to determine the optimal set of input and output prices for the pooled or general model. Then insert this into the yield response equation to get predicted yields that are then inserted into the objective function to get the profits for each individual trial. Note that this entails imposing a broad or average optimal N rate to all individual trials or locations. In order to make a comparison with the case where different N levels are recommended, we simply take the recommended N levels and insert it into the spatial disaggregated model. The difference in predicted yields and profits show the (dis)advantage of the disaggregated model (i.e., with regionally specific N recommendations) and thus the potential benefits or losses to spatially varying recommendations. Figure 13 illustrates the cumulative probability step functions for the different regions.

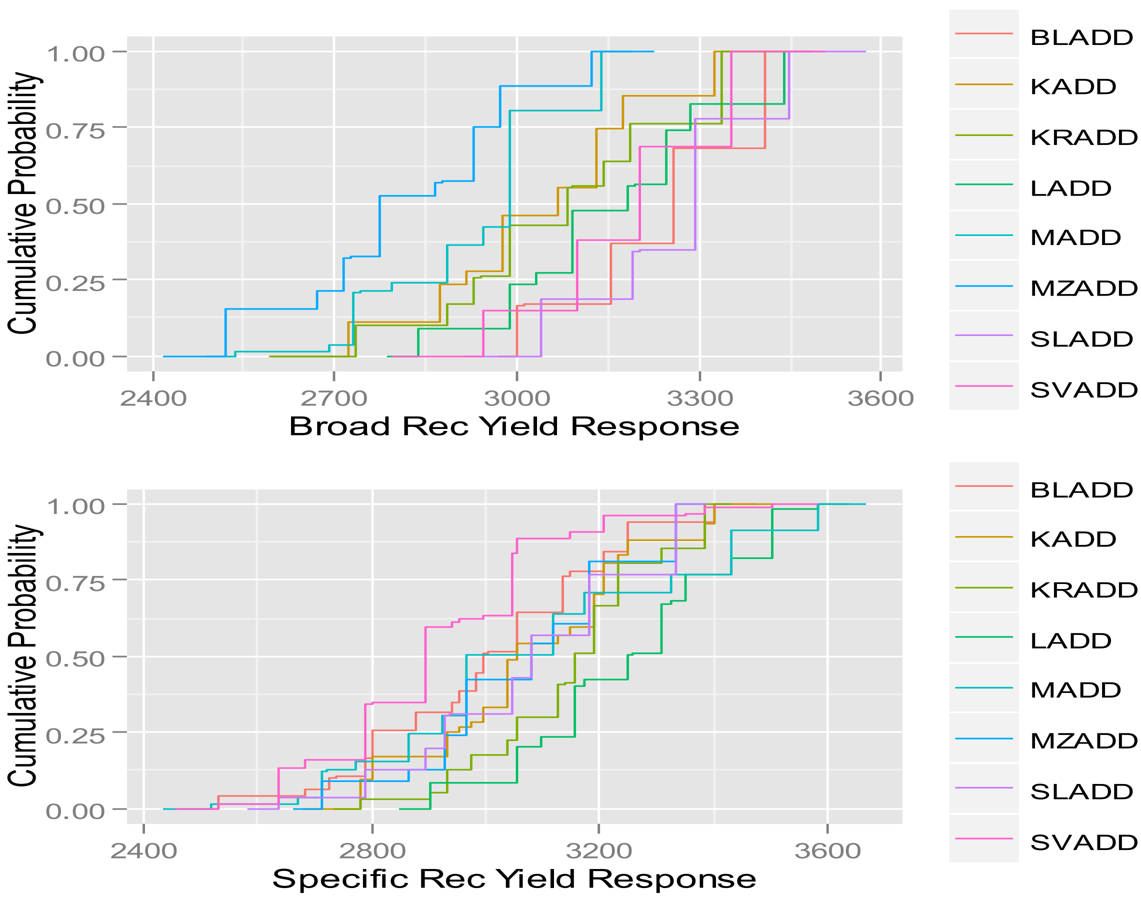


Figure 13: Cumulative distribution of profits for broad vs. specific recommendations

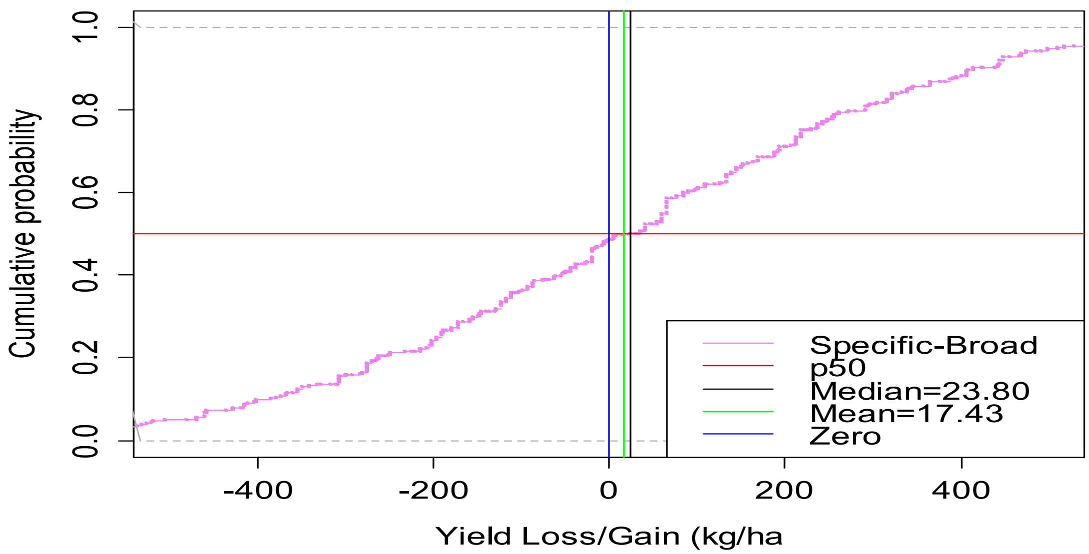
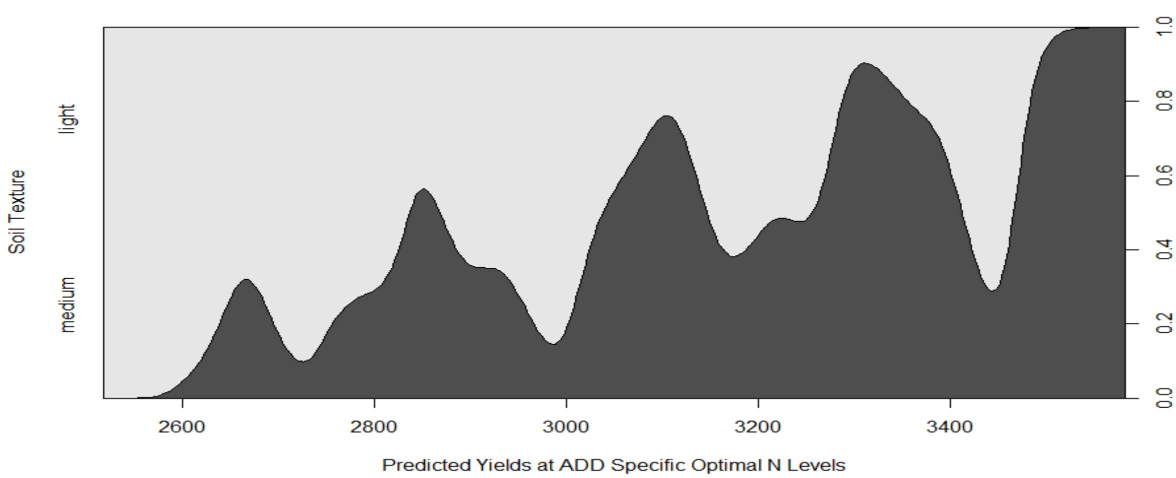


Figure 14: Overall yield loss/gain to specific recommendations

The terraces in the cumulative distribution functions are a result of categorical variables that were included in the model. These included: soil texture, maize variety, and season. Using the optimal regional specific recommendations, we can deduce the predicted yields at these optimal levels. The conditional density plot[6] shows how soil texture is distributed over the predicted yields at the optimum rate. It is apparent from the figure that medium textured soils were at the optimum higher yielding than light textured (coarse textured) soils.



[6] Conditional density plot shows the conditional densities of the random variable given levels of a categorical variable weighted by the marginal distribution of the categorical variable. The densities are derived cumulatively over the levels of the categorical variable.

Figure 15: Conditional density plot on the conditional distribution of soil texture over predicted yields at ADD Specific Optimal N Levels

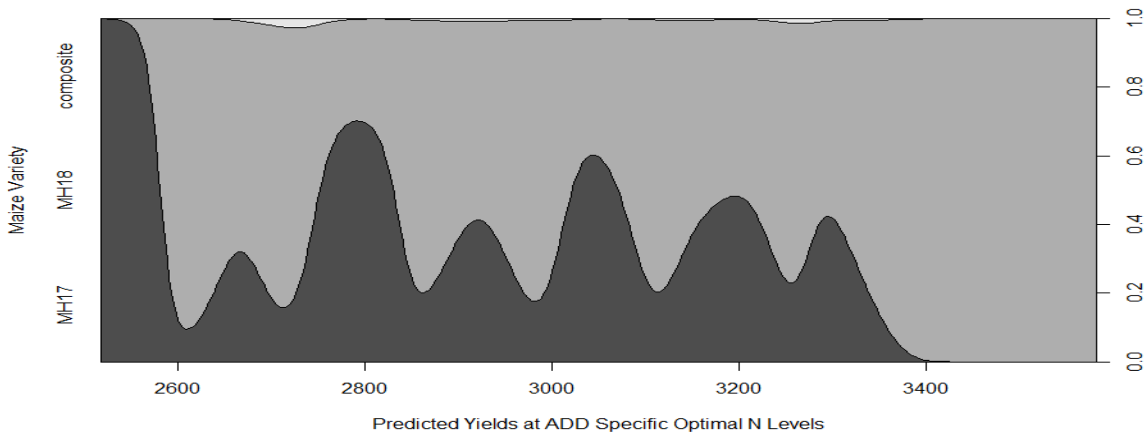


Figure 16: Conditional density plot on the conditional distribution of maize variety over predicted yields at ADD Specific Optimal N Levels

In terms of maize variety, Figure 16 shows that MH 17 had both extremes of high yields and low yields at the optimum. It is apparent that in general MH18 had the highest yields.

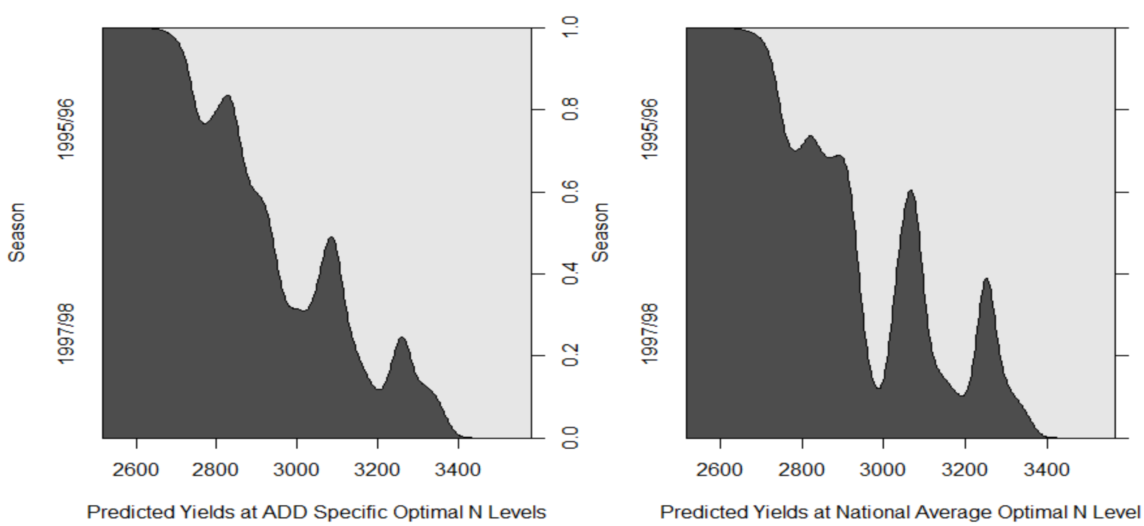


Figure 17: Seasonality and Predicted Yields at National and ADD Optimal N Levels

It is apparent in the Figure 17 that seasonality had an effect on the yields at the optimal N levels. For both broad and specific recommendations, 1995/96 had the highest yields than 1997/98.

4.5.2 Spatial Profit Maximization with Bayesian Methods

When using the Bayesian approach, we are integrating the profit function under the whole distribution using Markov Chain integration. Thus, the equation in the Bayesian model uses the economically optimal N rates at each sample location. In the analysis, we simply used the results from the MCMC chains to make conclusions as to whether broad or specific recommendations are better.

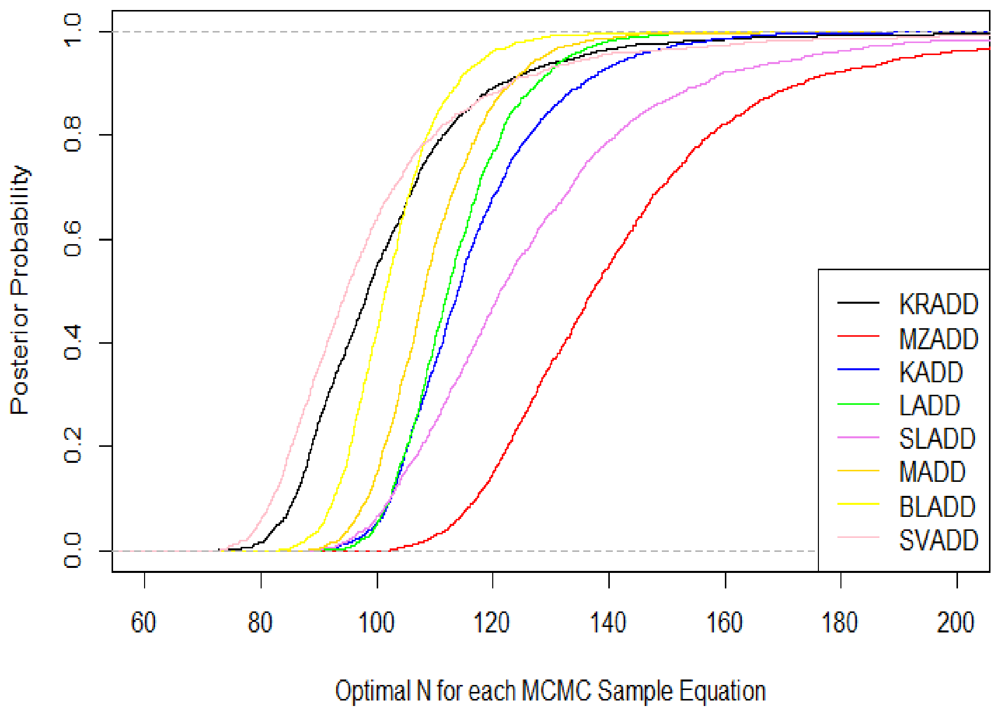


Figure 18: Optimal N for each MCMC Sample equation by ADD

It is evident from Figure 18 that the optimal specific N rates are different across regions. In addition, the figure shows that there is always a distribution of optimal N rates that make it may be misleading to simply use a single recommendation for any of the regions. Figure 19 compares the expected profits from applying an optimal broad recommendation with estimation risk accounted for against average expected profits from applying ADD specific recommendations with estimation risk accounted for.

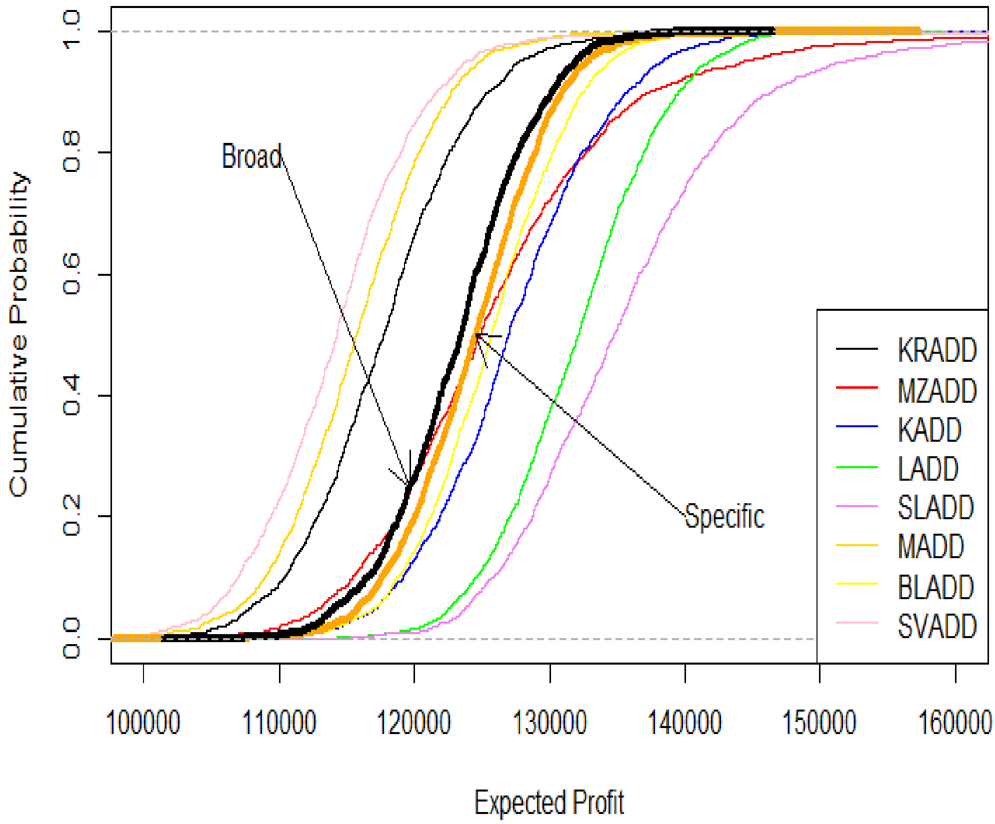


Figure 19: Posterior probability distribution of expected profits

The key result from the figure is that specific recommendations stochastically dominate broad recommendations. This is consistent with proposition IV which says that specific recommendations will always stochastically dominate broad recommendations for all increasing profit functions. It's also possible to make stochastic dominance interpretations about the differences of profits attained in some of the regions. For example, profits in SLADD clearly stochastically dominate profits in KRADD. This comparison of profit distributions across space illuminates interesting aspects in understanding the parameter uncertainty. As it can be seen, while considering only the coefficient on the N may suggest that the optimal levels are similar across regions, cumulative distribution functions above demonstrate that this interpretation may be misleading. The profit distributions depend on the stochastic magnitudes of the other variables included in the model including soil type and season. These results are consistent with the suggestion that precision agriculture/specific recommendations may have risk management potential (Lowenberg-DeBoer, 1999).

Optimal N Levels at Varying Prices

The Figure 20 shows the results of varying the price ratio of nitrogen to maize output. This is done by holding the maize output as numeraire and varying the prices of nitrogen upwards and downwards.

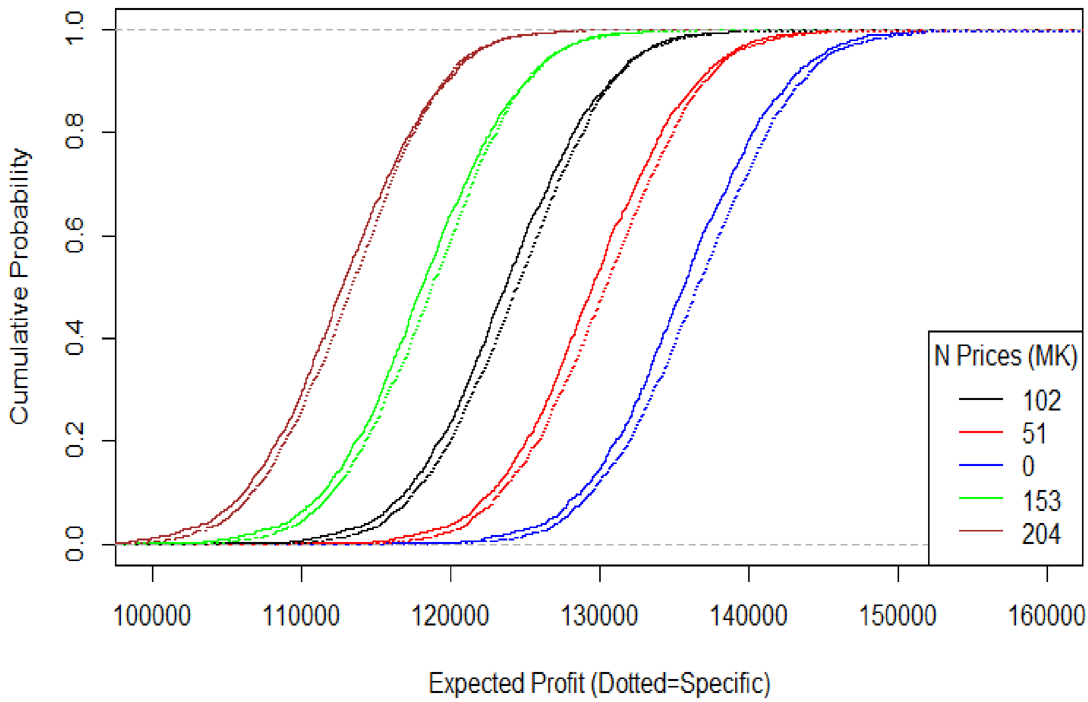


Figure 20: Posterior probability distribution of expected profits at different prices.

It is apparent from price ratio comparisons that while the optimal levels of nitrogen and profits are dependent on the price ratios; the decision on the spatial scale to make recommendations is independent of price ratio once we assumed the same price ratio across locations. This is consistent with the invariance property of stochastic dominance which says that stochastic dominance is preserved under multiplication with a constant or addition of another independent random variable. This conclusion applies in this context because we have assumed risk neutrality. An extension of the analysis with non-neutral risk preferences would be an important part of further research.

5.0 Conclusion

Though the agricultural economics literature is full of research papers on the economics of precision agriculture and site-specific crop response functions, the theory has been lacking. In most cases, it has been found that variable rate technology and other precision agriculture technologies are more profitable, yet only one paper, Bullock *et al*. (2009), to the knowledge of the author provided a theory on the value of precision agriculture and why this is normally the case. This study however develops an easier and direct approach to empirically testing the value of using specific recommendations.

In summary, this paper has analytically and numerically demonstrated that profits from using specific recommendations will always either be the same or first order stochastically dominate profits from using broad recommendations when ignoring the quasi-fixed cost differentials. This result has been proven directly by applying the theorem of first order spatial scale stochastic dominance and Jensen's Inequality assuming the profit function is concave. The paper has also shown that it is almost impossible in agriculture to achieve perfect aggregation of economically optimal nitrogen rates which implies that any broad recommendation cannot exactly reproduce specific recommendations unless the locations concerned are homogenous in all parameters. This logically means that specific recommendations will strictly stochastically dominate broad recommendations. In terms of agricultural research policy, these findings illustrate the importance of collecting better information on costs of generating specific recommendations and in defining a lower bound cost that can be provided together with the recommendations so that farmers are aware of the comparative benefits of using regional/national recommendations or searching for locally specific recommendations. Finally, it is apparent that agricultural researchers should first identify the key factors that drive yield responses in each particular location concerned and then determine the optimal level based on enough sample in that location other than relying on an optimal broad recommendation.

In practical situations, a specific recommendation would be adopted if the ex-ante costs of generating the specific recommendations are small enough not to offset the gains in profitability. This thesis did not attempt to quantify the costs of such information. This area of research remains of practical value to economists working on the value of precision agriculture. The likely avenue for that research is the calculation of lower bound costs that would offset these dominance results. Another research avenue would be using multi-location and multi-season data to exploit the spatio-temporal heterogeneity which inevitably adds a layer of complexity. The major weakness of the model presented is that it assumes exogenous input and maize output prices. This assumption is potentially problematic in the context of Malawi. Therefore, models that consider endogenous demand and supply relations in determining soil fertility recommendations are required. In our model, we also assumed a risk-neutral producer. Further research is needed for the case where farmers are assumed to be risk averse.

Bibliography

Anderson, Jock, and John Dillon. "Economic considerations in response research." *American Journal of Agricultural Economics*, 1968: 130-142.

Anselin, L. *Spatial econometrics: methods and models.* Dordrecht, Netherlands: Kluwer Academic, 1988.

Anselin, L, and R J Florax. "Small sample properties of tests for spatial dependence in regression models: some further results." In *New Directions in Spatial Econometrics*, by L Anselin and R J Florax, 21-74. Berlin: Springer-Verlag, 1995.

Anselin, Luc. *Exploring spatial data with with GeoDa: a workbook.* Urbana: Center for Spatially Integrated Social Science, 2005.

—. *Spatial econometrics: methods and models.* Dordrecht: Kluwer Academic Publishers, 1988.

Anselin, Luc, R Bongiovanni, and J Lowenberg-DeBoer. "A spatial econometric approach to the economics of site-specific nitrogen management in corn production." *American Journal of Agricultural Economics* 86 (2004): 675-687.

Antle, John M, and Jetse J Stoorvogel. "Incorporating systems dynamics and spatial heterogeneity in integrated assessment of agricultural production systems." *Environment and Development Economics*, 2006: 39-58.

Bagnoli, Mark, and Ted Bergstrom . "Log-concave probability and its applications." *Economic Theory*, 2004: 1-25.

Banerjee, Sudipto, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data.* New York: CRC Press, 2015.

Berger, James O. *Statistical decision theory and bayesian analysis.* New York: Springer, 1985.

Brorsen, Wade B. "Using Bayesian Estimation and Decision Theory to Determine the Optimal Level of Nitrogen in Cotton." *Southern Agriculltural Economics Association Annual Meeting.* Florida: Southern Agriculltural Economics Association, 2013. 1-13.

Bullock, David S, and James Lowenburg-DeBoer. "Using spatial analysis to study the values of variate rate technology and information." *International Association of Agricultural Economists Conference.* Gold Coast, Australia: International Association of Agricultural Economists, 2006. 1-41.

Bullock, David S, Jess Lowenberg-DeBoer, and Scott M Swinton. "Adding value to spatially managed inputs by understanding site-specific yield response." *Agricultural Economics* 27 (2002): 233-245.

Bullock, David S, Matias L Ruffo, Donald G Bullock, and German D Bollero. "The value of variable rate technology: an information theoretic approach." *American Journal of Agricultural Economics* 91, no. 1 (2009): 209-223.

Byerlee, D R, and J R Anderson. "Value of predictors of uncontrolled factors in response functions." *Australian Journal of Agricultural Economics*, 1969: 118-127.

Byerlee, Derek, and Jock R Anderson. "Risk, utility and the value of information." *Review of Marketing and Agricultural Economics* 50, no. 3 (1982): 231-246.

Cressie, Noel A.C. *Statistics for spatial data.* John Wiley & Sons, 1991.

Diggle, Peter J, and Paulo J Ribeiro. *Model-based geostatistics.* New York: Springer, 2007.

Dillon, John L, and Jock R Anderson. *The analysis of response in crop and livestock production.* Pergamon Press, 1990.

Dorfman, Jeffrey H. *Bayesian economics through numerical methods: a guide to econometrics and decision-making through prior information.* New York: Springer, 1997.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. "How high are rates of return to fertilizer? Evidence from field experiments in Kenya." *American Economic Reviews: Papers and Proceedings* 98, no. 2 (2008): 482-488.

Elhorst, Paul J. *Spatial econometrics: from cross-sectional data to spatial panels.* London: Springer , 2014.

Fiebig, Denzil G. "Seemingly unrelated regression." In *A companion to theoretical econometrics*, by Badi H Baltagi, 101-116. Blackwell Publishing, 2003.

Gamerman, Dani, and Ajax R.B Moreira. "Multivariate spatial regression models." *Journal of Multivariate Analysis* 91 (2004): 262-281.

Gelfand, A E, S E Hills, A Racine-Poon, and A F.M Smith. "Illustration of bayesian inference on normal data models using gibbs sampling." *Journal of the American Statistical Association* 85 (1990): 972-985.

Gelfand, Alan E, Hyon-Jung Kim, C F Sirmans, and Sudipto Banerjee. "Spatial modeling with spatially varying coefficient processes." *Journal of the American Statistical Association* 98, no. 462 (2003): 387-396.

Getis, Arthur, and Keith J Ord. "The analysis of spatial association by use of distance statistics." In *Perspectives on Spatial Data Analysis*, by Luc Anselin and Sergion J Rey, 127-145. New York: Springer, 2010.

Giller, K.E, et al. "Communicating complexity: Integrated assessment of trade-offs concerning soil fertility management within African farming systems to support innovation and devlopment." *Agricultural Systems*, 2010.

Goverment of Malawi;. *National Fertilizer Strategy.* Lilongwe: Government of Malawi, 2009.

Government of Malawi. *Fertilizer verification trials 1995/96: Trial Implementation Manual for Section Field Assistants.* Action Group I- Maize Productivity Taskforce, Ministry of Agriculture and Livestock Development, Lilongwe: Government of Malawi, 1995.

Government of Malawi. *Guide to Agricultural Production in Malawi.* Lilongwe: Ministry of Agriculture and Food Security, 2012.

Government of Malawi. *Maize fertilizer recommendations demonstration 1997/98.* Action Group I- Maize Productivity Taskforce, Ministry of Agriculture and Livestock Development , Lilongwe: Government of Malawi, 1997.

Government of Malawi. *The 1995/96 Fertilization verification trial-Malawi: Economic analysis of results for policy discussion.* Action Group I-Maize Productivity Taskforce, Ministry of Agriculture and Livestock Development, Lilongwe: Government of Malawi , 1997.

Greene, William H. *Econometric analysis.* Boston: Prentice Hall, 2012.

Griffin, T W. "The spatial analysis of yield data." In *Geostatistical applications for precision agriculture* , by M A Oliver. London: Springer, 2010.

Griffin, Terry W. *Decision-making from on-farm experiments: spatial analysis of precision agriculture data.* PhD dissertation, Department of Economics, Purdue University, Indiana: Purdue University, 2006.

Haining, Robert. *Spatial data analysis.* Cambridge University Press, 2004.

Hernandez, Jose A, and David J Mulla. "Estimating uncertainty of economically optimum fertilizer rates." *Agronomy Journal* 100, no. 5 (2008): 1221-1229.

Hurley, T M, G L Malzer, and B Kilian. "Estimating site-specific nitrogen crop response functions: a conceptual framework and geostatistical model." *Agronomy Journal* (American Society of Agronomy) 96 (2004): 1331-1343.

Hurley, Terrance M, Kikuo Oishi, and Gary L Malzer. *Estimating the potential value of variable rate nitrogen applications: a comparison of spatial econometric and geostatistical.* 2005.

Hyman, Glen, Dave Hodson, and Peter Jones. "Spatial analysis to support geographical targeting of genotypes to environments." *Frontiers in Physiology*, 2013.

Jauregul, Mario A, and Gustavo E Sain. *Continuous economic analysis of crop response to fertilizer in on-farm research.* CIMMYT Economics Paper No. 3, Mexico: CIMMYT, 1992.

Jaynes, Dan B. "Confidence bands for measured economically optima nitrogen rates." *Precision agriculture* 12 (2011): 196-213.

Johnson, Devin S. *Bayesian inference for geostatistical regression models.* Fairbanks: Department of Mathematics and Statistics, University of Alaska, 2005.

Johnson, Michael, and Regina Birner. *Understanding the Role of Research in the Evolution of Fertilizer Policies in Malawi.* IFPRI Discussion Paper 01266, Washington D.C: IFPRI, 2013.

Kaleijian, H H, and I R Prucha. "A generalized moments estimator for the autoregressive parameter in a spatial model." *International Economic Review* 40 (1999): 509-533.

Kennedy, Peter. *A guide to econometrics.* Oxford: MIT Press, 1992.

King, Robert P, Donald W Lybecker, Anita Regmi, and Scott M Swinton. "Bioeconomic models of crop production systems:design, development, and use." *Review of Agricultural Economics* 15, no. 2 (1993): 389-401.

Kyriakidis, Phaedon C, and Nicholas N Nagle. "Geostatistical regression for areal data." *Accuracy 2010 Symposium.* Leiceister, 2010. 293-296.

Lambert, Dayton M, and James Lowenberg-DeBoer. "Spatial regression models for yield monitor data: a case study from Argentina." *American Agricultural Economics Association Annual Meeting.* Montreal, Canada: American Agricultural Economics Association, 2003. 1-30.

Lesage, James, and Kelley R Pace. *Introduction to Spatial Econometrics.* New York: Chapman & Hall/CRC, 2009.

Levy, Haim. *Stochastic dominance: investment decision making under uncertainty.* New York: Springer, 2006.

Liu, Yanyan, Scott M Swinton, and Neil R Miller. "Is site-specific yield response consistent overtime? Does it pay?" *American Journal of Agricultural Economics* (American Agricultural Economics Association) 88, no. 2 (2006): 471-483.

Lowenberg-DeBoer, J., 1999. Risk management potential of precision farming technologies. *Journal of Agricultural and Applied Economics,* 31(2), pp. 275-285.

Makumba, W. *Nitrogen use efficiency and carbon sequestration in legume tree based agroforestry systems: a case study in Malawi.* Wageningen, Netherlands: Wageningen University, 2003.

Mann, Charles. *Higher yields for all smallholders through "best bet" technology: the surest way to restart economic growth in Malawi.* Network Research Results Working Paper 3, Network Research, 1998.

Mas-Colell, Andreu, Michael D Whinston, and Jerry R Green. *Microeconomic Theory.* Oxford University Press, 1995.

Munnell, Alicia H. "Why has productivity growth declined? Productivity and public investment." *New England Economic Review*, 1990: 3-22.

Muus, Lars, Hiek van der Scheer, and Tom Wansbeek. "A decision theoretic framework for profit maximization in direct marketing." In *Econometric Models in Marketing*, by P H Franses and A L Montgomery, 119-140. Amsterdam: Elsevier Science Ltd, 2002.

Oishi, Kikuo. *Does precision agriculture pay? Profitability potential of variable rate application of fertilizers, econometrics and geostatistical approaches.* PhD Dissertation, St. Paul: University of Minnesota, 2006.

Ojayi, O C, F Place, F K Akinnifesi, and G W Sileshi. "Agricultural success from Africa: the case of fertilizer tree systems in Southern Africa (Malawi, Mozambique, Zambia and Zimbabwe." *International Journal of Agricultural Sustainability* 9, no. 1 (2011).

Openshaw, Stan. "The Modifiable Areal Unit Problem." *Concepts and Techniques in Modern Geography 38*, 1983: 3-40.

Ord, Keith J. "Spatial autocorrelation: a stastician's reflections." In *Perspectives on Spatial Data Analysis*, by Luc Anselin and Sergio J Rey, 165-180. New York: Springer, 2010.

Ouedraogo, Frederic B, and Wade B Brorsen. "Bayesian estimation of optimal nitrogen rates with a non-normally distributed stochastic plateau function." *Southern Agricultural Economics Association* . Dallas: Southern Agricultural Economics Association , 2014.

Pringle, M J, T F.A Bishop, R M Lark, B M Whelan, and A B McBratney. "The analysis of spatial experiments." In *Geostatistical applications for precision agriculture*, by M A Oliver, 244-267. Springer, 2010.

Rosenberg, Lloyd. "Nonnormality of linear combinations of normally distributed random variables." *The American Mathematical Monthly*, 1965: 888-890.

Schabenberger, Oliver, and Francis J Pierce. *Contemporary statistical models for the plant and soil sciences.* London: CRC Press, 2002.

Seagraves, J A. "Economic considerations in response research." *American Journal of Agricultural Economics* 52, no. 4 (1970): 607-609.

Shanmugalingam, S. "On the analysis of the ratio of two correlated normal variables." *Journal of the Royal Statistical Society, Series D (The Statistician)* 31 (1982): 251-258.

Sheahan, Megan, Roy Black, and Thom S Jayne. *What is the scope for increased fertilizer use in Kenya?* Working Paper, Michigan: Michigan State University, 2012.

Smale, Melinda, Derek Byerlee, and Thom Jayne. *Maize Revolutions in Sub-Saharan Africa.* Development Research Group, World Bank, Washington DC: World Bank, 2011.

Snapp, S S, M J Blackie, and C Donovan. "Realigning research and extension to focus on farmers' constraints and opportunities." *Food Policy* 28 (2003): 349-363.

Srivastava, Virendra K, and David E.A Giles. *Seemingly unrelated regression equations models: estimation and inference.* Marcel Dekker, 1987.

Suri, Tavneet. "Selection and comparative advantage in technology adoption." *Econometrica* 79, no. 1 (January 2011): 159–209.

Tchale, Hardwick. *Agricultural policy and soil fertility management in the Maize based smallholder farming system in Malawi.* Berlin: Peter Lang, 2006.

Tembo, Gelson, Wade B Brorsen, Francis M Epplin, and Emilio Tostao. "Crop input response functions with stochastic plateaus." *American Journal of Agricultural Economics* 90, no. 2 (2008): 424-434.

Theil, Henri. *Linear aggregation of economic relations.* Amsterdam: North-Holland Publishing Company, 1954.

—. *Linear aggregation of economic relations.* Amsterdam: North-Holland Publishing Company, 1955.

Tumusiine, Emmanuel, Wade B Brorsen, Jadeesh Mosali, Jim Johnson, James Locke, and Jon T Biermacher. "Determining optimal levels of N fertilizer using random parameter models." *Journal of Agricultural and Applied Economics* 43, no. 4 (2011): 541-552.

Wallach, D. "Regional optimization of fertilization using a hierarchical linear model." *Biometrics* 51 (1995): 338-346.

Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data.* Cambridge: MIT Press, 2010.

Zellner, Arnold. "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias." *Journal of the American Statistical Association* 57, no. 298 (1962): 348-368.

Zhou, Bin, and Kara M Kockelman. "Predicting the distribution of households and employment: a seemingly unrelated regression model with two spatial processes." *Journal of Transport Geography*, 2009: 369-376.

Zingore, S, H K Murwira, R J Delve, and K E Giller. "Soil type, historical management and current resource allocation: three dimensions regulating variability of maize yields and nutrient use efficiencies on smallholder farms." *Field Crops Research*, 2007: 296-305.

Appendices

Table 10: Results for different model specifications with RDP Fixed Effects

	OLS	GLS with FE	GLS with RE
(Intercept)	1243.72***	1243.72***	1162.75***
	(129.94)	(129.94)	(123.40)
rdpBlantyre Shire Highlands	173.76**	173.76**	
	(58.34)	(58.34)	
rdpBwanje	410.99***	410.99***	
	(66.73)	(66.73)	
rdpCentralMzimba	-259.13*	-259.13*	
	(103.22)	(103.22)	
rdpChikwawa	163.96**	163.96**	
	(55.31)	(55.31)	
rdpChitipa	-203.45	-203.45	
	(106.40)	(106.40)	
rdpDedza Hills	0.12	0.12	
	(106.54)	(106.54)	
rdpDowa East	-149.23*	-149.23*	
	(75.20)	(75.20)	
rdpDowa West	26.30	26.30	
	(62.90)	(62.90)	
rdpKaronga	91.06	91.06	
	(59.84)	(59.84)	
rdpKasungu	-142.02	-142.02	
	(108.31)	(108.31)	
rdpKawinga	-522.32***	-522.32***	
	(64.64)	(64.64)	
rdpLilongwe East	111.94	111.94	
	(58.36)	(58.36)	
rdpLilongwe West	68.41	68.41	
	(101.41)	(101.41)	
rdpMangochi	61.64	61.64	
	(65.10)	(65.10)	
rdpMchinji	-119.09	-119.09	
	(106.68)	(106.68)	
rdpMulanje	-84.51	-84.51	
	(71.89)	(71.89)	
rdpMwanza	136.62*	136.62*	
	(68.30)	(68.30)	
rdpNamwera	-875.18***	-875.18***	
	(110.29)	(110.29)	
rdpNkhata Bay	-391.40***	-391.40***	
	(65.91)	(65.91)	
rdpNkhotakota	-152.36*	-152.36*	
	(67.66)	(67.66)	

rdpNsanje	-274.37***	-274.37***	
	(68.81)	(68.81)	
rdpNtcheu	-247.06***	-247.06***	
	(68.82)	(68.82)	
rdpNtchisi	-30.21	-30.21	
	(109.12)	(109.12)	
rdpPhalombe	155.01*	155.01*	
	(63.70)	(63.70)	
rdpRumphi / North Mzimba	-75.35	-75.35	
	(60.92)	(60.92)	
rdpSalima	152.63*	152.63*	
	(66.55)	(66.55)	
rdpSouthMzimba	-431.55***	-431.55***	
	(105.38)	(105.38)	
rdpThiwi-Lifidzi	125.55	125.55	
	(109.25)	(109.25)	
rdpThyolo	20.95	20.95	
	(67.35)	(67.35)	
rdpZomba	35.44	35.44	
	(56.89)	(56.89)	
Season (1997/98=1)	-258.92***	-258.92***	-259.55***
	(16.22)	(16.22)	(16.20)
Variety-MH17	88.85	88.85	45.60
	(124.77)	(124.77)	(120.42)
Variety-MH18	130.11	130.11	166.20
	(120.37)	(120.37)	(117.79)
Soil texture (Medium=1)	161.95***	161.95***	162.53***
	(16.00)	(16.00)	(15.98)
N	**29.40***	**29.40***	**29.40***
	(0.82)	(0.82)	(0.82)
N Squared	**-0.12***	**-0.12***	**-0.12***
	(0.01)	(0.01)	(0.01)
R^2	0.34		
Adj. R^2	0.33		
Num. obs.	15690	15690	15690
AIC		259916.82	260278.00
BIC		260207.84	260346.94
Log Likelihood		-129920.41	-130130.00
Num. groups			31

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 11: ADD Varying Intercept and N Response Coefficient Models

	OLS	GLS with FE	GLS with RE
(Intercept)	1471.16***	1471.16***	1599.07***
	(63.15)	(63.15)	(142.13)
Year (1995/6=1)	255.76***	255.76***	
	(16.36)	(16.36)	
mz_varMH18	194.68***	194.68***	194.70***
	(20.58)	(20.58)	(20.58)
mz_varcomposite	-50.70	-50.70	-50.31
	(116.54)	(116.54)	(116.54)
lite_txtlight	-152.16***	-152.16***	-152.13***
	(15.83)	(15.83)	(15.83)
addidKRADD	-181.38*	-181.38*	-181.38*
	(87.59)	(87.59)	(87.59)
NKRADD	**25.83*****	**25.83*****	**25.83*****
	(3.14)	(3.14)	(3.14)
NsqKRADD	**-0.12*****	**-0.12*****	**-0.12*****
	(0.03)	(0.03)	(0.03)
addidMZADD	-606.99***	-606.99***	-606.97***
	(74.47)	(74.47)	(74.47)
NMZADD	**27.85*****	**27.85*****	**27.85*****
	(2.15)	(2.15)	(2.15)
NsqMZADD	**-0.09*****	**-0.09*****	**-0.09*****
	(0.02)	(0.02)	(0.02)
addidKADD	-443.78***	-443.78***	-443.76***
	(76.79)	(76.79)	(76.79)
NKADD	**31.19*****	**31.19*****	**31.19*****
	(2.33)	(2.33)	(2.33)
NsqKADD	**-0.13*****	**-0.13*****	**-0.13*****
	(0.02)	(0.02)	(0.02)
addidLADD	-354.89***	-354.89***	-354.85***
	(72.18)	(72.18)	(72.18)
NLADD	**32.04*****	**32.04*****	**32.04*****
	(1.95)	(1.95)	(1.95)
NsqLADD	**-0.13*****	**-0.13*****	**-0.13*****
	(0.02)	(0.02)	(0.02)
addidSLADD	-313.87***	-313.87***	-313.84***
	(85.93)	(85.93)	(85.93)
NSLADD	**30.05*****	**30.05*****	**30.05*****
	(3.04)	(3.04)	(3.04)
NsqSLADD	**-0.11*****	**-0.11*****	**-0.11*****
	(0.03)	(0.03)	(0.03)
addidMADD	-572.58***	-572.58***	-572.54***
	(72.93)	(72.93)	(72.93)

NMADD	**30.25***	**30.25***	**30.25***
	(2.09)	(2.09)	(2.09)
NsqMADD	**-0.13***	**-0.13***	**-0.13***
	(0.02)	(0.02)	(0.02)
addidBLADD	-312.56***	-312.56***	-312.55***
	(74.01)	(74.01)	(74.01)
NBLADD	**31.58***	**31.58***	**31.58***
	(2.20)	(2.20)	(2.20)
NsqBLADD	**-0.14***	**-0.14***	**-0.14***
	(0.02)	(0.02)	(0.02)
NSVADD	**21.19***	**21.19***	**21.19***
	(2.84)	(2.84)	(2.84)
NsqSVADD	**-0.10***	**-0.10***	**-0.10***
	(0.03)	(0.03)	(0.03)
R^2	0.32		
Adj. R^2	0.32		
Num. obs.	15690	15690	15690
AIC		260563.69	260577.62
BIC		260785.80	260799.73
Log Likelihood		-130252.84	-130259.81
Num. groups			2

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 12: RDP Varying Coefficient Models

	OLS	GLS with FE	GLS with RE
(Intercept)	945.14^{***}	945.14^{***}	981.94^{***}
	(61.57)	(61.57)	(147.07)
Year (1995/96=1)	257.90^{***}	257.90^{***}	258.57^{***}
	(16.16)	(16.16)	(16.16)
Variety-MH18	167.90^{*}	167.90^{*}	219.26^{**}
	(68.90)	(68.90)	(75.90)
Variety-Composite	-15.19	-15.19	10.50
	(121.10)	(121.10)	(122.26)
Soil Texture(Light=1)	-162.34^{***}	-162.34^{***}	-162.56^{***}
	(15.96)	(15.96)	(15.96)
rdpid11	-12.42	-12.42	-19.72^{*}
	(9.65)	(9.65)	(10.01)
rdpid12	509.56^{***}	509.56^{***}	244.36
	(99.73)	(99.73)	(146.97)
rdpid21	19.52^{***}	19.52^{***}	22.54^{***}
	(5.62)	(5.62)	(5.84)
rdpid22	-23.64	-23.64	5.88
	(29.72)	(29.72)	(33.13)
rdpid23	-11.42	-11.42	-8.05
	(7.18)	(7.18)	(7.37)
rdpid24	-9.83	-9.83	-6.97
	(6.91)	(6.91)	(7.05)
rdpid31	6.29	6.29	-23.17
	(17.19)	(17.19)	(26.82)
rdpid32	12.84	12.84	-12.03
	(13.68)	(13.68)	(22.13)
rdpid33	8.12	8.12	-18.73
	(11.38)	(11.38)	(17.97)
rdpid34	-11.66	-11.66	-20.27^{*}
	(6.96)	(6.96)	(8.26)
rdpid35	16.91	16.91	-0.89
	(11.04)	(11.04)	(16.59)
rdpid41	28.37^{***}	28.37^{***}	42.00^{***}
	(5.46)	(5.46)	(9.02)
rdpid42	7.22	7.22	15.29^{*}
	(4.62)	(4.62)	(7.35)
rdpid43	11.71	11.71	23.44^{**}
	(6.38)	(6.38)	(8.82)
rdpid44	4.85	4.85	15.65^{*}
	(5.39)	(5.39)	(7.79)
rdpid45	-7.33	-7.33	0.53
	(5.51)	(5.51)	(7.82)
rdpid51	11.33^{**}	11.33^{**}	11.21

	(4.36)	(4.36)	(11.52)
rdpid52	-0.69	-0.69	-0.80
	(4.21)	(4.21)	(10.88)
rdpid53	14.35***	14.35***	14.24
	(3.87)	(3.87)	(10.21)
rdpid61	5.01	5.01	3.40
	(3.13)	(3.13)	(8.28)
rdpid62	-10.53**	-10.53**	-10.76
	(3.43)	(3.43)	(8.01)
rdpid63	4.12	4.12	2.67
	(2.77)	(2.77)	(7.41)
rdpid64	-8.44**	-8.44**	-9.80
	(2.58)	(2.58)	(6.99)
rdpid65	2.48	2.48	1.21
	(1.93)	(1.93)	(6.31)
rdpid71	4.08*	4.08*	1.78
	(1.73)	(1.73)	(5.53)
rdpid72	2.75	2.75	0.62
	(2.21)	(2.21)	(5.31)
rdpid73	5.87***	5.87***	3.84
	(1.77)	(1.77)	(4.95)
rdpid74	5.01**	5.01**	3.07
	(1.88)	(1.88)	(4.81)
rdpid75	1.63	1.63	-0.46
	(1.99)	(1.99)	(5.16)
rdpid81	8.75***	8.75***	5.81
	(1.14)	(1.14)	(4.01)
rdpid82	3.06	3.06	0.28
	(1.58)	(1.58)	(3.96)
N11	35.32***	35.32***	28.67***
	(3.92)	(3.92)	(4.62)
N12	24.27***	24.27***	24.27***
	(4.07)	(4.07)	(4.07)
N21	25.28***	25.28***	26.69***
	(3.44)	(3.44)	(3.52)
N22	29.97***	29.97***	30.18***
	(3.74)	(3.74)	(3.74)
N23	23.69***	23.69***	26.54***
	(3.71)	(3.71)	(3.97)
N24	25.41***	25.41***	26.53***
	(4.15)	(4.15)	(4.20)
N31	36.03***	36.03***	36.02***
	(5.09)	(5.09)	(5.09)
N32	29.92***	29.92***	29.83***
	(4.67)	(4.67)	(4.67)
N33	29.79***	29.79***	29.67***

	(4.56)	(4.56)	(4.56)
N34	38.07***	38.07***	32.79***
	(5.39)	(5.39)	(6.04)
N35	29.77***	29.77***	29.70***
	(5.46)	(5.46)	(5.46)
N41	25.55***	25.55***	26.35***
	(3.22)	(3.22)	(3.24)
N42	32.43***	32.43***	33.55***
	(3.66)	(3.66)	(3.74)
N43	35.89***	35.89***	35.91***
	(5.34)	(5.34)	(5.34)
N44	35.47***	35.47***	35.49***
	(4.79)	(4.79)	(4.78)
N45	34.41***	34.41***	34.41***
	(5.48)	(5.48)	(5.47)
N51	30.33***	30.33***	30.34***
	(5.11)	(5.11)	(5.11)
N52	27.22***	27.22***	27.22***
	(5.29)	(5.29)	(5.29)
N53	31.89***	31.89***	31.89***
	(5.14)	(5.14)	(5.14)
N61	32.59***	32.59***	32.59***
	(4.93)	(4.93)	(4.93)
N62	23.43***	23.43***	23.43***
	(5.79)	(5.79)	(5.79)
N63	28.24***	28.24***	28.24***
	(4.90)	(4.90)	(4.90)
N64	30.11***	30.11***	30.11***
	(4.81)	(4.81)	(4.81)
N65	32.59***	32.59***	32.58***
	(3.54)	(3.54)	(3.54)
N71	33.21***	33.21***	33.20***
	(3.81)	(3.81)	(3.81)
N72	26.90***	26.90***	26.90***
	(5.87)	(5.87)	(5.86)
N73	29.13***	29.13***	29.12***
	(4.68)	(4.68)	(4.68)
N74	32.72***	32.72***	32.72***
	(5.37)	(5.37)	(5.37)
N75	34.44***	34.44***	34.43***
	(5.24)	(5.24)	(5.23)
N81	20.86***	20.86***	20.85***
	(3.25)	(3.25)	(3.25)
N82	23.05***	23.05***	23.04***

	(5.46)	(5.46)	(5.46)
Nsq11	-0.18***	-0.18***	-0.13**
	(0.04)	(0.04)	(0.05)
Nsq12	-0.12**	-0.12**	-0.12**
	(0.04)	(0.04)	(0.04)
Nsq21	-0.06	-0.06	-0.07*
	(0.04)	(0.04)	(0.04)
Nsq22	-0.10**	-0.10**	-0.10**
	(0.04)	(0.04)	(0.04)
Nsq23	-0.06	-0.06	-0.08*
	(0.04)	(0.04)	(0.04)
Nsq24	-0.10*	-0.10*	-0.11*
	(0.04)	(0.04)	(0.04)
Nsq31	-0.18***	-0.18***	-0.18***
	(0.05)	(0.05)	(0.05)
Nsq32	-0.11*	-0.11*	-0.11*
	(0.05)	(0.05)	(0.05)
Nsq33	-0.10*	-0.10*	-0.10*
	(0.04)	(0.04)	(0.04)
Nsq34	-0.20***	-0.20***	-0.16**
	(0.06)	(0.06)	(0.06)
Nsq35	-0.11*	-0.11*	-0.11*
	(0.05)	(0.05)	(0.05)
Nsq41	-0.09**	-0.09**	-0.09**
	(0.03)	(0.03)	(0.03)
Nsq42	-0.13***	-0.13***	-0.14***
	(0.04)	(0.04)	(0.04)
Nsq43	-0.16**	-0.16**	-0.16**
	(0.05)	(0.05)	(0.05)
Nsq44	-0.15**	-0.15**	-0.15**
	(0.05)	(0.05)	(0.05)
Nsq45	-0.17**	-0.17**	-0.17**
	(0.05)	(0.05)	(0.05)
Nsq51	-0.13**	-0.13**	-0.13**
	(0.05)	(0.05)	(0.05)
Nsq52	-0.08	-0.08	-0.08
	(0.05)	(0.05)	(0.05)
Nsq53	-0.12*	-0.12*	-0.12*
	(0.05)	(0.05)	(0.05)
Nsq61	-0.15**	-0.15**	-0.15**
	(0.05)	(0.05)	(0.05)
Nsq62	-0.09	-0.09	-0.09
	(0.06)	(0.06)	(0.06)
Nsq63	-0.11*	-0.11*	-0.11*
	(0.05)	(0.05)	(0.05)

Nsq64	-0.13**	-0.13**	-0.13**
	(0.05)	(0.05)	(0.05)
Nsq65	-0.14***	-0.14***	-0.14***
	(0.03)	(0.03)	(0.03)
Nsq71	-0.14***	-0.14***	-0.14***
	(0.04)	(0.04)	(0.04)
Nsq72	-0.11	-0.11	-0.11
	(0.06)	(0.06)	(0.06)
Nsq73	-0.13**	-0.13**	-0.13**
	(0.05)	(0.05)	(0.05)
Nsq74	-0.17**	-0.17**	-0.17**
	(0.05)	(0.05)	(0.05)
Nsq75	-0.17**	-0.17**	-0.17**
	(0.05)	(0.05)	(0.05)
Nsq81	-0.09**	-0.09**	-0.09**
	(0.03)	(0.03)	(0.03)
Nsq82	-0.13*	-0.13*	-0.13*
	(0.05)	(0.05)	(0.05)
R^2	0.34		
Adj. R^2	0.34		
Num. obs.	15690	15690	15690
AIC		260099.32	260093.65
BIC		260857.12	260859.10
Log Likelihood		-129950.66	-129946.83
Num. groups			8

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

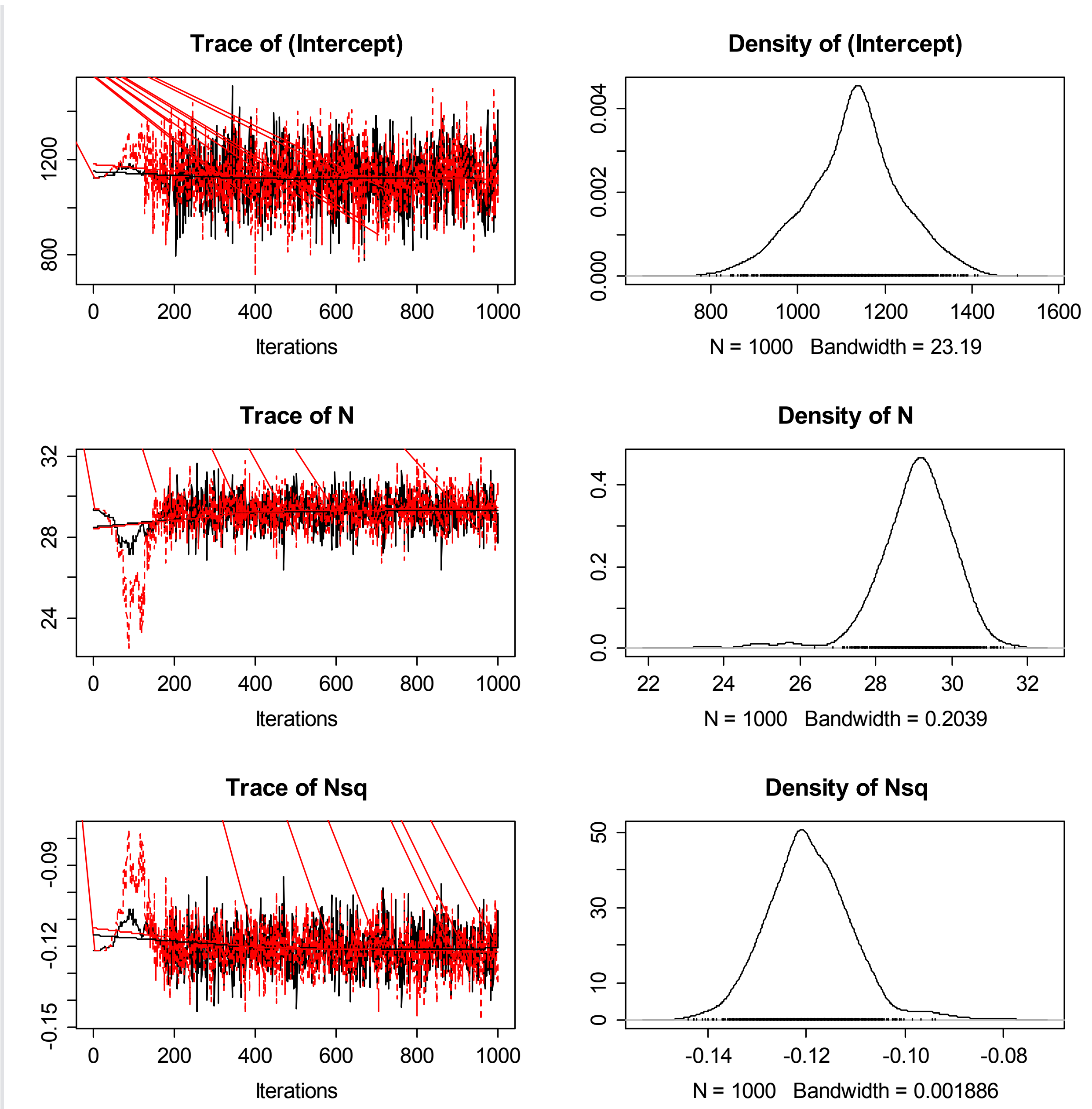


Figure 21: Trace plots of the intercept, N and Nsq at different starting values.