Learning from Neighboring Communities: A Spatial Analysis of Improved Seed Adoption

Juan Segundo Zapiola[1]
Universidad de San Andrés

Abstract

This study investigates how spatial interactions between neighboring communities influence decisions regarding the proportion of improved seed adoption in rural Malawi. Using georeferenced community panel data from 2010 to 2019, the study explores two key research questions: (1) how spatial interactions with neighboring communities influence the proportion of improved seed varieties adopted, and (2) how drought shocks affect the strength and direction of these peer effects. A Spatial Lag Model (SLM) is estimated to quantify this spatial interaction, where results show that neighboring communities' behavior significantly influences local adoption rates, and this influence becomes stronger in the aftermath of severe drought shocks.

Keywords: Improved seed adoption, spatial econometrics, peer effects, drought shocks, social learning, Geographically Weighted Regression (GWR), Spatial Lag Model (SLM), Malawi, technology diffusion, smallholder agriculture.

March, 2025

[1] Juan Segundo Zapiola, jzapiola@udesa.edu.ar.

Contents

1 Introduction

The adoption of new technologies have become crucial in tackling climate change, especially for those regions facing socioeconomic adversities. Sub-Saharan Africa (SSA) continues to face profound challenges as the most food-insecure region in the world, exacerbated by the limited adoption of advanced technological interventions (Arslan et al., 2016). Smallholder agriculture remains the primary driver of economic growth in most SSA countries, yet it is also the most vulnerable to the adverse effects of climate change (Barrios et al., 2008). Adopting improved seed varieties have shown to increase agricultural productivity accounting for about 50–90% of global crop yield increase (Kafle, 2010). The possibility of SSA farmers to adopt to new improved seed varieties and be aware of new agricultural technology is still very limited. Indeed, information, or the lack thereof, is often found to be the most important factor limiting rural development among farmers and communities (Fang and Richards, 2018).

The decision to adopt improved seed varieties is heavily influenced by both farmers and community characteristics. According to research, factors such as access to seeds, extension services, labor availability, and the location of farm households play a crucial role in determining adoption rates (Martey et al., 2020). Additionally, characteristics like farm size and membership in farmer organizations have been found to positively impact the adoption of improved seed varieties. Membership in these organizations enhances access to information and facilitates collective action, which strengthens the dissemination of agricultural technologies and encourages broader adoption of new practices (Kalinda et al., 2014). Furthermore, seed scaling programs implemented through community institutions have been shown to significantly increase adoption rates compared to non-collective approaches (Hossain et al., 2024), highlighting the importance of community structures in the adoption process.

The influence of neighboring farmers and communities plays a pivotal role in the adoption of improved seed varieties. Social learning, or learning by observing others, has been identified as a key mechanism that increases the likelihood of adoption (Fang and Richards, 2018; Li et al., 2021). Farmers tend to cluster around similar practices, where the decisions of neighboring farmers significantly impact their own choices regarding seed adoption. In addition, Fang and Richards (2018) showed that local networks act as important agents for information exchange, where farmers rely on their immediate neighbors for recommendations, and weigh the neighbors' opinions heavily. They also go beyond immediate neighbors to exploit extended social networks for gathering information. This spatial dependency is critical, as interactions within these networks facilitate the diffusion of information, leading to higher adoption rates.

The interaction effect of social learning offers valuable insights into how neighboring communities influence decisions regarding the quantity of improved seed varieties used. Based on the literature, I anticipate a positive spatial lag effect when a community follows the practices of its neighbors, and a negative spatial lag effect when it diverges from them. It is expected that social learning operates through interactions within each network, indicating that these interactions are inherently spatial.

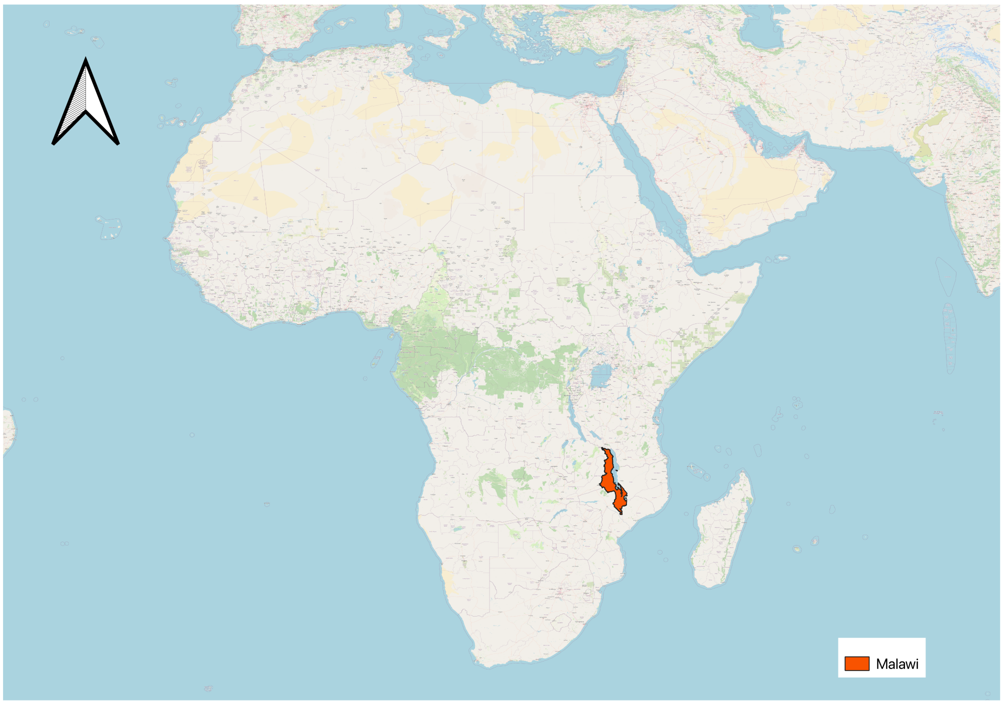
Smallholder farming systems are accustomed to coping with many challenges, including climate variability (Nordhagen and Pascual, 2013). However, climate change may pose formidable novel challenges for which traditional livelihood strategies may not be fully suited (FAO, 2008). Some of the key adaptation strategies that have been studied in the literature include switching toward crop varieties resistant to heat and drought, development and adoption of new cultivars, changing the farm portfolio of crops and livestock, integrating the use of forecasts into cropping decisions, uses on fertilizers and irrigation, increased storage of food/feed or reliance on imports, and many more (Nordhagen and Pascual, 2013). Under drought shocks, farmer's seed security becomes profoundly influenced. Seed security is often defined by the three parameters of availability (seed being available in space and time), access (physical and economic access), and utilization (seed quality meets user's needs and preferences) (Makate et al., 2023; Sperling et al., 2008; McGuire and Sperling, 2016). Seed purchasing and the quantity of what seed variety to use enables the farmer and farmer communities to respond to negative factors that result in chronic and temporary seed insecurity, like drought shocks. Some studies have found that drought shock exposure in prior seasons increases seed purchasing for both improved and local seeds in Malawi and Tanzania while encouraging (discouraging) local (improved) seed purchases in Ethiopia (Makate et al., 2023). Others, found that having experienced an adverse weather-related shock had a significantly positive impact on the likelihood of farmers purchasing local seeds (not improved) (Nordhagen and Pascual, 2013). Following this line, a study in Kenya found that frequent past climatic shocks, as manifested by drought incidence, reduce the maize area share per farm allocated to hybrid seeds (Bozzola et al., 2018). Building on this literature, I anticipate that a drought shock during the previous rainy season, when crops are grown, will increase the likelihood of farmers purchasing local seeds. However, there is a gap in understanding how such drought shocks might alter the influence of neighbors on your decision regarding the quantity of seed varieties to use.

Our study primary centers in Malawi, a representative nation within the Sub-Saharan Africa (SSA) region, where the agriculture sector plays a pivotal role in Malawi's socioeconomic wellbeing and poverty alleviation (Stevens and Madani, 2016). The study aims to: (1) assess how spatial interactions with neighboring communities influence the proportion of improved seed varieties used, and (2) examine how drought shocks shape the dynamics of peer effects on the adoption of improved seed varieties, using georeferenced community panel data from Malawi (2010–2019). The primary contribution of this study is to provide deeper insights into the learning dynamics among rural communities, where the proportion of land dedicated to improved seed varieties may be influenced by information shared with neighboring communities. Furthermore, despite the growing literature on this topic, there remains a significant gap in understanding the spatial dependence of seed adoption in Malawi. This study addresses this gap by being the first to examine how spatial interactions between communities influence the adoption of improved seed varieties in the region. Furthermore, it is among the first to explore how spatial dependence affects the quantity of improved seeds used and how drought shocks influence peer dynamics in decisions about the quantity of improved seeds to adopt, using available plot-level data.

2 Malawi

Malawi is a small, landlocked country in southern Africa, bordered by Tanzania, Mozambique, and Zambia as shown in Figure 1. It has a population of 19 million and is divided into 28 districts across three regions: northern, central, and southern. The Great Rift Valley runs through the country from north to south, with Lake Malawi (also known as Lake Nyasa) forming more than three-quarters of its eastern boundary.

Figure 1: Map of Malawi. *Own elaboration using QGIS.*



Malawi is one of the world's least developed countries, with 50.7% of the population living in poverty and 20.5% in extreme poverty (NSO, 2021). The economy is heavily reliant on agriculture, which has contributed over 40% of economic growth in recent years. More than 75% of Malawian adults work in agriculture, primarily in small-scale, rainfed farming. The rural economy, which encompasses various income-generating activities in rural areas, represents a major share of both employment and national output. According to the 2018 census, 84% of the population resides in rural areas and depends largely on subsistence farming (Engel et al., 2022).

Ninety-four percent of poor households in Malawi are located in rural areas, where the poverty headcount ratio reaches 57%—almost three times higher than in urban centers (19%) (Gross et al., 2021). This disparity is driven by low agricultural productivity, which is influenced by small average farm sizes of just 0.7 ha—significantly smaller than in other countries in the region (World Bank, 2022). Productivity among smallholder farmers is particularly low due to multiple constraints, including climate variability (floods and dry spells), limited adoption of improved agricultural technologies and practices, restricted access to quality inputs, and poor availability of credit and markets. These challenges hinder rural households from increasing yields and improving their economic conditions.

Malawi has two distinct seasons: a wet, warm season from October to April and a dry, cool season from May to September (CIAT and World Bank, 2018). Agricultural production is concentrated during the wet season, as 90% of farming is rain-fed and only 4% of the total cultivated area is irrigated (CIAT and World Bank, 2018). The seven major crops cultivated in the country are maize, groundnuts, pigeon peas, beans, soybeans, rice, and tobacco. Among smallholder farmers, 96% grow maize, while 30% cultivate groundnuts, 25% pigeon peas, 10% soybeans and rice, and 3% tobacco (National Statistical Office, 2020). Maize is the most widely produced crop, occupying 28% of total agricultural land, followed by groundnuts with 6% (CIAT and World Bank, 2018).

3 Methodology

This section presents the methodological framework for analyzing the spatial dynamics of improved seed adoption in Malawi. First, we introduce Geographically Weighted Regression (GWR) to examine spatial heterogeneity and assess regional variations in adoption patterns. Next, we define and construct spatial weight matrices to quantify spatial relationships between communities and apply Moran's I test to detect spatial autocorrelation. Finally, we outline the spatial econometric models that explicitly incorporate spatial dependence to better understand the diffusion of improved seed varieties.

3.1 Spatial Heterogeneity (GWR)

When studying the adoption of improved seeds, spatial heterogeneity plays a crucial role, as different regions may experience different levels of adoption. To examine if there is spatial heterogeneity on improved seed adoption in Malawi a geographically weighted regressions (GWR) estimation would be implemented. GWR allows us to examine how the proportion of improved seed adoption varies across space.

Geographically Weighted Regression (GWR), introduced by Brunsdon et al. (1996), is a technique that extends linear regression by allowing for localized analysis. This method identifies spatial variations in the relationships between variables, enabling a more detailed understanding of how these associations change across different regions. The core idea behind GWR is that relationships between variables are influenced by the local context rather than remaining constant across the entire geographic area.

With GWR we then estimate locally the impact of improved seed adoption for each year of this study. This estimation allows the relationships between variables to vary continuously across space:

$$y_i = \sum_{j=1}^{k} \beta_j(\text{lat}_i, \text{lon}_i) x_{ij} + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where y_i is the proportion of improved seed used geographically identified by its latitude and longitude $(\text{lat}_i, \text{lon}_i)$. The sample size is denoted as n. The coefficients $\beta_j(\text{lat}_i, \text{lon}_i)$ are unknown parameters that depend on the location coordinates, and ε_i represents an error term with a mean of zero and constant variance σ^2. In this study, we aim to estimate the local impact of improved seed adoption. Instead of including explanatory variables in the model, we use only the intercept. Equation (1) is estimated using Ordinary Least Squares (OLS) by minimizing the following loss function:

$$\sum_{i=1}^{n} \left[y_i - \sum_{j=1}^{k} \beta_j(lat_i, lon_i) x_{ij} \right]^2 w(d_{0i}), \tag{2}$$

where $w(.)$ represents a geographic weighting function, which follows a tricube kernel. The optimal bandwidth is determined through cross-validation. The results obtained using this method are presented in the Results section under Descriptive Statistics subsection.

3.2 Spatial Weight Matrix & Spatial Dependence

In spatial econometrics, assuming independence between observations is highly restrictive, as the observed independent values y for a given observation i are often statistically similar to those in nearby locations. This phenomenon is known as spatial dependence (Ahumada et al., 2018). To estimate it, a starting point is to define an appropriate weights matrix, $\mathbf{W}$. Based on Tobler (1970) "everything is related to everything else, but near things are more related than distant things", this weights matrix represents the independence and interactions strength between the spatial observations.

Since the community's proportion of improved seed varieties used are more likely to be influenced by their nearby neighbor communities rather than the remotes ones, the inverse distance between communities' locations is usually used as weights to measure the proximity between communities and neighboring communities (Elhorst, 2014; LeSage and Pace, 2009). To construct the distance-based weight matrix $\mathbf{W_d}$, following the studies of Zheng et al. (2021);

Fang and Richards (2018); Li et al. (2021), it is measured as

$$\mathbf{W_d} = w_{ij} = \begin{cases} 0 & \text{if} \;\; i = j \\ 1/d_{ij} & \text{if} \;\; d_{ij} \leq D \\ 0 & \text{if} \;\; d_{ij} > D \end{cases} \tag{3}$$

where d_{ij} is the arc-distance between the centroids of community i and the neighbouring community j. D is the critical distance cut-off, it would be assumed that there is no spatial relationship between communities that are beyond the distance D from each other, as indicated in equation (3). The critical distance cut-off D to be used would be 100 km.

An extra weight matrix would be defined to emphasize on immediate neighbors. This weight matrix $\mathbf{W_k}$ will be constructed using the k-Nearest Neighbors method, meaning that only the k closest observations or communities are considered to exert spatial influence. Both matrices would be row standardized. Following Fang and Richards (2018), for this weight matrix a $k = 5$ would be used,

$$\mathbf{W_k} = w_{ij} = \begin{cases} 1 & \text{if } d_{ij} \leq d_{ij}^k, \\ 0 & \text{if } d_{ij} > d_{ij}^k. \end{cases} \tag{4}$$

where d_{ij} is the distance between observations i and j. d_{ij}^k represents the distance to the k -th nearest neighbor of i, where $k = 5$ in this case. So, if j is within the k -th nearest neighbors of i (i.e., $d_{ij} \leq \mathrm{d}_{ij}^k$), then $w_{ij} = 1$. Otherwise, if $d_{ij} > d_{ij}^k$, then $w_{ij} = 0$.

To test whether there is spatial dependence on the proportion of improved seeds used, the Moran's I test would be employed for the panel dataset, which is averaged from all time periods:

$$\text{Moran's I}_t = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{S^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}} \tag{5}$$

Where $S^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2$, $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$, y_i indicates the observed value in the ith communities, w_{ij} is the spatial weight of the link between two communities i and j (as defined above), and n refers to the total number of communities. It is important that each community persists across each wave so that $\mathbf{W}$ is constant over time and the panel is balanced (Elhorst, 2014).

3.3 Spatial Econometric Models

The previous section introduced how to detect spatial dependency and autocorrelation. Spatial externalities require verification, specification, and estimation through spatial econometric models. Anselin (2002) proposes a taxonomy of formal models for spatial externalities, where the classification depends on how the spatial variable is incorporated.

To introduce spatial models, it is first necessary to identify the three interaction effects that explain why an observation at a particular location may depend on observations in different locations (Anselin, 2002; Elhorst, 2014).

1. Endogenous Spatial Lag: This occurs when the dependent variable of agent A depends on the dependent variable of agent B, and vice versa. In this case, the value of the dependent variable for an agent is jointly determined by the values of its neighbors.

2. Exogenous Spatial Lag: This arises when the dependent variable of a given agent is influenced by the independent variables of other agents.

3. Spatial Error Dependence: This effect occurs when omitted determinants of the dependent variable are spatially autocorrelated or when unobservable shocks follow a spatial pattern.

The most complete and general model that incorporates all spatial interaction effects is the Cliff-Ord Model. The endogenous spatial lag is specified as Wy (substantive spatial dependence), the exogenous spatial lag is specified as WX, and the spatial error lag is specified as Wu (residual spatial dependence). The parameter ρ is referred to as the spatial lag coefficient, θ represents the local spatial dependence coefficient, and λ is the spatial autocorrelation coefficient of the errors.

Following the most general model, there are three spatial econometric models that incorporate two of the three spatial lag coefficients. When the restriction $\rho = 0$ is imposed, the model simplifies to the Spatial Durbin Error Model (SDEM), whereas setting $\lambda = 0$ results in the Spatial Durbin Model (SDM). Similarly, if $\theta = 0$, the model becomes the SARAR Model (Spatial Autoregressive and Spatial Error Model). Additional restrictions lead to further simplifications: imposing $\rho = 0$ and $\lambda = 0$ yields the SLX Model (Spatial Lag in X Model), while setting $\rho = 0$ and $\theta = 0$ results in the Spatial Error Model (SEM). Likewise, if $\lambda = 0$ and $\theta = 0$, we obtain the Spatial Lag Model (SLM). Finally, when all three restrictions are imposed ($\rho = 0$, $\lambda = 0$, and $\theta = 0$), the model reduces to a non-spatial linear regression model.

To illustrate all of the spatial interaction effects, equation (4) shows the Cliff-Ord Model based on this study.

$$y_{it} = Wy_{it}\rho + X_{it}\beta + WX_{it}\theta + u_{it}; \quad u_{it} = \lambda W u_{it} + \epsilon_{it} \tag{6}$$

y_{it} is the proportion of improved seed adopted for the community i at year t. X_{it} represents a matrix of the community's characteristics i at year t. The endogenous spatial lag is specified as Wy, the exogenous spatial lag is specified as WX, and the residual spatial lag is specified as Wu for the community i at year t. β, ρ, θ, and λ are it's corresponding parameters to be estimated. u_{it} refers to the error term.

The Cliff-Ord model is the most general among spatial econometric models; however, it faces an identification problem, a mathematical issue that arises before estimation, making it impossible to recover all spatial lag coefficients. This limitation stems from the Reflection Problem, which prevents the separation of endogenous and exogenous spatial effects unless additional restrictions are imposed. As a result, the most complex spatial model can incorporate only two out of the three spatial effects (ρ, λ, or θ), requiring a restriction on the remaining effect to ensure proper identification (Ahumada et al., 2018).

To select the final model, various statistical tests are applied to determine the best fit given the spatial variability in the data. There are two main strategies for model selection. The first approach follows a specific-to-general (STGE) framework, starting with the most restrictive model and gradually expanding to a more general or nested specification. The second approach follows a general-to-specific (GETS) framework, beginning with the most comprehensive model and progressively simplifying it. In this study we are going to follow the specific-to-general (STGE) framework. The models are estimated using the Maximum Likelihood (ML) method, which assumes that the observed data is generated by an unknown probability distribution from a given family. The goal is to find the parameter values that most likely produced the observed sample. To estimate the parameters that maximize this likelihood, a joint likelihood function L must be defined, assuming a known distribution for the error term. The likelihood function is given by $L(\theta, y) = f(y, \theta)$, where f represents the joint density function and θ is the parameter of interest. Section 4.3 presents how this framework and tests are done with our data so to present the final model to obtain our ML regression estimates.

4 Data

This study utilizes two georeferenced datasets: the Malawi Integrated Household Panel Survey, which provides information on household, community, and agricultural characteristics, as well as data on improved seed usage; and the Climatic Research Unit (CRU) SPEI database, which supplies the drought shock index. These datasets are further described, along with the presentation of descriptive statistics.

4.1 Malawi Integrated Household Panel Survey

This study will use the Malawi Integrated Household Panel Survey (IHPS). This is a nationally representative panel data of household, community, fishery and agriculture surveys implemented by the Government of Malawi through the National Statistical Office. The data was collected by World Bank's Living Standards Measurements Survey (LSMS) program. It is a georeferenced database, meaning that each enumeration area (EA)/community has it's longitude and latitude coordinates. Given the increasing numbers of households to be tracked, as well as budget/resource constraints, starting in 2016, the IHPS target household sample was adjusted as the households that have been associated with 102 out of 204 baseline EAs. The

IHPS collects data about health, education, labor, crop production, fertilizers, weather conditions, among others. The survey conducts interviews over a year, encompassing four waves occurring biennially.

From this study, we are going to obtain the quantity of improved seed varieties used at the community level for the 2010, 2013, 2016 and 2019 year waves from the agriculture questionary focused on the main rainy season. We will focus exclusively on households that own at least one agricultural plot. Accordingly, our dataset includes 95 out of the 102 georeferenced enumeration areas (EAs) or communities, comprising 885 households per round. Additionally, it will collect data on community characteristics and the information they have acquired regarding agricultural practices. From the panel dataset, those household members who migrated or lost track from the original household during the waves would be erased.

We define improved seeds as hybrid seeds, which, compared to local seeds, have the potential to produce higher yields. Hybrid seeds are the result of controlled crossbreeding between two genetically distinct parent plants of the same species to produce offspring with enhanced agricultural traits. This process, known as hybridization, aims to combine desirable characteristics from both parent lines, such as higher yield potential, disease resistance, stress tolerance, and improved uniformity (Acquaah, 2009). Unlike open-pollinated or local/traditional seeds, hybrid seeds do not produce offspring with identical traits in subsequent generations, meaning farmers must purchase new seeds each planting season to maintain optimal performance (Fehr, 1991). From the Agriculture Section, improved seeds include: Maize Hybrid, Groundnut CG7, Groundnut Manipinta, Groundnut Mawanga, Groundnut JL24, Rice Faya, Rice Pussa, Rice TCG10, Rice IET4094 and Rice ITA [2].

4.2 Weather Data & Drought Index

To address for drought shocks we utilized the Climatic Research Unit (CRU) SPEI database[3]. This base is a global scale high-resolution drought index developed from a combination of precipitation and potential evapotranspiration datasets for the Hydro-JULES project (Solomon Gebrechorkos, 2023). The drought index is developed using the Standardized Precipitation Evapotranspiration Index (SPEI). SPEI classifies drought based on precipitation data for 3-, 6-,12-, 24-, and 48-month timescales. However, what sets SPEI apart from other drought indices is its consideration of reference evapotranspiration, which represents the amount of water that would evaporate under reference conditions. SPEI considers climatic factors such as temperature, humidity, solar radiation, and wind, providing a more comprehensive measure of the available water (climatic water balance). By incorporating climatic factors, SPEI offers insights into the combined effects of precipitation and evapotranspiration on drought severity in different locations and time periods (Beguería et al., 2014).

[2]See: Malawi Integrated Household Panel Survey, Agriculture and Fishery Enumerator Manual (ANNEX 3)
[3]SPEI Dataset: https://catalogue.ceda.ac.uk/uuid/ac43da11867243a1bb414e1637802dec/

This high-resolution global scale drought index is available from 1981-2022 at a monthly and 5km spatial resolution (0.05 degree). The SPEI index is available from 1-48 months timescales. Selecting a specific timescale depends on assessing drought impacts on different water resources. For example, meteorological drought requires a 1- or 2-month SPEI, while agricultural drought analysis utilizes SPEI ranging from 1 to 6 months. For hydrological drought investigations and applications, longer timescales, such as six months up to 24 months or more, are typically employed (McKee et al., 1993; Svoboda et al., 2012; Santiago Beguería, 2023).

As it was previously mentioned, in Malawi, there are two distinct seasons: a dry season from May to September and a wet season from October to May where the main crops are produced. Malawi's seven primary crops—maize, groundnuts, pigeon peas, beans, soybeans, rice, and tobacco—share similar agricultural calendars, with overlapping periods for planting, flowering, and harvesting. These crops are generally planted with the onset of the rainy season, typically between October and December, and begin flowering around three months later. The first three months after planting are particularly critical, as they represent the phase of highest water demand for the crops. Adequate and timely rainfall during this stage is essential, contributing to approximately 70% of the final yield [4]. In contrast, drought conditions during this period significantly increase the likelihood of yield losses and crop failure. Given the importance of this window, the study uses a 3-month Standardized Precipitation Evapotranspiration Index (SPEI) calculated for each grid cell, capturing drought variability during the crucial December to February period of the rainy season.

5 Results

This section presents the findings of our spatial analysis on improved seed adoption in Malawi. We begin by providing descriptive statistics, offering an overview of key variables and their spatial distribution. Next, we examine spatial heterogeneity through Geographically Weighted Regression (GWR) to assess how the relationship of improved seed adoption varies across locations and time. Following this, we focus on the final model selection and estimations, where we evaluate different spatial econometric models to identify the best specification for capturing the spatial dependencies. Finally, we estimate this final model to analyze peer effects in seed adoption, both under normal conditions and in response to drought shocks, highlighting how spatial interactions under these shocks can shape the weight your neighbor has on your own decision-making across communities.

5.1 Descriptive Statistics

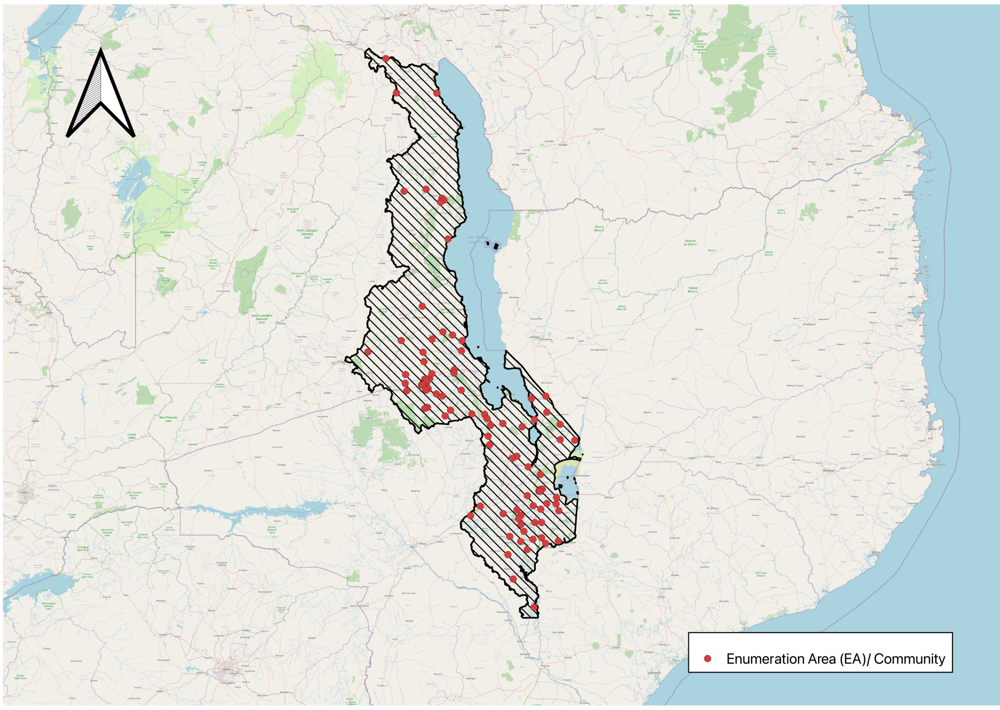
To provide an overview of the datasets used in this study, Figure 2 displays the geographic distribution of Enumeration Areas (EAs)/Communities in Malawi, while Table 1 presents key

[4] As stated by two agronomic engineers during an interview.

variables relevant to the analysis. The adoption variable measures the proportion of improved seed varieties used within communities, showing that, on average, 34% of plots have adopted improved seeds in each community. This metric serves as a key continuous variable for assessing the adoption of improved seeds, capturing variations in uptake across different regions.

Household and agricultural characteristics, following the frameworks of Fang et al. (2018) and Li et al. (2021), describe the socioeconomic context of the communities. On average, 25% of household heads are women, with a mean age of 44 years, and 20% are engaged in salaried employment. In terms of education, 54% of household heads have never received formal schooling, while only 11% have completed primary education. Regarding agricultural variables, the average household plot size is 16.14 acres. Access to subsidies and financial support remains limited, with only 4% of households receiving coupons to purchase improved seeds and 13% accessing credit for improved seed purchases. Additionally, 43% of households rely on leftover seeds from the previous season for planting, highlighting the constraints many farmers face in adopting improved seed varieties.

Figure 2: Map of Malawi with the Enumeration Areas (EAs)/ Communities. *Own elaboration using QGIS.*



5.1.1 Spatial Heterogeneity (GWR)

Figure 3 displays a map of Malawi with Geographically Weighted Regression (GWR) estimates for the proportion of improved seed adoption across the four rounds of this study. The map reveals clear spatial heterogeneity, indicating that the adoption rate varies across both location and time. This suggests that improved seed adoption is influenced by local contextual factors,

Table 1: Descriptive Statistics of Key Variables.

Variable	Description	Mean	SD
Adoption Var.			
prop_imp	Proportion of improved seed used	0.34	0.25
Household Var.			
prop_female_d	Proportion of female head in households	0.25	0.20
mean_age_h_d	Mean age of household head	44.73	6.17
porp_salar_d	Proportion of being salaried employed	0.20	0.25
prop_head_1	Proportion of head with no education	0.54	0.29
prop_head_2	Proportion of head with PSLC education	0.11	0.17
prop_head_3	Proportion of head with JCE education	0.09	0.16
prop_head_4	Proportion of head with MSCE education	0.07	0.15
prop_head_5	Proportion of head with non-university education	0.03	0.12
prop_head_6	Proportion of head with university degree education	0.01	0.08
prop_head_7	Proportion of head with post-grad degree education	0.01	0.05
Agriculture Var.			
total_plot_e	Plot size in Community (acres)	16.14	11.87
prop_coupon	Proportion of seeds bought with coupons	0.04	0.06
prop_credit	Proportion of seeds bought with credit	0.13	0.16
prop_left_s	Proportion who used left over seeds from previous season	0.43	0.22

Note: Mean and standard deviation are rounded to 2 decimal places.

rather than remaining uniform across the entire country.

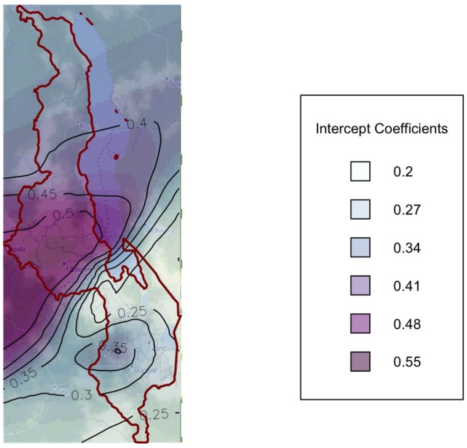
As mentioned before, adoption rates vary significantly by region, with some areas consistently exhibiting higher adoption levels than others. Examining the maps across rounds, we observe shifts in the spatial distribution of adoption rates over time. In Rounds 1 and 2, the highest adoption rates (darker shades of purple) are concentrated in the central and northern regions. However, in Rounds 3 and 4, we notice a redistribution of adoption intensity, with some previously high-adoption areas experiencing a decline, while others—particularly in the southern and central regions—show an increase. This spatial shift could be attributed to multiple factors, such as policy changes, climate variability, or shifts in farmer networks and peer influence.

Furthermore, the GWR results indicate that local environmental conditions and spatial spillovers may influence seed adoption. Farmers in certain regions may become more or less inclined to adopt improved seeds based on their experiences and risk assessments from their neighbors. We follow this section with the idea that exposure to climate risks, like drought shocks, can shape the adoption dynamics of improved seed varieties of communities and farmers.

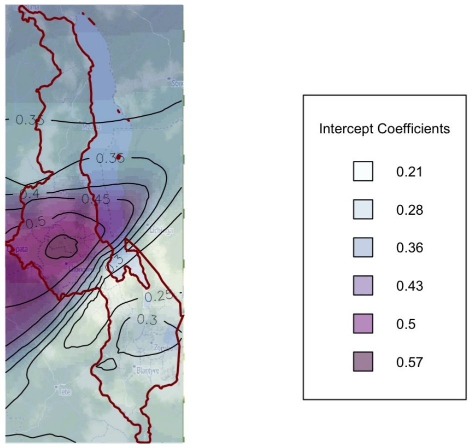
5.1.2 Drought Shock

Spatial interactions under drought shocks can shape the weight your neighbor has on your decision making across communities. Spatial interactions under drought shocks can influence

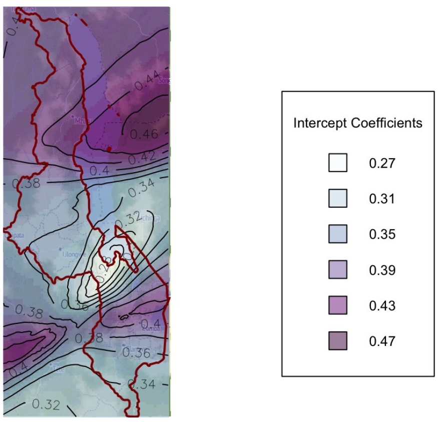
Figure 3: GWR estimations for the proportion of improved seed adoption in Malawi. *Own elaboration using R-Studio.*



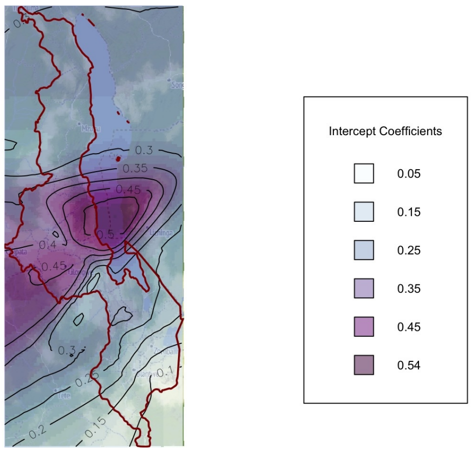
(a) Round 1 - 2010



(b) Round 2 - 2013



(c) Round 3 - 2016



(d) Round 4 - 2019

how much a neighbor's decision impacts your own across communities. As discussed in Section 2, agriculture plays a central role in Malawi's economy, and climate shocks like drought can affect decision-making, including the choice between improved or local seeds, while also shaping the influence of neighboring farmers' choices have on you.

Figure 4 presents the 3-month SPEI drought shocks in Malawi during the key planting period of the previous long rain season for each survey round. The year before the surveys were conducted, some areas experienced severe drought shocks during the key planting period, potentially influencing farmers' seed adoption decisions in the following season. Map (a) shows that many communities in the central region faced severe drought, likely leading to lower yields due to insufficient rainfall. In contrast, map (b) indicates that communities in the southern region received abundant rainfall, which may have resulted in higher yields, while those in the northern region experienced relatively low rainfall.

Table 2 shows the number of communities that experienced a drought shock during the previous year's planting season. In 2009, 22 communities faced a severe drought shock. In 2012, no community experienced a drought shock. In 2015, only one community faced a severe drought shock. Finally in 2018, 22 communities experienced a severe drought shock. In addition to this, Table 3 examines whether communities' pre-treatment characteristics are balanced across those who are never exposed to a drought shock in the previous year's planting season of each round (control) and those who are eventually exposed (eventually treated). For 12 out of 14 characteristics available there are no statistical significant differences across those never exposed and those eventually exposed to a drought shock.

Table 2: Communities under a 3-month SPEI Drought Shock

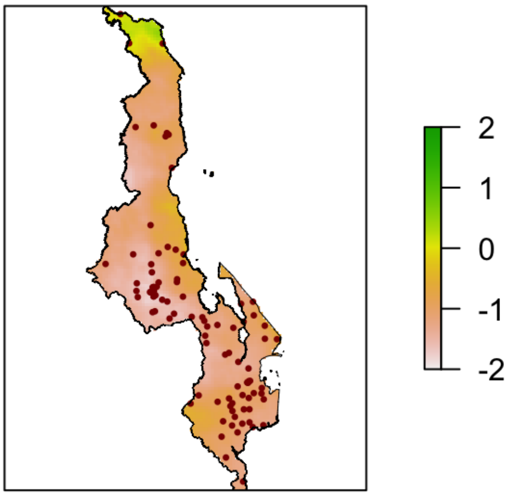
Drought Shock	2009	2012	2015	2018	Total
No Shock	73	95	94	73	335
Severe Shock	22	0	1	22	45

5.2 Model Selection and Estimations

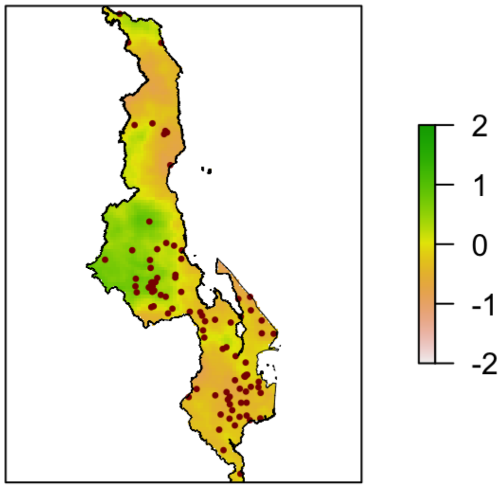
To identify the final spatial model to run our estimations that best fits the spatial variability effects of our data, we are going to follow the specific-to-general (STGE) framework, as specified in Section 3.3. As we are working with two spatial weight matrices, this framework is going to begin with $\mathbf{W_k}$ and be repeated for $\mathbf{W_d}$.

Let's remember that this strategy involves moving from a more restricted or simpler model to a more nested one. To incorporate spatial variables, we must use various statistical tests that provide sufficient evidence of the presence of spatial lags. These statistical tests include the Lagrange Multiplier (LM) tests and the Likelihood Ratio (LR) test. The LM tests detect

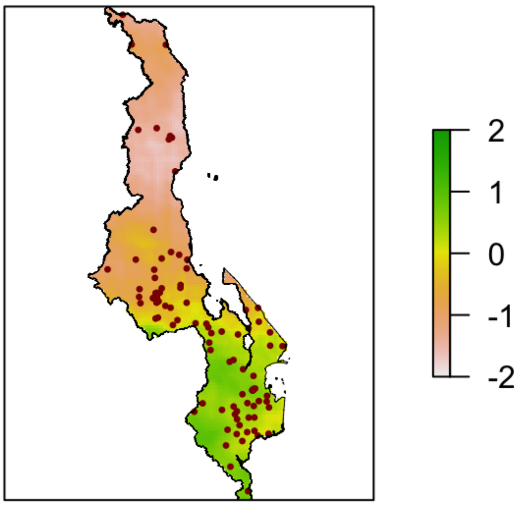
Figure 4: 3-month SPEI Drought Shock in Malawi. *Own elaboration using R-Studio.*



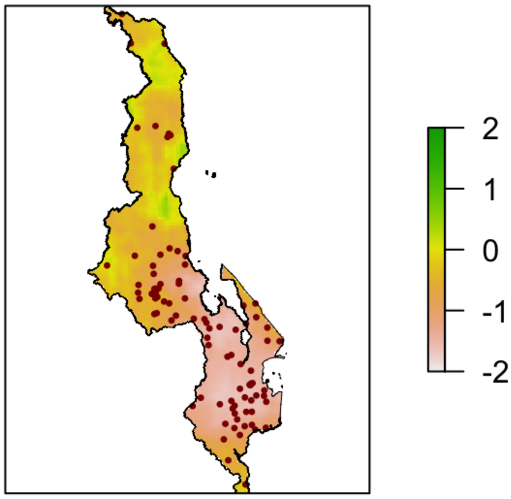
(a) Dec-Feb 2010



(b) Dec-Feb 2013



(c) Dec-Feb 2016



(d) Dec-Feb 2018

Table 3: Pre-Treatment Characteristics.

Variable	Control	Eventually Treated	Difference
prop_female_head	0.24	0.19	0.05
	(0.21)	(0.18)	[0.04]
mean_age_head	42.79	42.58	0.22
	(5.92)	(6.71)	[1.30]
prop_salaried_head	0.18	0.32	-0.13**
	(0.23)	(0.33)	[0.06]
prop_head_edu_1	0.55	0.46	0.09
	(0.22)	(0.33)	[0.06]
prop_head_edu_2	0.09	0.15	-0.05
	(0.15)	(0.24)	[0.04]
prop_head_edu_3	0.09	0.12	-0.03
	(0.17)	(0.24)	[0.04]
prop_head_edu_4	0.05	0.10	-0.05
	(0.11)	(0.21)	[0.03]
prop_head_edu_5	0.00	0.03	-0.03*
	(0.00)	(0.13)	[0.02]
prop_head_edu_6	0.00	0.01	-0.01
	(0.00)	(0.07)	[0.01]
prop_head_edu_7	0.00	0.01	-0.01
	(0.00)	(0.04)	[0.01]
total_plot_size	18.83	14.88	3.95
	(11.30)	(12.34)	[2.43]
prop_coupon	0.06	0.06	-0.00
	(0.05)	(0.06)	[0.01]
prop_credit	0.14	0.19	-0.05
	(0.14)	(0.23)	[0.04]
prop_left_seeds	0.48	0.42	0.06
	(0.21)	(0.26)	[0.05]

Note: Standard deviations are in parentheses. Standard errors are in brackets. Statistical significance levels: * p<0.1, ** p<0.05, *** p<0.01.

substantive or residual spatial dependence in the error term, while the LR test evaluates differences in the log-likelihood values of different models.

We start from our most simple model, the linear model with OLS (non-spatial model), where $\rho = 0 \wedge \lambda = 0 \wedge \theta = 0$. The first test done was Moran's I test to see if there is spatial dependence, presented in Table 4. The Moran's I statistic is 3.348, statistically significant at a 1% level, meaning that there is sufficient evidence of spatial dependence on the proportion of improved seed used. We repeat this test but with fixed effects, and also there is sufficient evidence of spatial dependence on the proportion of improved seed used (1% level). Based on this test, we run the LM tests for the lineal model with and without fixed effects, also presented in Table 4. Robust LM for either spatial error and spatial lag are a correction of the LM tests for the presence of $\rho = 0 \wedge \lambda = 0$ respectively. In both models, there is strong evidence of spatial lag dependence ($\rho \neq 0$) at the 1% significance level. However, there is no sufficient evidence to confirm spatial error dependence ($\lambda = 0$) in the model with fixed effects. Based on these results, the most competent spatial model is the SLM.

We repeat these tests but for the SLX model. This model imposes $\rho = 0$ and $\lambda = 0$, accounting only exogenous spatial lag. In Table 5, Moran's I and LM tests are presented for the SLX model. The Moran's I statistic is statistically significant at a 1% level, meaning that there is sufficient evidence of spatial dependence. Based on this result, the LM tests indicate that there is enough evidence of exogenous spatial lag but not sufficient for spatial error dependence. This indicate that SDM model is competent as well.

To verify which of the models is more competent, the LR tests between SLM vs SARAR and SLM vs SDM are presented in Table 6. In both tests, there is no strong evidence that $\lambda \neq 0 \wedge \theta \neq 0$, suggesting that the most competent model is SLM.

Table 4: Moran's I and LM Tests from No-Spatial Model with W_k.

	OLS			**OLS w/ FE**		
Test	Statistic	df	p-value	Statistic	df	p-value
Spatial error:						
Moran's I	3.348	1	0.001 ***	3.413	1	0.001 ***
Lagrange multiplier	9.989	1	0.002 ***	10.462	1	0.001 ***
Robust Lagrange multiplier	3.601	1	0.058 *	2.455	1	0.117
Spatial lag:						
Lagrange multiplier	16.159	1	0.000 ***	16.272	1	0.000 ***
Robust Lagrange multiplier	9.772	1	0.002 ***	8.265	1	0.004 ***

Note: Results obtained from an OLS regression. Statistical significance levels: * p<0.1, ** p<0.05, *** p<0.01.

We repeat this process for $\mathbf{W_d}$. Beginning with the linear model estimated using OLS

Table 5: Moran's I and LM Tests from SLX Model with W_k.

Test	Statistic	df	p-value
Spatial error:			
Moran's I	3.652	1	0.000 ***
Lagrange multiplier	7.491	1	0.006 ***
Robust Lagrange multiplier	1.831	1	0.176
Spatial lag:			
Lagrange multiplier	8.642	1	0.003 ***
Robust Lagrange multiplier	2.982	1	0.084 *

Note: Results obtained from a ML regression. Statistical significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 6: Likelihood-Ratio (LR) Tests with W_k.

Test Comparison	LR Chi2	p-value
SLM vs. SARAR	-3.24	1.00
SLM vs. SDM	10.82	0.70

(non-spatial model), Table 10 (see Appendix) presents the results of Moran's I test. The test is statistically significant at the 1% level for both the model with and without fixed effects, providing strong evidence of spatial dependence on the proportion of improved seed adoption. Given these results, we proceed with the Lagrange Multiplier (LM) tests, also shown in Table 10 (see Appendix). The results indicate significant spatial lag dependence ($\rho \neq 0$) (statistically significant at 5% level), but no sufficient evidence to confirm spatial error dependence ($\lambda = 0$). These findings suggest that, at this stage, the Spatial Lag Model (SLM) is the most appropriate specification.

In Table 11 (see Appendix), we repeat these tests for the SLX model, again finding strong evidence of spatial dependence. Based on this result, the LM tests indicate significant exogenous spatial lag dependence, suggesting that the Spatial Durbin Model (SDM) is also a viable option. To determine which model is the most appropriate, we perform a Likelihood Ratio (LR) test comparing SLM and SDM, with results presented in Table 12 (see Appendix). The findings provide strong evidence that the SLM model is the most suitable specification when using the $\mathbf{W_d}$ spatial weight matrix.

5.2.1 SLM Model - Peer Effect

To examine whether and how spatial interactions between neighboring communities influence the proportion of improved seeds adopted, we estimate Equation (7) using the Spatial Lag Model (SLM). Based on the findings from the previous section, where the SLM was identified as the most suitable model, we estimate it using Maximum Likelihood (ML).

$$y_{it} = Wy_{it}\rho + X_{it}\beta + \gamma_t + \alpha_i + u_{it} \tag{7}$$

In equation (7), y_{it} is the proportion of improved seeds adopted by a community i at year t. X_{it} represents a matrix of the community's characteristics. The endogenous spatial lag is specify as Wy_{it}, where ρ is the parameter to be estimated as peer effect. γ is a community fixed effect, α is a time fixed effect, and u_{it} refers to the error term.

To draw general conclusions, Table 7 presents results from regressing equation (7) under both spatial weight matrices ($\mathbf{W_k}$, $\mathbf{W_d}$) as well as the non-spatial model estimated via Ordinary Least Squares (OLS). In spatial econometrics, standard estimation methods such as OLS are unsuitable for spatial models. If the Spatial Lag Model (SLM) were estimated using OLS, the parameters would be biased and inconsistent due to the presence of spatial dependence (Anselin, 1988). To address this issue, the literature has developed alternative estimation methods, including parametric, non-parametric, and semi-parametric approaches. Among these, Maximum Likelihood (ML) is the most widely used and will be employed in this study.

Columns (1) and (2) present the estimated parameters of the non-spatial model using OLS, both without and with fixed effects, respectively. However, we already know that this model is biased, as it fails to account for spatial dependence in the error term, leading to inconsistent estimates. Column (3) presents the estimated parameters of the SLM model under the k-nearest neighbor spatial weight matrix ($\mathbf{W_k}$). The results clearly indicate the presence of positive spatial dependence, meaning that the proportion of improved seed adoption in one community is influenced by adoption levels in neighboring communities. Specifically, a 10% increase in improved seed adoption in neighboring communities leads to a 1.81 percentage point average increase in local adoption, highlighting the impact of spatial spillovers in the diffusion of improved seeds. This results is statistically significant at a 5% level. As variables that have a statistical significance effect over the proportion of improved seeds adoption, those communities with a higher proportion of head of households with university degree education tend to increase the proportion of improved seed adoption. Regarding financial access, an increase of seeds bought with coupons and credit by 10% increases in average a 6.13% and 1.51% of improved seed adoption respectively. Column (4) presents, as well, the estimated parameters of the SLM model but under the inverse arc-distance spatial weight matrix ($\mathbf{W_d}$). Just like in the previous model, the results clearly indicate the presence of positive spatial dependence, meaning that the proportion of improved seed adoption in one community is influenced by adoption levels in neighboring communities. Specifically, a 10% increase in improved seed adoption in neighboring communities leads to a 2.05 percentage point average increase in local adoption, highlighting the impact of spatial spillovers in the diffusion of improved seeds. This results is statistically significant at a 5% level. Regarding education levels among household heads, it is noteworthy that an increase in the proportion of uneducated household heads is associated with higher adoption rates. This effect is even more pronounced when primary education is completed and becomes strongest with an increase in the proportion of household heads with a university degree. Regarding financial access, similar to the previous model, more access to credit and coupons enhance improved seed adoption.

When analyzing the adoption of improved seeds across communities, spatial interactions play a crucial role in shaping outcomes. The SLM model accounts for these dependencies by incorporating a spatially lagged variable, allowing us to measure not only the direct impact of explanatory factors within a community but also the indirect spillover effects from neighboring communities. The estimated indirect effects of the independent explanatory variables should eventually be used to test the hypothesis of whether or not spatial spillovers exists (Elhorst, 2014). Indirect effects measure the impact of an explanatory variable in neighboring communities and how it feeds back into the focal community through spatial interactions. If the indirect effect is significant, it suggests that changes in a variable in one community also influence the adoption behavior in nearby communities. This is crucial for understanding the diffusion process of improved seed adoption.

Table 8 presents the direct, indirect and total effects of the SLM models that capture both local influences and broader regional spillovers, providing a more comprehensive view of how different factors drive improved seed adoption. For the SLM ($\mathbf{W_k}$) model, the only variable with a significant and positive indirect effect is the proportion of seeds obtained through coupons. This indicates that seed subsidies or distribution programs in one community also impact neighboring communities, likely through informal seed exchanges, knowledge sharing, or demonstration effects.

5.2.2 SLM Model - Peer Effect under Drought Shock

To examine how drought shocks have a direct effect over the proportion of improved seeds adopted and how that drought shock might affect how the spatial interactions between neighboring communities influence the adoption, we estimate equation (8) using the Spatial Lag Model (SLM).

$$y_{it} = Wy_{it}\rho + DroughtShock\beta_1 + (DroughtShock * Wy_{it})\beta_2 + X_{it}\beta_3 + \gamma_t + \alpha_i + u_{it} \qquad (8)$$

In this equation, $DroughtShock$ is a binary variable that takes the value of 1 if the community i in the previous long rainy season experienced severe drought shock. Otherwise, it takes the value of zero, being β_1 the parameter of interest. In addition, $(DroughtShock * Wy_{it})$ is an interaction effect capturing how the drought shock modifies the influence neighboring communities have over the proportion of adoption. γ is a community fixed effect, α is a time fixed effect, and u is the error term.

Table 7: Regression Estimates: No-Spatial and SLM models.

	No-Spatial	**No-Spatial**	**SLM ($\mathbf{W_k}$)**	**SLM ($\mathbf{W_d}$)**
Main Variables	(1)	(2)	(3)	(4)
prop_female_head	-0.153*	-0.072	-0.073	-0.073
	(0.087)	(0.097)	(0.080)	(0.081)
mean_age_head	0.003	0.000	0.000	0.000
	(0.002)	(0.003)	(0.003)	(0.003)
prop_salaried_head	0.183**	0.014	0.007	0.016
	(0.080)	(0.090)	(0.075)	(0.075)
prop_head_edu_1	0.180*	0.182*	0.178**	0.185**
	(0.095)	(0.094)	(0.078)	(0.079)
prop_head_edu_2	0.277**	0.243**	0.245**	0.241**
	(0.121)	(0.120)	(0.100)	(0.100)
prop_head_edu_3	0.075	0.052	0.063	0.063
	(0.132)	(0.132)	(0.110)	(0.110)
prop_head_edu_4	0.180	0.020	0.040	0.041
	(0.151)	(0.168)	(0.140)	(0.140)
prop_head_edu_5	0.300	0.081	0.114	0.114
	(0.188)	(0.187)	(0.156)	(0.157)
prop_head_edu_6	0.724***	0.521**	0.568***	0.584***
	(0.186)	(0.230)	(0.193)	(0.195)
prop_head_edu_7	0.053	-0.036	0.040	0.007
	(0.249)	(0.335)	(0.281)	(0.280)
total_plot_size	0.000	0.002	0.001	0.002
	(0.002)	(0.003)	(0.002)	(0.002)
prop_coupon	0.550***	0.626***	0.613***	0.624***
	(0.159)	(0.205)	(0.171)	(0.171)
prop_credit	0.127	0.148	0.151*	0.146*
	(0.086)	(0.096)	(0.080)	(0.080)
prop_left_seeds	0.049	0.083	0.090	0.086
	(0.093)	(0.099)	(0.082)	(0.082)
Spatial Effects				
ρ	-	-	0.181**	0.205**
			(0.077)	(0.103)
Variance				
σ_e^2	0.201***	0.183***	0.030***	0.031***
	(0.002)	(0.002)	(0.002)	(0.002)
Fixed Effects	No	Yes	Yes	Yes
Observations	380	380	380	380
Log-Likelihood	-	-	123.114	122.40
AIC	-6.20	-211.03	-214.22	-212.80

Note: (1) Presents the No-Spatial model (OLS). (2) Presents the No-Spatial model (OLS) with Fixed Effects. (3) Presents the ML coefficients of the SLM Model with $\mathbf{W_k}$. (4) Presents the ML coefficients of the SLM Model with $\mathbf{W_d}$. Robust standard errors in parenthesis. Statistical significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 8: Direct, Indirect, and Total Effects of Spatial Models

	SLM ($\mathbf{W_k}$)			SLM ($\mathbf{W_d}$)		
Variable	Direct	Indirect	Total	Direct	Indirect	Total
prop_female_head	-0.070	-0.016	-0.086	-0.071	-0.019	-0.090
	(0.083)	(0.021)	(0.102)	(0.083)	(0.027)	(0.106)
mean_age_head	0.0002	0.00004	0.0002	0.0004	0.0001	0.0005
	(0.0028)	(0.0007)	(0.0034)	(0.0028)	(0.0008)	(0.0036)
prop_salaried_head	0.015	0.003	0.018	0.023	0.007	0.030
	(0.072)	(0.017)	(0.088)	(0.072)	(0.022)	(0.091)
prop_head_edu_1	0.178**	0.039	0.216**	0.184**	0.049	0.233**
	(0.076)	(0.027)	(0.095)	(0.076)	(0.038)	(0.103)
prop_head_edu_2	0.243**	0.053	0.296**	0.239**	0.063	0.302**
	(0.096)	(0.036)	(0.121)	(0.096)	(0.048)	(0.129)
prop_head_edu_3	0.063	0.014	0.077	0.063	0.016	0.079
	(0.103)	(0.026)	(0.127)	(0.104)	(0.034)	(0.133)
prop_head_edu_6	0.581***	0.127	0.708***	0.596***	0.160	0.756***
	(0.188)	(0.078)	(0.239)	(0.190)	(0.115)	(0.266)
prop_head_edu_7	0.042	0.012	0.054	0.009	0.005	0.014
	(0.278)	(0.068)	(0.340)	(0.278)	(0.085)	(0.354)
total_plot_size	0.0014	0.0003	0.0017	0.0016	0.0004	0.0020
	(0.0022)	(0.0006)	(0.0027)	(0.0022)	(0.0007)	(0.0028)
prop_coupon	0.630***	0.137*	0.766***	0.641***	0.169	0.810***
	(0.180)	(0.079)	(0.230)	(0.181)	(0.114)	(0.252)
prop_credit	0.152*	0.033	0.185*	0.146*	0.038	0.184*
	(0.081)	(0.025)	(0.100)	(0.081)	(0.032)	(0.105)
prop_left_seeds	0.095	0.021	0.116	0.091	0.024	0.116
	(0.080)	(0.022)	(0.099)	(0.081)	(0.027)	(0.103)

Note: Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 9: Regression Estimates under Drought Shock.

	SLM ($\mathbf{W_k}$)	**SLM ($\mathbf{W_d}$)**
Main Variables	(1)	(2)
Severe Drought Shock	-0.142**	-0.242**
	(0.057)	(0.094)
Severe Drought Shock * Wy	0.392**	0.726**
	(0.193)	(0.289)
Spatial Effects		
ρ	0.170***	0.202***
	(0.063)	(0.075)
Variance		
σ_e^2	0.030***	0.030***
	(0.005)	(0.005)
Fixed Effects	Yes	Yes
Observations	380	380

Note: Clustered standard errors in parentheses. The complete table with characteristic's estimates are presented in Table 13 at the Appendix Section. Statistical significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 9 presents the results from regressing equation (8) under both spatial weight matrices ($\mathbf{W_k}$, $\mathbf{W_d}$). Column (1) presents the estimated parameters of the SLM model using the k-nearest neighbors spatial weight matrix. The results indicate that experiencing a drought shock in the previous crop season significantly affects the proportion of improved seed adoption. Specifically, a severe drought shock leads to a 14.2 percentage point decrease in improved seed adoption, with statistical significance at the 5% level. This pattern is consistent in Column (2), which estimates the SLM model using the inverse arc-distance spatial weight matrix. Under this specification, a severe drought shock results in an even larger reduction, decreasing the proportion of improved seed adoption by 24.2 percentage points on average. Building on these results, we find that experiencing a severe drought shock in the previous crop season significantly strengthens the influence of neighboring communities on the proportion of improved seed adoption, with statistical significance at the 5% level. Specifically, in Column (1), a severe drought shock leads to a 39.2 percentage point increase in the proportion of improved seeds adopted on average. A similar pattern is observed in Column (2), where the effect is even stronger, with an average increase of 72.6 percentage points. These findings suggest that after experiencing a drought shock during a critical crop period, communities tend to rely more on neighboring communities' decisions when making their own adoption choices.

6 Robustness Check

La idea de esta sección es validar mis resultados, principalmente aquellos causales (drought shock). Ya el usar otra matriz de pesos espaciales (W_d) valida mis primeros resultados. Para el shock de sequía puedo hacer un leave-one-out, o usar otro indice que mida la sequía. Puedo hacer un Falsification Test, para ver que el efecto del shock sea del año en que pasa y no de otro año.

7 Conclusion

A Appendix

Table 10: Moran's I and LM Tests from No-Spatial Model with W_d.

	OLS			OLS w/ FE		
Test	Statistic	df	p-value	Statistic	df	p-value
Spatial error:						
Moran's I	3.784	1	0.000 ***	3.802	1	0.000 ***
Lagrange multiplier	12.567	1	0.000 ***	12.799	1	0.000 ***
Robust Lagrange multiplier	1.410	1	0.235	1.063	1	0.302
Spatial lag:						
Lagrange multiplier	17.258	1	0.000 ***	17.428	1	0.000 ***
Robust Lagrange multiplier	6.100	1	0.014 **	5.692	1	0.017 **

Note: Results obtained from an OLS regression. Statistical significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 11: Moran's I and LM Tests from SLX Model with W_d.

Test	Statistic	df	p-value
Spatial error:			
Moran's I	1.969	1	0.049 *
Lagrange multiplier	0.668	1	0.414
Robust Lagrange multiplier	2.352	1	0.125
Spatial lag:			
Lagrange multiplier	1.466	1	0.226
Robust Lagrange multiplier	3.151	1	0.076 *

Note: Results obtained from an ML regression. Statistical significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 12: Likelihood-Ratio (LR) Tests with W_d.

Test Comparison	LR Chi2	p-value
SLM vs. SARAR	-1.34	1.00
SLM vs. SDM	12.31	0.42

Table 13: Regression Estimates under Drought Shock.

	SLM ($\mathbf{W_k}$)	**SLM ($\mathbf{W_d}$)**
Main Variables	(1)	(2)
Severe Drought Shock	-0.142**	-0.242**
	(0.057)	(0.094)
Severe Drought Shock * Wy	0.392**	0.726**
	(0.193)	(0.289)
prop_female_head	-0.074	-0.065
	(0.148)	(0.149)
mean_age_head	0.001	0.001
	(0.005)	(0.005)
prop_salaried_head	-0.007	0.004
	(0.106)	(0.106)
prop_head_edu_1	0.175	0.186
	(0.149)	(0.148)
prop_head_edu_2	0.233	0.246
	(0.166)	(0.167)
prop_head_edu_3	0.070	0.091
	(0.208)	(0.205)
prop_head_edu_4	0.054	0.060
	(0.215)	(0.215)
prop_head_edu_5	0.163	0.177
	(0.175)	(0.178)
prop_head_edu_6	0.576*	0.603**
	(0.301)	(0.305)
prop_head_edu_7	0.090	0.076
	(0.297)	(0.311)
total_plot_size	0.002	0.002
	(0.002)	(0.002)
prop_coupon	0.602***	0.599***
	(0.160)	(0.161)
prop_credit	0.157	0.156
	(0.100)	(0.100)
prop_left_seeds	0.081	0.073
	(0.093)	(0.093)
Spatial Effects		
ρ	0.170***	0.202***
	(0.063)	(0.075)
Variance		
σ_e^2	0.030***	0.030***
	(0.005)	(0.005)
Fixed Effects	Yes	Yes
Observations	380	380

Note: Clustered standard errors in parentheses. Statistical significance levels: * p<0.1, ** p<0.05, *** p<0.01.

References

Acquaah, G. (2009). *Principles of plant genetics and breeding*. John Wiley & Sons.

Ahumada, H. A., Gabrielli, M. F., Herrera Gomez, M. H., and Sosa Escudero, W. (2018). Una nueva econometría: Automatización, big data, econometría espacial y estructural.

Anselin, L. (1988). Model validation in spatial econometrics: a review and evaluation of alternative approaches. *International Regional Science Review*, 11(3):279–316.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics*, 27(3):247–267.

Arslan, A., Belotti, F., Asfaw, S., Karfakis, P., and Lipper, L. (2016). Welfare impacts of climate shocks: evidence from Tanzania.

Barrios, S., Ouattara, B., and Strobl, E. (2008). The impact of climatic change on agricultural production: Is it different for Africa? *Food policy*, 33(4):287–298.

Beguería, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B. (2014). Standardized precipitation evapotranspiration index (spei) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *International Journal of Climatology*, 34(10):3001–3023.

Bozzola, M., Smale, M., and Di Falco, S. (2018). Climate, shocks, weather and maize intensification decisions in rural kenya. In *Agricultural Adaptation to Climate Change in Africa*, pages 107–128. Routledge.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4):281–298.

CIAT and World Bank (2018). Climate-smart agriculture in malawi. CSA Country Profiles for Africa Series. International Center for Tropical Agriculture (CIAT); Washington, D.C. 30 p.

Elhorst, J. P. (2014). Linear spatial dependence models for cross-section data. *Spatial econometrics: From cross-sectional data to spatial panels*, pages 5–36.

Engel, J., Hirano, Y., Nyirenda, Y. L., Nur, H. D. A., Chilima, E. Z., Thapa, D., Botha, B. N., and Eliste, P. (2022). Malawi economic monitor: Planning beyond the next harvest, advancing economic stability and agricultural commercialization. Disclosed on December 8, 2022.

Fang, D. and Richards, T. J. (2018). New maize variety adoption in Mozambique: a spatial approach. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 66(3):469–488.

FAO (2008). Climate change adaptation and mitigation: Challenges and opportunities for food security. *Report from high-level conference on world food security: The challenges of climate change and bioenergy.*

Fehr, W. (1991). Principles of cultivar development: theory and technique.

Gross, M., Muller, M., Caruso, G. D., and Cardona Sosa, L. M. (2021). Poverty and gender in malawi. World Bank Publications - Reports 37070, The World Bank Group.

Hossain, M., Nayak, S., Mohapatra, S., Trivedi, P., and Waza, S. A. (2024). Women-led community institutions as a potential vehicle for the adoption of varieties and improved seed practices: an impact case from India. *CABI Agriculture and Bioscience*, 5(1):53.

Kafle, B. (2010). Determinants of adoption of improved maize varieties in developing countries: A review. *International Research Journal of Applied and Basic Sciences*, 1(1):1–7.

Kalinda, T., Tembo, G., Kuntashula, E., and Lusaka, Z. (2014). Adoption of improved maize seed varieties in southern Zambia. *Asian Journal of Agricultural Sciences*, 6(1):33–39.

LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Li, F., Feng, S., Lu, H., Qu, F., and D'Haese, M. (2021). How do non-farm employment and agricultural mechanization impact on large-scale farming? a spatial panel data analysis from Jiangsu Province, China. *Land Use Policy*, 107:105517.

Makate, C., Angelsen, A., Holden, S. T., and Westengen, O. T. (2023). Rainfall shocks and inequality have heterogeneous effects on farmers' seed purchase decisions in east africa. *Climate Risk Management*, 42:100556.

Martey, E., Etwire, P. M., and Kuwornu, J. K. (2020). Economic impacts of smallholder farmers' adoption of drought-tolerant maize varieties. *Land use policy*, 94:104524.

McGuire, S. and Sperling, L. (2016). Seed systems smallholder farmers use. *Food security*, 8:179–195.

McKee, T. B., Doesken, N. J., Kleist, J., et al. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, volume 17, pages 179–183. California.

National Statistical Office (2020). The fourth integrated household panel survey (ihps4) 2019 report. Technical report, National Statistical Office, Zomba, Malawi.

Nordhagen, S. and Pascual, U. (2013). The impact of climate shocks on seed purchase decisions in malawi: implications for climate change adaptation. *World Development*, 43:238–251.

NSO (2021). Malawi poverty report 2020.

Santiago Beguería, Borja Latorre, F. R. S. M. V.-S. (2023). About the SPEI. Accessed: 23 December 2024.

Solomon Gebrechorkos, Jing Peng, E. D. D. G. M. (2023). Hydro-jules: Global high-resolution drought datasets from 1981-2022. *NERC EDS Centre for Environmental Data Analysis*. 07 July 2023.

Sperling, L., Cooper, H. D., and Remington, T. (2008). Moving towards more effective seed aid. *The Journal of Development Studies*, 44(4):586–612.

Stevens, T. and Madani, K. (2016). Future climate impacts on maize farming and food security in malawi. *Scientific reports*, 6(1):36241.

Svoboda, M., Hayes, M., and Wood, D. (2012). Standardized precipitation index: user guide.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46(sup1):234–240.

World Bank (2022). Malawi economic monitor: Strengthening agricultural commercialization and rural labor markets.

Zheng, H., Ma, W., and Li, G. (2021). Learning from neighboring farmers: Does spatial dependence affect adoption of drought-tolerant wheat varieties in China? *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 69(4):519–537.