Universidad de San Andrés

Department of Economics

Master in Economics

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

Author: Juan Segundo Zapiola

Mentor: Marcos Herrera Gómez

August 16th, 2025

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. To address the second question, the analysis incorporates climate variability using the Standardized Precipitation Evapotranspiration Index (SPEI), which allows us to examine how drought shocks alter the dynamics of peer effects in the adoption process.

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

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; Läpple and Kelley, 2015; Singha, 2019; Zheng et al., 2021; Ward and Pede, 2015). 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 weight 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.

These social learning dynamics and interactions with neighboring communities are commonly referred to as peer effects. As defined by Bramoullé et al. (2020), peer effects occur when individuals are influenced by others across various dimensions and contexts, meaning that the behavior, choices, or outcomes of peers affect one's own decision-making process. In the context of agricultural technology adoption, peer effects can emerge through mechanisms such as social learning, conformity to local norms, or social pressure, where farmers adapt their behavior based on what others in their community are doing. However, identifying causal peer effects poses a significant empirical challenge. The reflection problem, as outlined by Manski (1993), arises when it becomes difficult to determine whether individuals are truly influencing each other, reacting to a common external factor, or simply exhibiting similar behavior due to self-selection into the same environment. To address this challenge, the present study applies spatial econometric techniques, which allow us to control for unobserved spatial correlation and isolate the effect of neighboring communities on local adoption decisions. By explicitly modeling spatial interactions, we aim to provide more credible evidence on peer effects and gain a deeper understanding of how these dynamics shape the diffusion of improved seed varieties.

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 dependence effect when a community follows the practices of its neighbors, and a negative spatial dependence 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 focuses on Malawi, a representative case within Sub-Saharan Africa where agriculture plays a central role in economic development and poverty alleviation. We aim to: (1) assess how spatial interactions with neighboring communities influence the proportion of improved seed varieties adopted, and (2) examine how drought shocks shape the dynamics of peer effects on adoption, using georeferenced panel data from 2010 to 2019.

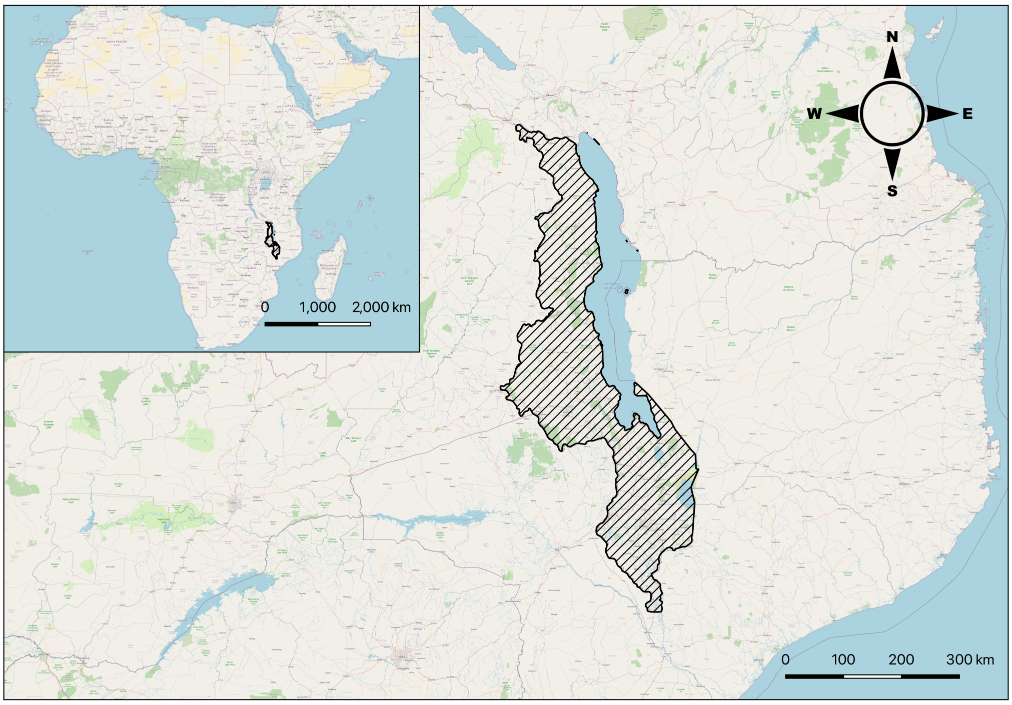
The key contribution of this study lies in the explicit use of spatial econometric techniques to identify and quantify peer effects in the adoption of improved seed varieties—something that has not yet been explored in the context of Malawi. While previous studies have acknowledged the role of social networks and neighborhood effects in technology adoption, they have largely relied on non-spatial methods, such as regression models with interaction terms or qualitative case studies Makate et al. (2023); Kalinda et al. (2014); Hossain et al. (2024). In contrast, this study uses a Spatial Lag Model (SLM) to formally capture spatial dependence in adoption behavior across georeferenced communities. A novel aspect of this study is its focus on community-level dynamics, rather than individual farmer decisions (Fang and Richards, 2018; Läpple and Kelley, 2015; Li et al., 2021; Zheng et al., 2021; Skevas et al., 2018), allowing us to explore how the behavior of neighboring communities influences collective adoption outcomes. Furthermore, this study is among the first to examine how climatic shocks—specifically droughts—interact with spatial effects, providing new insights into how environmental stress reshapes local decision-making processes through peer influence.

The study is organized as follows. Section 2 provides background information on Malawi, highlighting its agricultural context and vulnerability to climate shocks. Section 3 presents the methodological framework, including the construction of spatial weight matrices, testing for spatial dependence, and the estimation strategy using spatial econometric models. Section 4 describes the data source and section 5 presents the descriptive statistics. Section 6 reports the main results, including model selection tests to assess the presence of additional spatial effects and the estimation of peer effects on improved seed adoption. Section 7 presents further results, examining whether drought shocks shape the dynamics of peer effects on adoption. Section 8 presents robustness checks to validate the main findings. Finally, Section 9 concludes with a summary of the results and policy implications.

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.[1]



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

[1] Own elaboration using QGIS.

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 used to analyze spatial dependence in the adoption of improved seed varieties across rural communities in Malawi. The analysis is based on spatial panel econometric techniques, which are well-suited to identify how adoption decisions in one community are influenced by those in neighboring areas over time.

We begin by presenting the spatial econometric model used to study how spatial interactions with neighboring communities influence the proportion of improved seed varieties adopted. Next, we construct spatial weight matrices to formally capture the geographic relationships between communities, and apply Moran's I test to assess the presence of spatial autocorrelation. Finally, we introduce a broader set of spatial econometric models to explore whether additional spatial components should be incorporated into the analysis of improved seed adoption. To determine the most appropriate model specification, we conduct Lagrange Multiplier (LM) and Likelihood Ratio (LR) tests, which help identify the presence of endogenous, exogenous, or error-based spatial dependence. Based on the results of these tests, we estimate the selected model using Maximum Likelihood, allowing us to formally quantify peer effects and capture spatial spillovers in the adoption process.

3.1 Empirical Framework - SLM Model

The decision of a farmer to adopt an improved seed variety is ultimately an investment decision, often made under uncertainty. In making this choice, the behavior of neighboring farmers plays a crucial role: what others decide influences, directly or indirectly, the farmer's own willingness to adopt. These spatial interactions, whether through observation, shared information, or social influence, shape the adoption decision, introducing what is commonly referred to as a peer effect. Based on this idea, we assume that the decision to adopt an improved seed variety is part of a spatially dependent process. To formally account for this spatial dependence, we use a Spatial Lag Model (SLM), which is specified as follows:

$$y_i = Wy_i\rho + X_i\beta + u_i \tag{1}$$

where y_{it} is the proportion of improved seeds adopted by a community i. X_i represents a matrix of the community's characteristics. The endogenous spatial lag is specify as Wy_i, where ρ is the parameter to be estimated as peer effect. The u_{it} refers to the error term.

The decision to use the Spatial Lag Model (SLM) is supported by previous studies such as Fang and Richards (2018), Singha (2019), and Wollni and Andersson (2014), which also examine peer effects and spatial interactions in agricultural technology adoption. These studies demonstrate the relevance of capturing spatially lagged dependent variables to model decision-making influenced by neighboring agents. Other research in the field has employed more general spatial econometric models, such as the Spatial Durbin Model (SDM) and the SARAR model, which incorporate additional spatial components like exogenous spatial lags or spatially correlated error terms (Zheng et al., 2021; Läpple and Kelley, 2015; Skevas et al., 2018; Ward and Pede, 2015). To determine whether such additional spatial effects are present in our context, we conduct a series of specification tests to assess whether these alternative models provide a better fit than the baseline SLM.

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); Ward and Pede (2015), it is measured as

$$\mathbf{W_d} = w_{ij} = \begin{cases} 0 & \text{if } \; i = j \\ 1/d_{ij} & \text{if } \; d_{ij} \leq D \\ 0 & \text{if } \; d_{ij} > D \end{cases} \tag{2}$$

where d_{ij} is the arc-distance between the centroids of community i and the neighboring 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 (2). The critical distance cut-off D to be used would be 100 km.

Additionally, to test the robustness of our results to different definitions of neighborhood proximity, we construct alternative versions of the matrix $\mathbf{W_d}$ using buffer distances of 80 km and 120 km. These allow us to evaluate how sensitive the model results are to the choice of spatial threshold, while maintaining the same inverse-distance structure.

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:

$$\text{Moran's I}_t = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{S^2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \tag{3}$$

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

This section explores whether additional spatial components should be incorporated into the analysis of improved seed adoption. To do so, we consider a range of spatial econometric models that allow for different types of spatial interactions. These models are evaluated using formal specification tests, which help determine the most appropriate structure for capturing spatial dependence in the data.

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.

$$y_{it} = Wy_{it}\rho + X_{it}\beta + WX_{it}\theta + u_{it}; \quad u_{it} = \lambda W u_{it} + \epsilon_{it} \tag{4}$$

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 explore whether additional spatial components should be incorporated, 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 uses a georeferenced dataset, the Malawi Integrated Household Panel Survey, which includes detailed information on household, community, and agricultural characteristics, as well as data on the use of improved seeds. This dataset is described in more detail in the following section, 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].

5 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 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 and Richards (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.

[2] See: Malawi Integrated Household Panel Survey, Agriculture and Fishery Enumerator Manual (ANNEX 3)
[3] Own elaboration using QGIS.

Figure 2: Map of Malawi with the Enumeration Areas (EAs)/ Communities. [3]

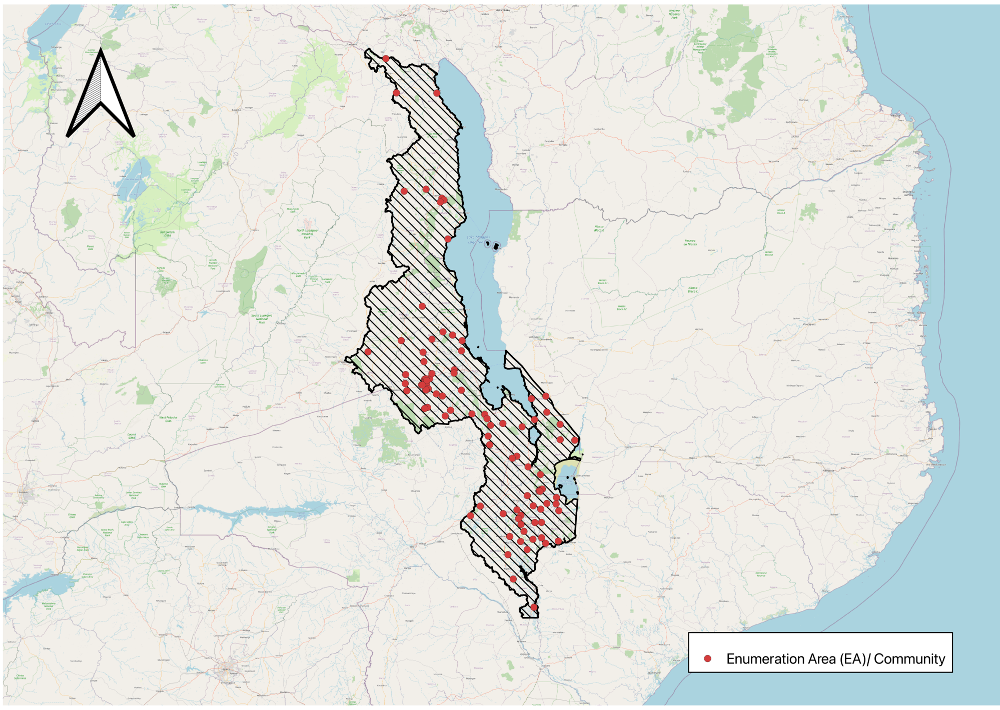


Table 1: Descriptive Statistics of Key Variables.

Variable	Description	Mean	SD
Adoption Var.			
Prop. imp.	Proportion of improved seed used	0.34	0.25
Household Var.			
Prop. female head	Proportion of female head in households	0.25	0.20
Mean age head	Mean age of household head	44.73	6.17
Prop. salaried head	Proportion of being salaried employed	0.20	0.25
Prop. head edu. 1	Proportion of head with no education	0.54	0.29
Prop. head edu. 2	Proportion of head with PSLC education	0.11	0.17
Prop. head edu. 3	Proportion of head with JCE education	0.09	0.16
Prop. head edu. 4	Proportion of head with MSCE education	0.07	0.15
Prop. head edu. 5	Proportion of head with non-university education	0.03	0.12
Prop. head edu. 6	Proportion of head with university degree education	0.01	0.08
Prop. head edu. 7	Proportion of head with post-grad degree education	0.01	0.05
Agriculture Var.			
Total plot size	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 seeds	Proportion who used left over seeds from previous season	0.43	0.22

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

6 Results

This section presents the findings of the spatial analysis on improved seed adoption in Malawi. We begin by verifying the presence of spatial dependence in the data using Moran's I statistic. Once spatial dependence is confirmed, we assess whether additional spatial effects such as spatial error or exogenous spatial lags should be incorporated into the model. To do this, we apply diagnostic tools including the Lagrange Multiplier (LM) and Likelihood Ratio (LR) tests. Based on the outcomes of these tests, we identify the most appropriate model specification. We then proceed with the estimation of the model, which allows us to formally quantify peer effects across communities and examine how neighboring behavior influences the adoption of improved seeds.

6.1 Testing for Additional Spatial Effects

We begin this section by defining our baseline specification: the Spatial Lag Model (SLM). This model assumes that the adoption of improved seeds in a given community is directly influenced by the adoption behavior of neighboring communities. Formally, the SLM includes a spatially lagged dependent variable, represented by ρWy, where W is the spatial weight matrix and ρ is the spatial autoregressive parameter capturing peer effects.

To evaluate whether additional spatial effects, such as spatial error dependence (λ) or exogenous spatial lags (θ), should be incorporated, we follow the specific-to-general (STGE) framework, as described in Section 3.3. This strategy begins with a simpler model and progressively tests more complex, nested alternatives. The full set of spatial models is estimated using $\mathbf{W}_d$ (inverse arc-distance of 100 km).

To detect spatial dependence in the baseline (non-spatial) model, we first compute Moran's I statistic, as shown in Table 2. The statistic is significant at the 1% level, indicating the presence of spatial autocorrelation in the proportion of improved seed used. The result holds both with and without fixed effects.

We then conduct Lagrange Multiplier (LM) tests to determine the nature of this spatial dependence. The LM tests, also presented in Table 2, provide strong evidence of spatial lag dependence ($\rho \neq 0$) at the 1% level. However, there is no sufficient evidence of spatial error dependence ($\lambda = 0$), particularly in the fixed effects specification. Based on these findings, the SLM appears to be the most appropriate model at this stage.

To explore whether exogenous spatial lags should also be considered, we test the SLX model, which imposes $\rho = 0$ and $\lambda = 0$. Results from Moran's I and LM tests for this model (Table 3) also suggest significant spatial dependence. The LM tests indicate the presence of exogenous spatial lag dependence ($\theta \neq 0$), suggesting that the Spatial Durbin Model (SDM) may also be a viable alternative.

Finally, to determine whether the more general models (SARAR or SDM) offer a better fit than the SLM, we conduct Likelihood Ratio (LR) tests (Table 4). In both comparisons, SLM vs SARAR and SLM vs SDM, we find no strong evidence that including λ or θ significantly improves the model fit. Thus, we conclude that the SLM is the most competent specification for capturing spatial dynamics in improved seed adoption.

Table 2: 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 3: 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 4: Likelihood-Ratio (LR) Tests with W_d.

Test Comparison	LR Chi2	p-value
SLM vs. SARAR	0.38	0.54
SLM vs. SDM	12.31	0.42

6.2 SLM Model - Peer Effect

To examine whether and how spatial interactions between neighboring communities influence the proportion of improved seeds adopted, we estimate Equation (5) 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).

$$y_{it} = Wy_{it}\rho + X_{it}\beta + \gamma_t + \alpha_i + u_{it} \tag{5}$$

In equation (5), 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 5 presents results from regressing equation (5) under the distance-based spatial weight matrices ($\mathbf{W_{d80}}$, $\mathbf{W_{d100}}$, $\mathbf{W_{d120}}$). 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.

Column (1) presents the estimated parameters of the SLM model using the inverse 80 km arc-distance spatial weight matrix ($\mathbf{W_{d80}}$). The results confirm the presence of positive spatial dependence, as indicated by a statistically significant spatial autoregressive parameter (ρ) (1% level), suggesting that higher adoption levels in neighboring communities are associated with higher local adoption of improved seeds. While the coefficients from the ML estimation do not directly reflect marginal effects, their signs and significance provide insight into the direction and relevance of each variable's association with adoption. For example, communities with a higher proportion of household heads holding a university degree are more likely to adopt improved seeds. Similarly, greater access to seed obtained through coupons and agricultural credit is positively associated with adoption: higher shares of seed purchased with coupons and credit are linked to increased adoption rates, highlighting the importance of financial accessibility. Column (2) reports results for the SLM model estimated with a 100 km distance matrix ($\mathbf{W_{d100}}$). Again, positive spatial dependence is evident at a 5% level, reinforcing the conclusion that improved seed adoption in one area is positively associated with adoption in nearby areas. These findings collectively underscore the role of spatial processes and local conditions in shaping adoption behavior, even though the precise magnitudes of effects are better captured in the direct, indirect, and total impact decomposition.

Column (3) presents the estimated parameters of the SLM model under the inverse 120 km arc-distance spatial weight matrix ($\mathbf{W_{d120}}$). Here, even-though that the estimation results has a positive spatial dependence is not statistically significant. One of the possible explanation of this result is that there might be a dilution of spatial influence over distance. As distance band increases, the additional neighbors are less likely to influence a community's seed adoption behavior. This dilutes the strength of the spatial autocorrelation, making ρ harder to estimate precisely (larger robust standard error) and thus statistically insignificant. In addition, When you increase the distance to 120 km, you're potentially connecting units that are not meaningfully related in practice, making the spatial dependence to become weaker.

Regarding education levels among household heads in Column (2) and (3), 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 Spatial Lag Model (SLM) accounts for these dependencies by incorporating a spatially lagged dependent variable, allowing us to distinguish between the direct impact of explanatory factors within a given community and the indirect spillover effects originating from neighboring communities. To better capture and interpret these effects, we decompose the estimated coefficients into direct, indirect, and total effects. This decomposition allows us to test the hypothesis of whether spatial spillovers exist, as suggested by Elhorst (2014). In this context, indirect effects measure how changes in an explanatory variable in one community influence adoption behavior in nearby communities through spatial linkages. If significant, they provide strong evidence of peer influence or regional diffusion mechanisms. The total effects, meanwhile, reflect the overall impact of each variable, combining both local and neighboring influences.

Table 6 presents the direct, indirect and total effects of the SLM model with $\mathbf{W_{d80}}$ that capture both local influences and broader regional spillovers, providing a more comprehensive view of how different factors drive improved seed adoption. As shown in the indirect effect column, the only variables with a significant and positive indirect effect are the proportion of seeds obtained through coupons and the proportion of household heads with a university degree. A 10 percent point increase of households with university-educated heads in community A leads to an average increase of 1.98 percentage points in the proportion of improved seed adoption in neighboring communities, this is statistically significant at a 10% level. In addition, A 10 percent point increase in the proportion of seeds obtained via coupons in one community leads to an average increase of 2.05 percentage points in improved seed adoption in neighbor-

Table 5: Regression Estimates: SLM model with $\mathbf{W_d}$.

	SLM ($\mathbf{W_{d80}}$)	**SLM ($\mathbf{W_{d100}}$)**	**SLM ($\mathbf{W_{d120}}$)**
Main Variables	(1)	(2)	(3)
Prop. female head	-0.084	-0.073	-0.076
	(0.080)	(0.081)	(0.081)
Mean age head	0.001	0.000	0.000
	(0.003)	(0.003)	(0.003)
Prop. salaried head	0.013	0.016	0.016
	(0.074)	(0.075)	(0.075)
Prop. head edu. 1	0.182**	0.185**	0.184**
	(0.078)	(0.079)	(0.079)
Prop. head edu. 2	0.232**	0.241**	0.237**
	(0.100)	(0.100)	(0.101)
Prop. head edu. 3	0.056	0.063	0.055
	(0.109)	(0.110)	(0.110)
Prop. head edu. 4	0.037	0.041	0.033
	(0.139)	(0.140)	(0.140)
Prop. head edu. 5	0.132	0.114	0.101
	(0.156)	(0.157)	(0.157)
Prop. head edu. 6	0.597***	0.584***	0.568***
	(0.193)	(0.195)	(0.195)
Prop. head edu. 7	0.036	0.007	-0.007
	(0.279)	(0.280)	(0.280)
Total plot size	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)
Prop. coupon	0.620***	0.624***	0.638***
	(0.170)	(0.171)	(0.171)
Prop. credit	0.142*	0.146*	0.146*
	(0.080)	(0.080)	(0.080)
Prop. left seeds	0.091	0.086	0.085
	(0.082)	(0.082)	(0.082)
Spatial Effects			
ρ	0.244***	0.205**	0.182
	(0.088)	(0.103)	(0.111)
Variance			
σ_e^2	0.030***	0.031***	0.031***
	(0.002)	(0.002)	(0.002)
Fixed Effects	Yes	Yes	Yes
Observations	380	380	380
Log-Likelihood	124.10	122.40	121.78
AIC	-216.20	-212.80	-209.56

Note: This table presents ML estimates from the Spatial Lag Model (SLM) using distance-based spatial weights matrices: $\mathbf{W_{d80}}$, $\mathbf{W_{d100}}$, and $\mathbf{W_{d120}}$. Robust standard errors in parentheses. Statistical significance: * p<0.1, ** p<0.05, *** p<0.01.

ing communities, also statistically significant at a 10% level. These results indicate that both farmer education and coupon-based interventions generate positive spillover effects. Educated farmers may influence surrounding communities by sharing knowledge, setting examples, or participating in local networks. Similarly, seed subsidies do not only benefit the direct recipients—they also increase adoption in neighboring communities, likely through informal seed sharing, observable yield improvements, or word-of-mouth awareness of the program. This highlights the importance of designing interventions with the potential for regional diffusion, not just localized impact. Notably, these indirect effects are statistically significant under the 80 km spatial weight matrix ($\mathbf{W}_{d80}$), while models with broader distance thresholds—100 km and 120 km—do not show statistically significant spillovers, suggesting that spatial influence might weaken as distance increases [4].

Table 6: Direct, Indirect, and Total Effects of SLM Model with $\mathbf{W_{d80}}$

Variable	Direct	Indirect	Total
Prop female head	-0.081	-0.027	-0.109
	(0.083)	(0.032)	(0.111)
Mean age head	0.0008	0.0003	0.0011
	(0.0028)	(0.0010)	(0.0037)
Prop salaried head	0.0207	0.0068	0.0275
	(0.0718)	(0.0249)	(0.0951)
Prop. head edu. 1	0.182**	0.058	0.240**
	(0.0760)	(0.0383)	(0.1057)
Prop. head edu. 2	0.231**	0.074	0.305**
	(0.0961)	(0.0469)	(0.1311)
Prop. head edu. 3	0.056	0.017	0.074
	(0.1032)	(0.0373)	(0.1375)
Prop. head edu. 4	0.035	0.012	0.047
	(0.1385)	(0.0509)	(0.1859)
Prop. head edu. 5	0.127	0.043	0.170
	(0.1484)	(0.0549)	(0.1974)
Prop. head edu. 6	0.611***	0.198*	0.809***
	(0.1885)	(0.1136)	(0.2701)
Prop. head edu. 7	0.038	0.015	0.053
	(0.2776)	(0.0982)	(0.3692)
Total plot size	0.0016	0.0005	0.0021
	(0.0022)	(0.0008)	(0.0030)
Prop. coupon	0.639***	0.205*	0.844***
	(0.1800)	(0.1116)	(0.2570)
Prop. credit	0.143	0.045	0.189
	(0.0807)	(0.0334)	(0.1083)
Prop. left seeds	0.096	0.031	0.128
	(0.0804)	(0.0309)	(0.1079)

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

[4] *See Tables 11 and 12 at the Appendix Section.*

7 Further Results

We find that higher levels of improved seed adoption in neighboring communities are associated with higher local adoption rates, indicating that spatial interactions play a significant role in shaping adoption decisions. However, an important question remains: how are these peer effects dynamics influenced by external shocks, such as climate-related events? In this section, we examine how drought shocks modify the strength and dynamics of peer effects, providing insights into how environmental stress may alter the diffusion of improved seed technologies across communities.

We begin by introducing the dataset used to construct the drought shock index, which allows us to capture climate variability across communities. Next, we present the descriptive statistics related to drought exposure and seed adoption. Finally, we outline the empirical model used to assess the impact of drought shocks on the proportion of improved seeds adopted, and to examine whether these shocks influence the strength of spatial interactions, that is, how neighboring communities affect each other's adoption behavior under climate stress.

7.1 Weather Data & Drought Index

To address for drought shocks we utilized the Climatic Research Unit (CRU) SPEI database[5]. 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).

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).

[5] SPEI Dataset: https://catalogue.ceda.ac.uk/uuid/ac43da11867243a1bb414e1637802dec/

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 [6]. 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.

7.2 Descriptive Statistics - 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 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.

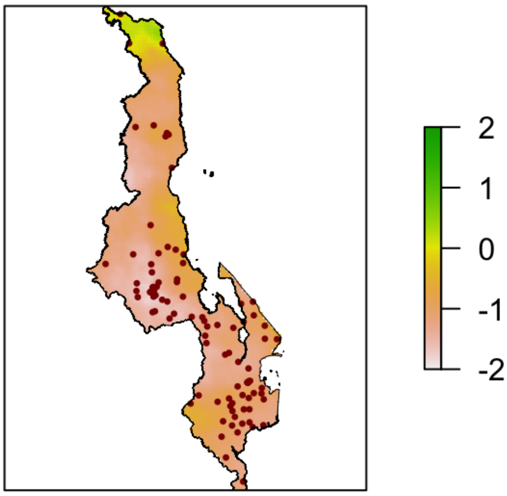
Figure 3 presents the 3-month SPEI drought shocks in Malawi during the key planting period of the previous long rain season for each survey round. The year before the surveys were conducted, some areas experienced severe drought shocks during the key planting period, potentially influencing farmers' 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 7 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 11 (in the Appendix Section) 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

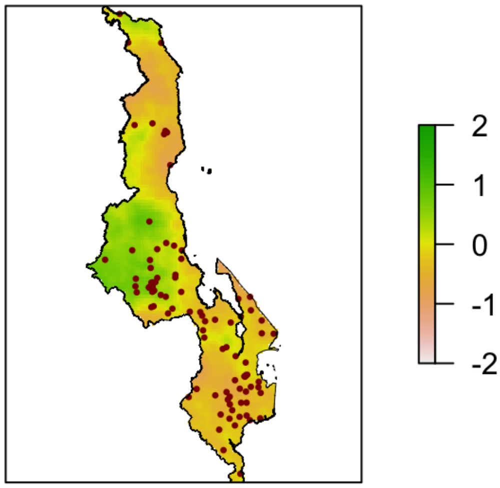
[6] As stated by two agronomic engineers during an interview.

[7] Own elaboration using R-Studio.

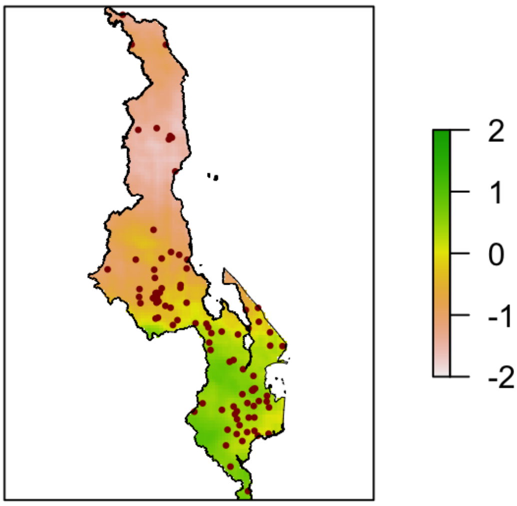
Figure 3: 3-month SPEI Drought Shock in Malawi. [7]



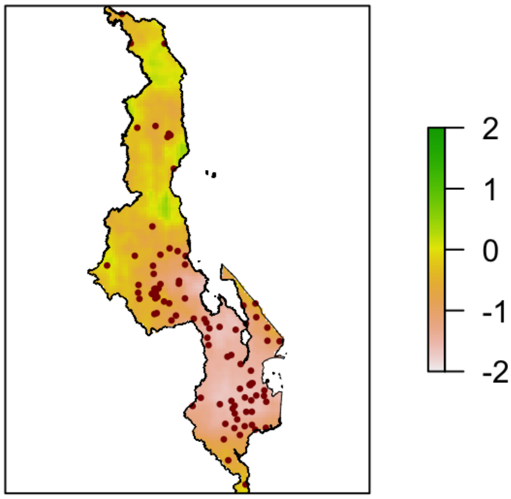
(a) Dec-Feb 2010



(b) Dec-Feb 2013



(c) Dec-Feb 2016



(d) Dec-Feb 2018

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 7: 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

7.3 Instrumental Variables - 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, formally we estimate the following model:

$$y_{it} = Wy_{it}\rho_1 + (DroughtShock * Wy_{it})\rho_2 + X_{it}\beta_1 + \gamma_t + \alpha_i + u_{it} \tag{6}$$

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. 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.

Equation (6) cannot be estimated with the usual methods as there exists an endogenous variable, Wy, that is correlated with the error term. Following the IV model presented by Herrera Gómez (2015), the alternative is to find an instrument H that is relevant on explaining Wy and satisfies the exclusion restriction. That is, the instrument H has to satisfy:

$$plim\left(\frac{1}{n}\right)H'Wy = M_H Wy, \tag{7}$$

$$plim\left(\frac{1}{n}\right)H'u = 0, \tag{8}$$

being $M_H Wy$ a non-singular finite matrix.

As Herrera Gómez (2015) mention, the $E(Wy)$ is linearly related to $WX, W^2X, W^3X, ..., W^qX$, which serve as valid instruments for the endogenous variable Wy. Monte Carlo simulations suggest that using instruments up to the second order ($q = 2$) yields good performance. Therefore, in this study, these two sets of spatial lags of the exogenous variables (WX and W^2X) are used as instruments. Including higher-order terms increases the risk of multicollinearity with the regressors. Applying this idea, we obtain the estimator of instrumental variables for Wy:

$$\hat{\rho} = (\hat{Z}'Z)^{-1}\hat{Z}'y, \tag{9}$$

where $\hat{Z} = P_H\widehat{Wy} = \left[X, \widehat{Wy}\right]$, $\widehat{Wy} = P_H Wy$ and $P_H = H(H'H)^{-1}H'$. Instrumental Variable (IV) estimations are carried out through the Two-Stage Least Squares (2SLS) method.

Table 8 presents the estimated 2SLS results under the distance-based spatial weight matrices ($\mathbf{W_{d80}}$, $\mathbf{W_{d100}}$, $\mathbf{W_{d120}}$). Column (1) presents the estimated parameters using the $\mathbf{W_{d80}}$ spatial weight matrix. The results indicate that there is no enough evidence to suggest that experiencing a drought shock in the previous crop season significantly affects the proportion of improved seed adoption. This pattern is consistent in Column (2) and (3), which estimates the SLM model using the $\mathbf{W_{d100}}$ and $\mathbf{W_{d120}}$ spatial weight matrix respectively. Building on these results, we do not find enough evidence to say 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.

8 Robustness Check

To further validate the robustness of our findings, we incorporate additional spatial weight matrices with varying specifications. In the main analysis, we already tested different distance-based cutoffs, specifically 80 km and 120 km using inverse arc-distance matrices ($\mathbf{W}_{d80}$ and $\mathbf{W}_{d120}$). These variations served to verify that our results were not sensitive to the specific threshold used to define spatial proximity. In addition to these, to validate even more our findings, we are going to construct and define an extra weight matrix to emphasize on immediate neighbor. Following the study of Fang and Richards (2018), we are going to construct a k-neareast neighbor weight matrix ($\mathbf{W_k}$), where the k closest observations or communities are considered to exert spatial influence. Based on the average number of neighbors under the 80km inverse arc distance matrix used in the study, a $k = 8$ would be used,

$$\mathbf{W_k} = w_{ij} = \begin{cases} 1 & \text{if } d_{ij} \leq d_{ij}^k \\ 0 & \text{if } d_{ij} > d_{ij}^k \end{cases} \tag{10}$$

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 $k = 19$ in this case. So, if j is within the k-th nearest neighbors of i (i.e., $d_{ij} \leq d_{ij}^k$, then $w_{ij} = 1$. Otherwise, if $d_{ij} > d_{ij}^k$, then $w_{ij} = 0$.

Following the methodology framework outlined, we find that the SLM model is the most competent specification for capturing spatial dynamics in improved seed adoption for this weight matrix $\mathbf{W_k}$ (See Table 12, 13 and 14 at the Appendix Section for Moran's I, LM and LR tests respectively). We then estimate Equation (5) using the Spatial Lag Model (SLM).

Table 8: 2SLS Regression Estimates – Drought Shock Interaction Effect

Variable	(1)	(2)	(3)
WY_1	0.552***	0.732***	0.793***
	(0.179)	(0.191)	(0.204)
WY_2	0.106	-0.231	-0.083
	(0.315)	(0.374)	(0.328)
Prop. female head	-0.094	-0.085	-0.068
	(0.081)	(0.079)	(0.078)
Mean age head	0.004	0.004	0.004
	(0.003)	(0.003)	(0.003)
Prop. salaried head	0.167**	0.183**	0.186**
	(0.085)	(0.084)	(0.087)
Prop. head edu. 1	0.081	0.071	0.075
	(0.093)	(0.100)	(0.098)
Prop. head edu. 2	0.190*	0.195*	0.187*
	(0.100)	(0.104)	(0.101)
Prop. head edu. 3	-0.013	-0.020	-0.019
	(0.143)	(0.130)	(0.132)
Prop. head edu. 4	0.097	0.088	0.089
	(0.149)	(0.145)	(0.147)
Prop. head edu. 5	0.260*	0.204	0.210
	(0.143)	(0.148)	(0.146)
Prop. head edu. 6	0.606***	0.607***	0.606***
	(0.186)	(0.174)	(0.181)
Prop. head edu. 7	0.040	0.062	0.047
	(0.146)	(0.141)	(0.146)
Total plot size	0.000	0.000	0.000
	(0.002)	(0.002)	(0.001)
Prop. coupon	0.694***	0.642***	0.684***
	(0.149)	(0.158)	(0.152)
Prop. credit	0.165*	0.149	0.164*
	(0.097)	(0.095)	(0.093)
Prop. left seeds	0.071	0.024	0.043
	(0.090)	(0.086)	(0.087)
Constant	-0.263*	-0.269**	-0.313**
	(0.140)	(0.136)	(0.138)
Fixed Effects	Yes	Yes	Yes
Observations	380	380	380

Note: Clustered standard errors are reported in parentheses. WY_1 and WY_2 are instrumented using first- and second-order spatial lags of X characteristics, based on three different spatial weight matrices. Column (1) uses W_{80}, Column (2) uses W_{100}, and Column (3) uses W_{120}. Statistical significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 15 (see Appendix) presents the results from regressing equation (5) and Table 16 (see Appendix) presents the direct, indirect and total effects under the k-nearest neighbor weight matrix ($\mathbf{W_k}$). The findings remain consistent with our main results: higher adoption levels of improved seeds in neighboring communities are associated with higher local adoption. Moreover, both localized peer effects and broader regional spillovers persist under this alternative specification. Specifically, the proportion of seeds obtained through coupons and the proportion of household heads with a university degree continue to generate positive spillover effects on adoption behavior.

9 Conclusion

This study investigated how spatial interactions among neighboring communities influence the adoption of improved seed varieties in rural Malawi. Using georeferenced panel data from the Malawi Integrated Household Panel Survey (IHPS) between 2010 and 2019, the research examined two central questions: (1) whether the adoption behavior in one community is affected by that of its neighbors, and (2) whether drought shocks influence the strength and nature of these peer effects. The empirical strategy centered around spatial econometric techniques, particularly the Spatial Lag Model (SLM), to capture spatial dependence in the adoption process and provide robust evidence of peer effects in a rural, agriculture-dependent context, where technology adoptions are becoming more relevant.

Our results reveal strong and consistent evidence of peer effects: communities tend to adopt improved seed varieties in alignment with the behavior of neighboring communities. This spatial dependence is particularly significant under proximity-based weight matrices, where neighboring influence is strongest at shorter distances (e.g., 80 km). The SLM estimation shows that improved seed adoption is not only shaped by local characteristics such as education levels, financial access, or plot size, but also by external, spatially mediated influences from surrounding communities. Notably, we find positive and statistically significant spillover effects from factors like seed subsidies (coupons) and the share of university-educated household heads, highlighting the regional impact of localized interventions and the value of investing in education.

We also explored how exposure to drought shocks affects these peer dynamics. Using the 3-month Standardized Precipitation Evapotranspiration Index (SPEI) to capture severe climate shocks during critical crop stages, we tested whether such shocks amplify or weaken neighboring influence on seed adoption. Contrary to our expectations, the instrumental variables analysis showed no statistically significant interaction between drought exposure and spatial spillovers. That is, although drought shocks are a major constraint to agricultural productivity, they do not appear to significantly alter how much neighboring communities influence one another in terms of seed adoption decisions. This suggests that peer effects between communities are relatively stable across normal and adverse weather conditions, or that other mechanisms (e.g., credit access, availability of local varieties) may mediate the drought response more strongly than spatial learning.

This work contributes to the growing literature on agricultural technology adoption by explicitly modeling spatial dependence at the community level, something that has received limited attention, particularly in the context of Malawi. The use of spatial panel econometrics allows us to move beyond traditional linear models and address the reflection problem by controlling for unobserved spatial correlation. Still, there are areas for further research. One key limitation of this study is the use of community-level rather than individual-level data, which may obscure heterogeneity in adoption behavior. Future research could benefit from integrating household-level or individual-level data to better identify causal peer effects, especially in the context of climate shocks. This more granular approach may help uncover whether and how drought exposure shapes peer dynamics at a finer scale than community averages allow. Additionally, exploring how other shocks, such as flooding, pests, or fires, interact with peer dynamics could offer a more comprehensive understanding of resilience in agricultural systems.

Ultimately, this study underscores the importance of designing agricultural policies and extension programs that recognize and leverage spatial spillovers. Interventions that target influential communities or build on local peer networks could have amplified regional effects, enhancing the efficiency and reach of policy efforts aimed at boosting the adoption of climate-resilient technologies in vulnerable rural settings.

A Appendix

Table 9: Direct, Indirect, and Total Effects of SLM Model with $\mathbf{W_{d100}}$

Variable	Direct	Indirect	Total
Prop. female head	-0.074	-0.020	-0.094
	(0.083)	(0.028)	(0.106)
Mean age head	0.0004	0.0001	0.0006
	(0.0028)	(0.0008)	(0.0036)
Prop. salaried head	0.025	0.007	0.032
	(0.072)	(0.022)	(0.091)
Prop. head edu. 1	0.183**	0.048	0.231**
	(0.076)	(0.038)	(0.103)
Prop. head edu. 2	0.237**	0.062	0.299**
	(0.096)	(0.047)	(0.128)
Prop. head edu. 3	0.059	0.015	0.074
	(0.104)	(0.033)	(0.132)
Prop. head edu. 4	0.039	0.012	0.051
	(0.139)	(0.046)	(0.180)
Prop. head edu. 5	0.106	0.030	0.135
	(0.148)	(0.048)	(0.189)
Prop. head edu. 6	0.591***	0.158	0.749***
	(0.189)	(0.113)	(0.264)
Prop. head edu. 7	0.004	0.003	0.008
	(0.278)	(0.085)	(0.353)
Total plot size	0.0017	0.0004	0.0021
	(0.0022)	(0.0007)	(0.0029)
Prop. coupon	0.650***	0.172	0.821***
	(0.181)	(0.117)	(0.255)
Prop. credit	0.142*	0.036	0.178*
	(0.081)	(0.031)	(0.104)
Prop. left seeds	0.089	0.024	0.113
	(0.081)	(0.027)	(0.103)

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

Table 10: Direct, Indirect, and Total Effects of SLM Model with $\mathbf{W_{d120}}$

Variable	Direct	Indirect	Total
Prop. female head	-0.074	-0.018	-0.092
	(0.083)	(0.026)	(0.104)
Mean age head	0.0003	0.0001	0.0004
	(0.0028)	(0.0008)	(0.0035)
Prop. salaried head	0.024	0.006	0.030
	(0.072)	(0.020)	(0.089)
Prop. head edu. 1	0.183**	0.043	0.226**
	(0.076)	(0.038)	(0.101)
Prop. head edu. 2	0.235**	0.054	0.289**
	(0.096)	(0.047)	(0.125)
Prop. head edu. 3	0.055	0.012	0.066
	(0.104)	(0.031)	(0.129)
Prop. head edu. 4	0.031	0.009	0.039
	(0.139)	(0.042)	(0.175)
Prop. head edu. 5	0.096	0.024	0.119
	(0.148)	(0.044)	(0.183)
Prop. head edu. 6	0.579***	0.137	0.716***
	(0.189)	(0.113)	(0.259)
Prop. head edu. 7	-0.005	0.0005	-0.005
	(0.278)	(0.079)	(0.345)
Total plot size	0.0017	0.0004	0.0021
	(0.0022)	(0.0007)	(0.0028)
Prop. coupon	0.655***	0.154	0.809***
	(0.181)	(0.121)	(0.255)
Prop. credit	0.146*	0.033	0.179*
	(0.081)	(0.031)	(0.103)
Prop. left seeds	0.089	0.021	0.111
	(0.081)	(0.026)	(0.101)

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

Table 11: 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: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 12: Moran's I and LM Tests from No-Spatial Model with W_k.

	OLS w/ FE		
Test	Statistic	df	p-value
Spatial error:			
Moran's I	3.65	1	0.000 ***
Lagrange multiplier	11.68	1	0.001 ***
Robust Lagrange multiplier	2.24	1	0.134
Spatial lag:			
Lagrange multiplier	18.35	1	0.000 ***
Robust Lagrange multiplier	8.91	1	0.003 ***

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

Table 13: Moran's I and LM Tests from SLX Model with W_k.

Test	Statistic	df	p-value
Spatial error:			
Moran's I	3.01	1	0.003 ***
Lagrange multiplier	3.01	1	0.0.83
Robust Lagrange multiplier	2.623	1	0.101
Spatial lag:			
Lagrange multiplier	4.67	1	0.031 **
Robust Lagrange multiplier	4.43	1	0.035 **

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

Table 14: Likelihood-Ratio (LR) Tests with W_k.

Test Comparison	LR Chi2	p-value
SLM vs. SARAR	0.37	0.57
SLM vs. SDM	15.59	0.34

Table 15: Regression Estimates: SLM model with $\mathbf{W_k}$.

	SLM ($\mathbf{W_k}$)
Main Variables	(1)
Prop. female head	-0.080
	(0.080)
Mean age head	0.001
	(0.003)
Prop. salaried head	0.007
	(0.074)
Prop. head edu. 1	0.182**
	(0.078)
Prop. head edu. 2	0.251**
	(0.099)
Prop. head edu. 3	0.070
	(0.109)
Prop. head edu. 4	0.043
	(0.139)
Prop. head edu. 5	0.131
	(0.155)
Prop. head edu. 6	0.581***
	(0.191)
Prop. head edu. 7	0.038
	(0.278)
Total plot size	0.002
	(0.002)
Prop. coupon	0.611***
	(0.169)
Prop. credit	0.146*
	(0.079)
Prop. left seeds	0.083
	(0.082)
Spatial Effects	
ρ	0.302***
	(0.092)
Variance	
σ_e^2	0.030***
	(0.002)
Fixed Effects	Yes
Observations	380
Log-Likelihood	114.54

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

Table 16: Direct, Indirect, and Total Effects of SLM Model with $\mathbf{W_k}$

Variable	Direct	Indirect	Total
Prop. female head	-0.078	-0.035	-0.113
	(0.083)	(0.041)	(0.120)
Mean age head	0.0005	0.0003	0.0008
	(0.0028)	(0.0013)	(0.0040)
Prop. salaried head	0.0145	0.0061	0.0207
	(0.0717)	(0.0327)	(0.1027)
Prop. head edu. 1	0.182**	0.079	0.261**
	(0.0759)	(0.0496)	(0.1161)
Prop. head edu. 2	0.251**	0.108*	0.359**
	(0.0960)	(0.0654)	(0.1478)
Prop. head edu. 3	0.070	0.030	0.100
	(0.1031)	(0.0501)	(0.1497)
Prop. head edu. 4	0.041	0.019	0.060
	(0.1383)	(0.0676)	(0.2020)
Prop. head edu. 5	0.126	0.056	0.182
	(0.1476)	(0.0715)	(0.2126)
Prop. head edu. 6	0.596***	0.257*	0.853***
	(0.1869)	(0.1392)	(0.2908)
Prop. head edu. 7	0.040	0.020	0.061
	(0.2767)	(0.1291)	(0.3986)
Total plot size	0.0015	0.0006	0.0022
	(0.0022)	(0.0011)	(0.0032)
Prop. coupon	0.631***	0.271*	0.902***
	(0.1797)	(0.1397)	(0.2822)
Prop. credit	0.147*	0.063	0.210*
	(0.0806)	(0.0445)	(0.1187)
Prop. left seeds	0.088	0.038	0.127
	(0.0802)	(0.0394)	(0.1161)

Note: Robust standard errors in parentheses. * 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.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2020). Peer effects in networks: A survey. *Annual Review of Economics*, 12(1):603–629.

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.

Herrera Gómez, M. H. (2015). Econometría espacial usando stata. breve guía aplicada para datos de corte transversal.

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.

Läpple, D. and Kelley, H. (2015). Spatial dependence in the adoption of organic drystock farming in ireland. *European Review of Agricultural Economics*, 42(2):315–337.

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.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

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.

Singha, C. (2019). Impact of the adoption of vegetative soil conservation measures on farm profit, revenue and variable cost in darjeeling district, india. *Environment and Development Economics*, 24(5):529–553.

Skevas, T., Skevas, I., and Swinton, S. M. (2018). Does spatial dependence affect the intention to make land available for bioenergy crops? *Journal of Agricultural Economics*, 69(2):393–412.

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.

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.

Ward, P. S. and Pede, V. O. (2015). Capturing social network effects in technology adoption: the spatial diffusion of hybrid rice in bangladesh. *Australian Journal of Agricultural and Resource Economics*, 59(2):225–241.

Wollni, M. and Andersson, C. (2014). Spatial patterns of organic agriculture adoption: Evidence from honduras. *Ecological economics*, 97:120–128.

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.