MASTER

APPLIED ECONOMETRICS AND FORECASTING

MASTER´S FINAL WORK

DISSERTATION

PRECISION AGRICULTURE: A SPATIAL ECONOMETRIC ANALYSIS OF A PORTUGUESE MAIZE YIELD

MARIANA CANAVARRO REIS TEIXEIRA ESTEVES

SUPERVISION:

PROFª DOUTORA ISABEL PROENÇA

PROFº NUNO SANTOS

OCTOBER – 2021

ABSTRACT

Nowadays, maize is the most important cereal in the world and its production has been increasing both worldwide and in Portugal, over the years. The constant technological development has led to the creation of new techniques such as precision agriculture, to better meet the global needs of this primordial cereal as well as optimize its production.

This research was developed jointly with the firm Agro Analítica from the agriculture sector, whose area of expertise is Precision Agriculture and System Optimization. In this manner, the present work aims to estimate a function for the maize yield identifying the relevant determining factors, and their effect, on maize productivity on an exploitation of a firm in Azinhaga, Golegã, district of Santarém, Portugal for the year 2020.

Using appropriate software, this dissertation applies the most recent spatial econometric methods to cross-sectional data, in order to properly include spatial dependence in the estimation. Thus, the appropriate models were estimated: *Spatial Lag Model* (SLM), *Spatial Error Model* (SEM) and *SARAR(1,1) Model*, whose use was recommended by the diagnosis to OLS (*Ordinary Least Square*) residuals. The elected model was the SARAR(1,1), capturing the spatial dependence and heteroscedasticity in the data, with an accuracy of approximately 90%. In this framework, it was concluded that maize yield, in the year and area under study, is positively influenced by factors such as the sowing density, applied sulfur trioxide (SO3) and a specific variety of seed. Regarding the fertilization, nitrogen and potassium, and irrigation of the crop, presented a non-linear (quadratic) relationship with the maize yield. Also influencing the yield, there are weather-related variables measured by stage of the maize life cycle, that proved to be significant at explaining the variable under study such as the relative humidity, the temperature, and the wind velocity.

Keywords: Maize Yield; Spatial Econometrics; Spatial Regression; Cross-section data; Spatial Dependence.

Resumo

Atualmente, o milho é o cereal mais importante do mundo tendo a sua produção vindo a aumentar tanto a nível mundial como em Portugal ao longo dos anos. O constante desenvolvimento tecnológico resultou na criação de novas técnicas, como a agricultura de precisão, para melhor satisfazer as necessidades globais deste cereal primordial bem como otimizar a sua produção.

Esta investigação foi desenvolvida em conjunto com a empresa *Agro Analítica* do sector da agricultura, cuja área de especialização é Agricultura de Precisão e Otimização de Sistemas. Desta forma, o presente trabalho visa estimar uma função que explique a produtividade do milho identificando os fatores, e o seu efeito na produtividade do milho, numa exploração da empresa em Azinhaga, Golegã, distrito de Santarém, Portugal, para o ano 2020.

Utilizando o software apropriado, esta dissertação aplica os mais recentes métodos e ferramentas econométricas espaciais para dados *cross-section* (dados transversais), de modo a incluir devidamente a dependência espacial na estimação. Assim, foram estimados os modelos apropriados: Modelo *Spatial Lag* (SLM), Modelo *Spatial Error* (SEM) e Modelo *SARAR(1,1)* (Kelejian & Prucha, 2010), cuja utilização foi recomendada pelo diagnóstico aos reísduos OLS (*Ordinary Least Square*). O modelo escolhido foi o *SARAR(1,1)*, captando a dependência espacial e heterocedasticidade presente nos dados. Desta forma, concluiu-se que a produtividade do milho, no ano e na área em estudo, encontra-se positivamente influenciado por fatores como a densidade de sementeira, o trióxido de enxofre aplicado (SO3) e uma variedade específica de sementes. Quanto à fertilização, com azoto e potássio, e quanto à irrigação da cultura, estes fatores apresentaram uma relação não linear (quadrática) com a produtividade do milho. Também influenciando a produtividade, existem variáveis relacionadas com o clima, medidas pela fase do ciclo de vida do milho, que provaram ser significativas para explicar a variável em estudo, tal como a humidade relativa, a temperatura e a velocidade do vento.

Palavras-passe: Produtividade do Milho; Econometria Espacial; Regressão Espacial; Dados de corte transversal; Dependência Espacial.

Acknowledgments

I want to thank everyone who accompanied me along this path, and who in some way helped me in this research.

First of all, I would like to thank my supervisor Professor Nuno Santos, for accepting the supervision of my dissertation and his tireless support, guidance and transmission of knowledge essential to this project.

To my co-supervisor Professor Isabel Proença, for her observations as well as all the help, dedication and availability given throughout the preparation of this work.

To both, thank you for the support and great ability to teach.

To Dr. António Coimbra, from the firm Agro Analítica, for his thoughtful availability during the execution of this project.

To the firm *Quinta da Cholda | Competitividade Sustentável* for making data available for the execution of this dissertation.

To the Lisbon School of Economics and Management and all the professors I came across during my years as a student once all contributed in part to this project.

And finally, to my parents, family and friends, a "thank you" will never be enough, for all the love, support and advice throughout my academic and personal life.

I feel grateful to all of you.

To my grandmother,
in loving memory.

Table of contents

TABLE OF TABLES AND FIGURES

1. INTRODUCTION

Understanding phenomena resulting from the spatial distribution of data in space presents a major challenge for several areas of knowledge. It wasn't until recently that the use of spatial econometrics started to being incorporated in the field of agriculture. The 1990's came with the realization that crops and, especially their yields, are influenced by factors as soil characteristics, type of seeds planted, applied fertilization and weather conditions, among others (Bockstael, 1996). Meanwhile, the evolution of technology and the introduction of precision farming in this field has brought more efficiency and optimization of crop production.

In PortugaL, in the agricultural sector, maize (*Zea Mays L.*) is considered the most important cereal in the world (ANPROMIS, 2017). The productivity of this crop is the result of the interaction of many factors controllable, or not, by the farmer, ranging from the choice of seed and, hence, the genetic traits of the plant, to climatic conditions in a given growing season (Cruz et al, 2008). It is within this framework and with the data provided by the firms Agro Analítica and Quinta da Cholda that this research on maize yield in the Quinta da Cholda exploitation, in 2020, arises. In view of the above, the main objective of this essay is to identify the factors and their effect on maize yield in the area under study for the year of 2020.

To this regard, this dissertation applies the most recent spatial statistical methods and tools to the cross-sectional data, in order to properly include factors such as spatial dependence and heterogeneity, commonly observed in studies related to agricultural econometrics. As a result, it's possible to correctly investigate the spatial causal relationship between the maize yield and soil classification, soil attributes and the agrometeorological variables. To analyze the maize yield and its explanatory variables, in conjunction with the study of spatial dependence using the global Moran´s I, spatial regression models are applied such as the SLM (*Spatial Lag Model*), SEM (*Spatial Error Model*) and the *SARAR(1,1) Model* (Kelejian & Prucha, 2010), after spatial correlation has been detected by means of hypothesis testing (Anselin, 2005).

This thesis is structured in seven chapters, including this introduction as the first. Chapter two contains a brief framework in which this research was developed. Chapter three presents a summary of the literature on the maize crop, a brief historical review of spatial econometric analyses carried out over the years in agriculture, with special attention to maize cereal, as well as the existing literature on the maize crop. Chapter four outlines the methodology adopted in this work. Chapter five describes the data, the steps for building the data base and the description of all variables. Chapter six contains the estimated models as well as the detailed analysis of the results. And, finally, the conclusions of the study, together with its limitations and suggestions for future work are displayed in chapter seven.

2. DISSERTATION FRAMEWORK

This research was jointly developed with the firm Agro Analítica, whose area of expertise is Precision Agriculture and System Optimization. This firm, from the agriculture sector, is based in Lisbon and was recently founded in 2017. Its work in Smart Farming | Precision Agriculture aims to fill the technological gap that exists in Portugal in this area (Agro Analítica, 2021).

The exploitation where the data for this study was collected is situated in *Quinta da Cholda* in the region of Azinhaga, Golegã, district of Santarém, Portugal (Quinta da Cholda, 2021). The firm holds a wide range of data regarding maize production as well as other variables, namely weather variables that may influence the production of this cultivar. Specifically, the exploitation under study cultivates grain maize (monoculture) that is later used for animal feed. Figure 1 below illustrates the aerial map of Portugal zooming in on the holding under study depicted in orange.

Figure 1 – Aerial map of the maize crop exploitation under study

Source: Own Elaboration; Software: QGIS.

3. LITERATURE REVIEW

Maize is considered the most important cereal in the world (ANPROMIS, 2017), currently exceeding the annual production of both rice and wheat. This cereal presents now numerous applications whether for silage, animal feed or the food industry, such as flour and starches, or even to produce renewable energy (bioethanol and biogas) and biodegradable materials (fibers and bioplastics). Nowadays, according to the FAOSTAT (2021), maize is grown in more than 160 countries, from the most advanced to the self-subsistent, being one of the most productive crops with an annual world production, in 2019, of 1148 million tonnes per hectare (Figure 5, Appendix A).

In the Portuguese agricultural context, the cultivation of maize appears intimately linked to irrigation, especially crucial in Mediterranean environments (ANPROMIS, 2017). Presenting itself as the most important arable crop in Portugal, it occupied, in 2019, around 83360 hectares of cultivated area with an annual production of 748780 tonnes per hectare, in the same year (FAOSTAT, 2021).

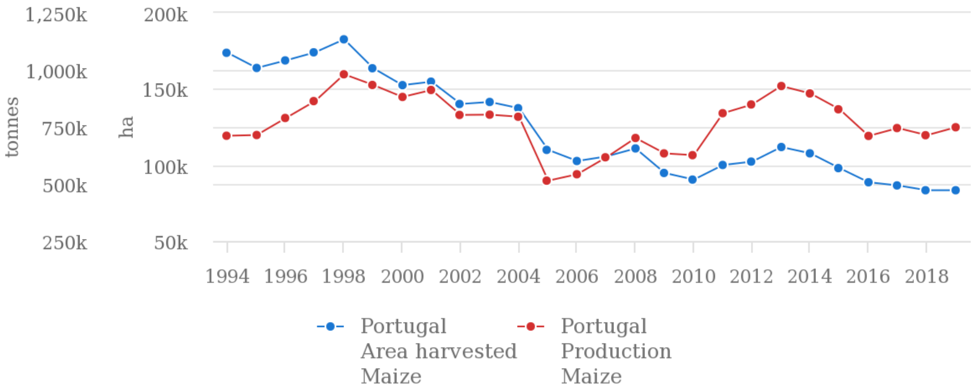

Figure 2 - Production quantities of maize in Portugal, from 1994 to 2019

Source: FAOSTAT (2021)

Figure 2 shows that, in the Portuguese case, there are many oscillations, and it shows precisely the opposite with respect to the world case. Both the planted area and the production of maize have been decreasing. It also shows that maize production decreased until 2004 and then exhibits an increasing trend with oscillations until 2019. However, it shows that with less planted acreage, production is higher, contrary to what happened in the 1990s. This outcome can be explained by the beginning of the use of more sophisticated and complex techniques in the field of agriculture, as well as the introduction of new varieties of maize better adapted to the soil and climatic conditions.

In Figure 6, in Appendix A, one can now look at the yield of maize, whereas in the previous figures only the total production was measured. A similar conclusion to the previous one follows. Maize production has become more efficient. Since yield is measured by planted area, usually in hectares, as the harvested area decreases, the production has been increasing. Both Portugal and the world follow an increasing trend, with some oscillations registering, in 2018, an average maize yield of 8.56 and 5.92 tonnes/ha per year, respectively.

It is within this context that the concept of precision farming is introduced. ISPA (2021) states that *"Precision Agriculture is a management strategy that takes account of temporal and spatial variability to improve sustainability of agricultural production"*. Typically, indicators of the crop's potential yield are built using soil studies. However, the emergence of precision farming has brought a more precise and thorough analysis of spatial variations with the use of complex technologies such as the global positioning system (GPS) and the geographical information systems (GIS) (Stafford & Bolam, 1998). Based on this, and on the complexity of the interactions between variables influencing maize yield and quality, in time and space, a multivariate approach to the problem is required.

3.1. Time and Space in Agricultural Econometrics

Despite the extensive literature about planted acreage by agricultural economists, there are gaps in literature that remain to address. Most of the studies conducted disregard the spatial dependence and heterogeneity present in the data, thus ignoring the spatial and over time variability of crops, which an econometric analysis could explain. Only a small number of studies applied spatio-temporal regression and techniques to analyze and understand the complex phenomena studied in precision agriculture (Bongiovanni & Lowenberg-Deboer, 2002; Lambert et al., 2006; Liu et al., 2006).

It was in the early 1990s, that farmers started to use yield monitors to produce yield maps for their fields (Bockstael, 1996). However, the interpretation of these maps can be complicated since crop yield is associated with both transient and permanent crop factors. Transient factors, include insects, diseases, planter or applicator malfunctions and measurement errors that result from the transport, mixing and cycling of the grain (Lark et al., 1997). This are site-specific factors that vary from year to year. Permanent spatial effects, landscape position, terrain attributes, erosion, and soil properties can also alter, alongside the transient factors, the spatial patterns in yield maps (Kravchenko & Bullock, 2000; Stone et al., 1985). According to Sudduth et al. (1997), data from multiple years are needed to identify recurring spatial yield patterns and, therefore, understand the effect of this factors in the crop yield.

As for the terrain attributes, topography is one of the most obvious causes for yield variation. Being mostly part unchangeable, can be used to explain variation. For example, maize silage yields are highest at lower positions rather than at mid-slope or summit positions (Afyuni et al., 1993; Spomer & Piest, 1982). Usually, the combination of the effect of terrain attributes, such as elevation, slope and curvature, with the plant available water, highly influence the crops yield. In years with below-normal rainfall, areas with greater slopes and convex curvatures normally have less available water and lower yields than areas lower on the hillslope and concave curvatures (Kaspar et al., 2003).

Still with regard to soils, the area of soil nutrition is where the greatest difficulties and expenses arise. Soil fertilization, especially with the maize macronutrients such as nitrogen (N), potassium (K) and phosphorus (P), is linked to several factors. Depending on the crop sowing density, soil type, i.e., its texture, and the characteristics of the macronutrient itself and its absorption by the plant, the crop will be more or less productive that year (Barros & Calado, 2014).

In the field of weather data, there’s still no agreement regarding the appropriate spatial or temporal aggregation of the data. In a study done by Dixon et al. (1994), these variables were measured differently. Typically, monthly measurements are used in most maize yield response models (Huff & Neill, 1980; Offutt et al., 1987). However, a monthly data proxy doesn´t provide a good specification for the climatic effects because of the year-to-year variability of the crop. Each month varies by location and year since the planting dates and weather events also vary, putting the maize at different development stages at different months each year. Hence, the author suggests measuring them by growth stage of the crop allowing for a better specified model where all the different crop planting dates can be taken into account.

Typically, studies only included precipitation and temperature as weather variables in regression analysis, mainly due to the lack of estimates available for other climatic data as is the case with solar radiation. According to Daughtry et al. (1983), there’s a positive relationship between the final maize yield and the cumulative solar radiation available which can be observed, especially in the 3rd and 4th stage of maize's life cycle, since the plant's leaves are fully developed intercepting more efficiently solar radiation for the photosynthesis.

3.2. Maize Crop

Maize, scientifically named *Zea Mays L.*, is a species belonging to the family Gramineae/Poaceae family, whose subspecies, *Zea mays*, originated in South and Central America more than 8000 years ago. This extremely adaptable crop is grown in climates that are tropical, subtropical and temperate, and also supporting altitudes of over 3600 meters (Barros & Calado, 2014).

It's great adaptability and successful cultivation depends mostly on the right choice of varieties, so that the length of the growing stage of the crop matches the length of the growing season and the purpose for which the crop will be grown. The optimal choice of sowing date is the cheapest tool to improve the grain yield. Each variety has an optimal sowing date and the greater the deviation from this optimal date (early or late sowing), the greater the yield losses (Chhetri et al., 2018; Sárvári & Futó, 2001).

Regarding the types of soil and, especially in Mediterranean environments, under irrigation conditions, a good circulation of water and air, a high usable capacity for water, the availability of nutrients in the soil and ideal weather conditions, grant this crop a better response.

With respect to the ideal temperature, it differs according to the stage in which the crop is at. In the germination stage the optimum temperature is 15ºC and should always be higher than 10ºC. In the vegetative development and flowering stage, the optimum temperatures range between 24 and 35ºC. According to Bellido (1991), in negative temperatures, the growth of the plant is compromised, and its aerial part ends up dying. Likewise, maximum temperatures above 35ºC, especially during flowering and fertilization, greatly diminish the productivity of the crop, due to the decrease in the number of grains.

Furthermore, high temperatures, especially in the south of Portugal, combined with water deficiency may compromise the development of the plant and, thus, causing a drop in productivity. Being a spring-summer crop, sown in the months of March to May, and under the climatic conditions in Portugal, it is extremely important to pay attention to the crop's water requirements.

The stages of plant development that are most critical to water deficiency correspond to the beginning of flowering, the fertilization period and, finally, the grain filling phase. Maize growth stages are divided into vegetative stages (V) and reproductive stages (R) as illustrated in the figure 3 bellow.

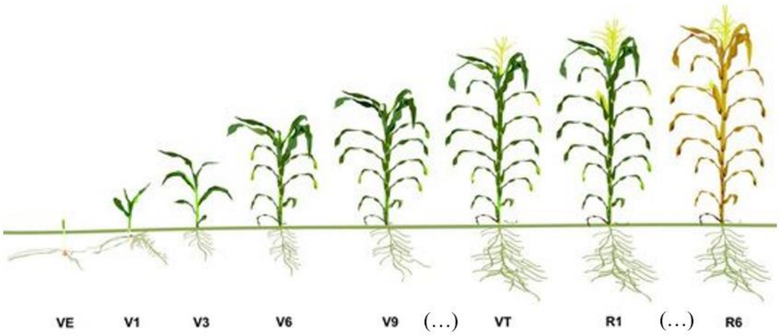

Figure 3 - Growth and development stages of the maize

Source: Pionner (2021)

The vegetative phases are divided by the number of leaf collars visible on the plant, presenting, normally, 6 sub-stages (Dekalb Asgrow Deltapine, 2020):

1. VE (emergence), when maize seedlings emerge from the soil occurring up to 5 days under ideal conditions, or up to 2 weeks in dry/cool conditions;
2. V1-V5, from the appearance of the first collar leaf until the fifth;
3. V6-V8, 4 to 6 weeks after emergence the number of kernel rows is defined, reaching approximately 60 cm high;
4. V9-V11, maize begins rapid and steady growth followed by dry matter accumulation;
5. V12-Vnth, the tassel is almost visible and water and nutrients deficiency results in a lower yield of maize;
6. And, finally, in VT, the plant has reached full size and the tassels are visible. A successful pollination is critical to make the kernels viable during for 1 to 2 weeks.

Maize plants normally develop up to the V18 stage before reaching maximum height and transitioning to the next step of growth.

Concerning the reproductive growth stages, this are now determined by kernel development (Dekalb Asgrow Deltapine, 2020):

1. R1, also called Silking, is one of the most critical stages for yield starting when the silks are visible, and the pollination begins;
2. R2, or Blister, kernels are white and cointain a clear fluid, and the silks dry out;
3. R3, also called Milk since the clear fluid turns milky and the kernels turn yellow;
4. R4, or Dough, it’s when kernels begin to dent at the top, and the milky fluid turns into a dough-consistency;
5. R5, the kernels only present about 55% moisture regarding the silking stage and harvesting may start;
6. And, finally, in R6, maturity is reached, and yield determined (kernels cointain a maximum of 30 to 35% moisture).

An extremely important climatic factor identified in several agronomic studies for predicting maize yield is solar radiation (Daughtry et al., 1983). A large part of maize's dry matter comes from the fixation of CO2 by the photosynthetic process being considered a highly efficient plant in the use of light. Long periods of cloudiness associated with frequent rainfall, thus suppressing active photosynthesis, are associated with a decrease in maize yield (Cruz et al, 2008).

Regarding the physical attributes of the crop, there are a few matters to consider. The planting density defined as the number of seeds per unit area, plays an important role in the yield of a maize crop, since the maize is the most sensitive *Gramineae* to variations in plant density. Depending on water availability, humidity level, soil characteristics, the maize’s life cycle, sowing time and inter-row spacing, the optimum number of seeds per hectare varies substantially. Generally, the cause of low maize yields is the low number of plants per area (Cruz et al, 2008).

On to elevation, which is the heigh above the sea level, several studies indicate that lower elevation areas have a lower maize yield (Seffrin, 2017). This can be due to water accumulations in lower regions. With the crop being waterlogged and coupled with lower/higher temperatures and solar radiation exposure, the maize yield diminishes.

Finally, fertilization is very important to obtain the potential yield of the maize crop. The most absorbed nutrients by this plant, and fundamental to its growth, hence being called macronutrients, are Nitrogen, Potassium, and Phosphorus. Starting with Nitrogen, its management is something difficult to do. Because it is a very soluble compound, it is easily lost by washing along the soil profile. Especially in irrigated conditions, this can happen if the amount of water used for irrigation is very high causing surface runoff, dragging the nitrogen and, consequently, its leaching (Barros & Calado, 2014). Because of this, it is very difficult, or practically impossible, to forecast a precise amount of nitrogen fertilization. To compensate for nitrogen excess, phosphorus has the function of stimulating root growth, increasing the mechanical resistance of the stems, and positively influencing flowering. This macronutrient is now poorly soluble and can easily become unavailable to plants. In addition, if the soil has an acid pH, phosphorus tends to bind to the iron and aluminum present in the soil, thus becoming unavailable for plant uptake. If the soil is alkaline, phosphorus binds to calcium forming a poorly soluble compound becoming difficult for plants to absorb. Lastly, potassium is the macronutrient most absorbed after nitrogen, contributing to the improvement of the quality of the maize. In other words, it is less washed out than nitrogen and more than phosphorus. In addition, if bound to clays, it becomes unavailable and impossible for plants to absorb (Barros & Calado, 2014).

4. METHODOLOGY

This chapter describes the methodology applied to achieve the objectives presented in this research, where the details of the procedures used to treat and build the database, and the applied spatial econometrics techniques, are discussed.

4.1. Underlying Hypothesis of the Climatic Variables

For the weather-related variables, it will be followed the approach previously mentioned used by Dixon et al. (1994). These authors decided to include in their models, climate variables measured by maize development stage rather than by month, as is commonly seen in most agricultural econometric studies. The reason for this is the year-to-year variability of the crop and all factors with it associated. In this manner, in order to create variables by growth stages rather than by month, information was gathered on important dates in the maize life cycle, including: Sowing date, emergence date, start of irrigation, 2nd leaf date, 4th leaf date, 6th leaf date, 8th leaf date, weeding date, 10th leaf date, 12th leaf date, 14th leaf date, flowering start date, flowering end date, irrigation end date, black spot date and harvest date. From these, and in agreement with the literature reviewed previously, four stages of maize growth were defined:

Table I - Definition of the maize growth stages

Stage	Plant Activity	Starting date	Ending date
1	Emergence of the seedling from below the soil	Planting date (March/April/May)	Emergence date (April/May/June)
2	Early vegetative growth	Emergence date (April/May/June)	Flowering start date (June/July)
3	Flowering	Flowering start date (June/July)	Flowering end date (June/July/August)
4	Grain fill until maturity (harvest)	Flowering end date (June/July/August)	Harvest date (September)

Note: Both stages 1 and 3 only last about 15 and 10 days, respectively.
Source: Own elaboration.

Each parcel of the exploitation, and hence, each spatial unit (id), has its own sowing date. Depending on the variety of maize planted, i.e., whether the maize cycle is longer or shorter, yield results from previous years and weather conditions, sowing takes place in different months. In the year under study, maize was all sown between April and May.

4.2. Spatial Econometric Methods

4.2.1. Spatial Weights Matrix W

The first notions of spatial dependence, or, more precisely, spatial autocorrelation, introduced by Moran (1948) and Geary (1954), were based on the simple concept of binary contiguity between spatial units. The idea of neighborhood based on contiguity implies that two contiguous regions are neighbors if they share a common physical border (Almeida, 2012). In this case, the value of 1 is assigned, which otherwise would be 0. In 1981, this notion was extended by Cliff and Ord in order to include a more general measure of the potential interaction between two spatial units, resulting in the Cliff-Ord weight matrix, also kwon as the spatial weight matrix *W*.

Later defined by Anselin and Bera (1998),

> *A spatial weights matrix is a N by N positive and symmetric matrix which expresses for each observation (row) those locations (columns) that belong to its neighborhood set as nonzero elements. More formally, w_{ij} = 1 when i and j are neighbors, and w_{ij} = 0 otherwise.*

(Anselin & Bera, 1998, p.243)

It is also to be noted that the elements of the weight matrix are nonstochastic and exogenous to the model, being defined *a priori* to avoid *spurious* relationships.

Giving the geographical regularity of the spatial units in this research, it is essential to correctly define the spatial weights matrix in order for the econometric model to be well specified. The geographic connection criterion of the spatial weight matrix relies on the idea of proximity, which, in turn, can be defined according to contiguity or geographic distance in accordance with a given metric (Almeida, 2012). For this study, the idea of contiguity will be applied, where two regions are neighbors if they share a common physical boundary.

There are several ways of defining contiguity that are distinguished simply by the way in which the concept of geographical boundary is defined. The queen criterion, in addition to common physical boundaries, also considers vertices in common. The tower criterion only takes into account common physical boundaries. And finally, if only vertices are considered to define contiguity, we are faced with the bishop criterion (Almeida, 2012).

4.2.2. Spatial Effects

After the spatial weights matrix has been specified, it is now possible to analyze the spatial effects of spatial dependence and heterogeneity in a spatial model.

According to Anselin and Florax (1995), Moran´s I statistic is one of the most widely used techniques to measure spatial autocorrelation. This statistic was first suggested by Moran in 1948 and then popularized through the classic work on spatial autocorrelation by Cliff and Ord (1973). In essence, it’s the cross-product statistic between the vectors of values observed at time and the weighted average of the neighborhood values, or spatial lags, with the variable expressed in deviations from its mean. In other words, is a measure of linear association between a variable and its special lag (Almeida, 2012).

This statistic provides a general measure of the linear degree of spatial association between the variable in time and weighted average of neighborhood values, or spatial lag, of the variable in question. This value can be visualized in the Moran diagram, first proposed by Anselin in 1996, were the slope of the linear fit of the scatter plot between the spatially lagged variable, on the y-axis, and the original variable, on the x-axis, equals

Moran's I. Values close to zero indicate the nonexistent significant spatial autocorrelation, that is, the closer to the unitary value, the more autocorrelated it will be. Regarding the sign of this linear association, if the coefficient value is positive, then it reveals positive spatial autocorrelation, contrasting with a negative coefficient value that represents negative spatial autocorrelation.

Even more, the scatter plot can easily be decomposed into four quadrants. The upper-right quadrant and the lower-left, also known as the *high-high* and *low-low* quadrants, respectively, correspond to *positive* spatial autocorrelation (similar values at neighboring locations). In contrast, the lower-right and upper-left quadrant, or, respectively, the *low-high and high-low* quadrant, correspond to *negative* spatial autocorrelation (dissimilar values at neighboring locations).

However, it is important to refer that the classification above listed does not imply significance. In order to assess significance, inference for Moran's I is performed under the null hypothesis of spatial randomness, that is, the evidence of no spatial autocorrelation (Anselin, 1996).

The second type of spatial effects, spatial heterogeneity, manifests itself when structural instability occurs in space. In other words, this implies that parameters vary with location, i.e., are not homogeneous throughout the data set (Anselin, 1988). Due to the large heterogeneity associated with agriculture, there are unobservable factors that are specific to the terrain and are maintained over time.

4.3. Spatial Econometric Models

As most geo-referenced variables are spatially autocorrelated or/and present spatial heterogeneity, spatial regression models are more appropriate than models that do not take spatial autocorrelation into account, as is the case of the Linear model estimated by OLS. As a non-spatial model, is referred to as the Best Linear Unbiased Estimator (BLUE) only when the assumptions of homoscedasticity and no autocorrelation are satisfied. Even when a lagged dependent variable is introduced, the OLS estimator remains consistent, as long as the error term does not show any serial autocorrelation. Hence, even though the estimator may no longer be unbiased, it can still be used as the basis for asymptotic inference (Anselin, 1988). The OLS model is first fitted to obtain regression diagnostics for the spatial dependence of the residuals, being then conducted four statistical tests to detect the presence of this spatial effect in linear models such as the simple Lagrange Multiplier (LM Lag and LM Error) and the Robust version (Robust LM lag and Robust LM Error) (Anselin, 1988; Anselin et al., 1996). Thus, this section discusses some of the most commonly used models in cross-section data analysis that can be estimated when the assumption of absence of spatial autocorrelation of errors is violated.

The first model discussed is the *Spatial Lag Model* (SLM), also known as the *spatial autoregressive* (SAR) *Model*. Spatial dependence is incorporated in this model as an additional regressor in the form of a spatially lagged dependent variable (Anselin, 2003). Formally, a SLM is expressed as

$$y = \rho W y + X\beta + \varepsilon \tag{1}$$

where ρ is a spatial autoregressive coefficient, W is a N×N spatial weights matrix, Wy represents a vector (N×1) of a weighted sum of the outcomes of the neighbouring locations of y, and ε is a vector of the error terms. Thereby, this model is specified so that the value of the dependent variable observed in a given region, is determined by the weighted sum of the values of the dependent variable observed in the neighboring region (Wy), by the values of the exogenous explanatory variables (X) and, also, randomly influenced by an error term (ε) (Almeida, 2012). As a result, the spatial lag term, Wy, is correlated with the disturbances, arising an endogeneity problem, even when the latter are independent and identically distributed. Consequently, the spatial lag term must be treated as an endogenous variable and this model cannot be estimated using the OLS method since this estimator would be biased and inconsistent due to the simultaneity (Anselin, 2003). In turn, the Generalized Methos of Moments (GMM), originally presented by Kelejian and Prucha (1998, 2010), is used to deal with the spatial effect, and the equation (1) can be written the reduced form,

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon\ . \tag{2}$$

Moving on to the next model, the *Spatial Error Model* (SEM), known for the spatial dependence in the regression disturbance term, is referred to as nuisance dependence. This model can be represented by

$$y = X\beta + \varepsilon \ with\ \varepsilon = \lambda W\varepsilon + u\ , \tag{3}$$

were λ represents the spatial lag coefficient of the error term (Almeida, 2012). The spatial dependence in this model now manifests itself in the error term rather than in the dependent variable. According to Almeida (2012), the errors associated with any observation represent the weighted average of the errors in the neighboring locations plus a random error component. Therefore, the assumptions of uncorrelated errors and/or the assumptions of homoscedastic (constant variance) are not satisfied and the OLS estimation loses its optimal properties. It is necessary to resort to another estimator as the GMM (GS2SLS).

And finally, Kelejian and Prucha (1998), advocated models that include both endogenous interaction effects and interaction effects among the error terms. The *SARAR(1,1) Model* assumes both hypotheses of the models presented above and determines those outcomes simultaneously (Kelejian & Prucha, 2010),

$$y = \rho W y + X\beta + \varepsilon \; with \; \varepsilon = \lambda M\varepsilon + u \tag{3}$$

with the (N×N) spatial weigh matrices W and M taken to be known and nonstochastic, may be different or equal. Note that in order to avoid unstable behavior, the constraints on the spatial parameters require that $|\rho| < 1$ and $|\lambda| < 1$. Here, besides the errors being autocorrelated, they are heteroscedastic. In what concerns the ordinary least-squares estimator, it will be inconsistent due to the presence of the variable Wy (Anselin, 1988). However, because of its properties, in a more restricted model becomes the most appropriate to properly estimate the regression at hand since it accounts for both spatial effects.

This will suggest that a robust inference will be carried out, based in LM tests of Anselin (1988) of spatial autocorrelation and heteroscedasticity. For the linear OLS Model and the SLM, White Standard Errors were performed with the estimation and, for the SEM and SARAR(1,1) Model it’s also necessary to estimate a robust inference of the estimator covariance matrix in presence of both spatial heteroskedasticity and autocorrelation (KP-HET) (Kelejian & Prucha, 2010).

5. EXPLORATORY SPATIAL DATA ANALYSIS

5.1. The Data Base

The data used in this study has the primary and only source the data provided by Agro Analítica and Quinta da Cholda firms. The maize exploitation in question is considered to be large, with approximately 542.5 hectares. The mapping of the area was done by using a grid divided into 10 by 10-meter squares (100 square meters) each representing a spatial unit (id). In addition to having data about the spatial unit, id, for the year 2020, information is also available on the spatial coordinates of each id and its insertion on each Parcel and Sub-Parcel. This way, the exploitation has a total of 54265 geo-referenced spatial units (ids).

The analysis is carried out at a spatial unit level, using cross-sectional data from the growing season of the year 2020. After an initial cleaning of the data and for it to be correctly interpreted in *Geoda* and *GeodaSpace* software, observations with missing values were eliminated, resulting in a data base with 53891 observations. This way, only 0.7% of the observations were dropped.

Concerning the data collection method, at the end of the season, the harvesters enter the farm and collect the maize. From these machines, with a width of 6 meters, a shapefile is created with the kilograms of the harvested maize on those meters. This data is then processed and filtered, and the errors due to the fragility of the machines are corrected.

Among the vast data collected, the two main sources used in this study are the production results obtained by the harvesters and the climatic data from the firm's own weather station. Concerning the first source, i.e., the structural information of the parcels, the different data refers to: the different soils textures and families, elevation and irrigation information and nutrient maps. Regarding the climatic data, the weather station provided a set of daily observations measured every 10 or 15 minutes, since January first of 2020 until January first of 2021, with a total average of 123 daily observations. For this data a daily average was taken, and the data aggregated by stage as stated before.

As stated earlier, the spatial matrix used in this study is based in contiguity. For this purpose, it was chosen the queen's criterion given that, besides considering the regions that have a physical border in common, also considers the ones with vertices in common. Presented in the appendix A, Figure 7, it's possible to observe the histogram of the number of neighbors of the contiguity matrix constructed from the queen's criterion. It's possible to conclude that the more frequent number of neighbors of a spatial unit is six, followed by five neighbors and then 7 neighbors. Because the grid is too small and there is a lot of spatial units, the W matrix has the following properties presented in table II.

Table II - Spatial W matrix properties

Minimum number of neighbors	3
Maximum number of neighbors	35
Median number of neighbors	6

Source: Own elaboration; Sofware: GeoDa.

In Figure 8, from Appendix A, it’s possible to analyze the Moran dispersion diagram for the variable maize yield in tonnes per hectare, by id, through which the slope of its line corresponds to the value of Moran’s index (I=0.876). In order to access significance, it was computed a test with randomization and 999 permutations for the global Moran’s index of the variable Yield under the null hypothesis of no spatial autocorrelation between spatial units. Given the Moran’s I-statistic (t) is 0.8762 and its standardized value (z) is 359.9061, the null hypothesis is rejected at a 1% level of significance with a p-value of 0.001, and a spatial specification is the most appropriate to handle spatial dependency and/or heterogeneity in the data. That is, the test suggests positive and significant spatial dependence (Figure 9, Appendix A).

In Figure 8 it is also visible that there is a higher density of observations in the first and third quadrant, meaning that, spatial units with high values of maize yield are surrounded by spatial units that also have above average values for the variable, and spatial units with low values of maize yield are surrounded by spatial units that also have below average values for the variable. This is to say, that there is a need to include spatial dependence in the estimation.

5.2.Variables Explanation

Dependent Variable Selection

For this study, the dependent variable is the one representing the average annual maize yield measured in tonnes per hectare by id. This is determined by the dry weight of the harvested maize in a parcel/sub-parcel indexed by the number of hectares of maize planted over the time. Analyzing the spatial distribution map of the variable Yield per quantile, in Figure 4, it is possible to observe a great variability in the space of the dependent variable. Spatial units with a high maize yield appear as neighbors of spatial units with the minimum maize yield.

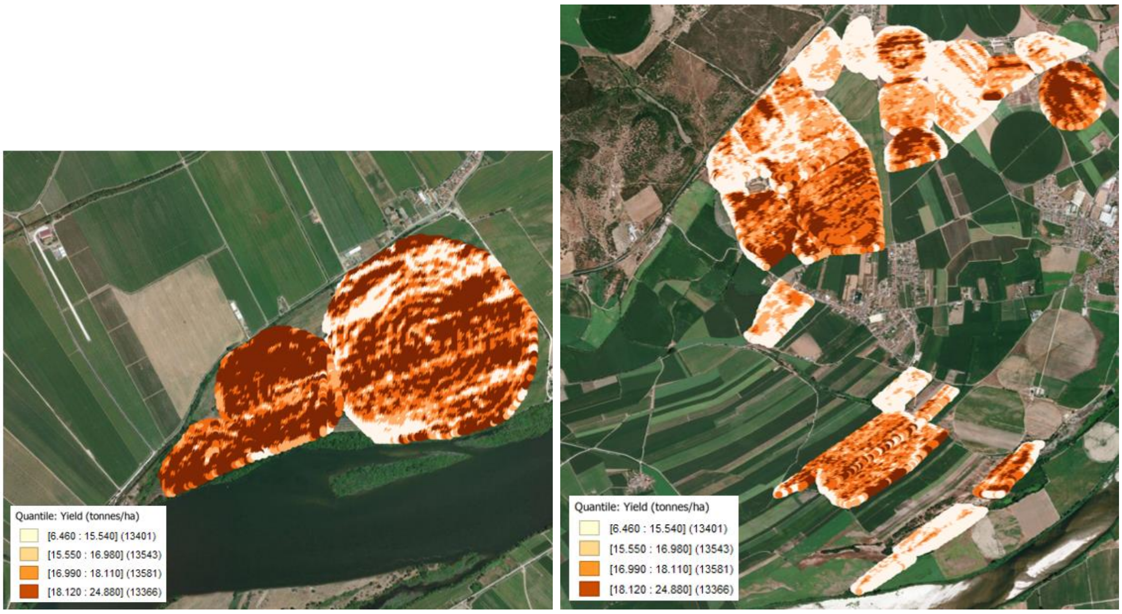

Figure 4 - Spatial distribution map of the variable Yield per quartile, in 2020

Source: Own Elaboration; Software: QGIS

Explanatory Variables Selection

Starting with the climatic data, after obtaining the daily measurements of the average temperature [T] (ºC), precipitation [P] (mm), relative humidity [H] (%RH), global solar radiation [R] (W/m²) and wind velocity [W] (km/h) for each growth stage in the Dixon et al. (1994) approach, as mentioned before, by spatial unit, this information was standardized as follows:

a) It was calculated the maximum and minimum mean and median values of each climatic variable;

b) Then, measures of dispersion such as the first and third quantile of maximum, minimum and average temperatures were calculated. The same was performed to the remaining 4 weather climatic variables;

c) Finally, variables of amplitude, interquartile and between maximum and minimum, were created, as well as a variable representing the sum of the number of days when the maximum daily temperature was above 35ºC and another one when the minimum daily temperature was below 0ºC, in all 4 stages, for all climatic variables. It´s important to mention that these values were chosen based on previously mentioned literature.

In this regard, more than one hundred climatic variables were analyzed and tested. The majority was drooped due to being a constant or insignificant in relation to the dependent variable. Therefore, for the remaining climate variables, the specification of each and their possible nonlinearity was explored.

In relation to the structural variables, the data collected and delivered by the firm Agro Analítica and Quinta da Cholda, refers to:

a) Land parcel and sub-parcel;

b) Coordinates (X and Y);

c) Sowing density (seeds/ha);

d) Variety of the maize (Dekalb or Pionner);

e) Type of soil (clay or sandy);

f) Elevation;

g) Total irrigation (mm/ha);

h) Conductivity of the soil;

i) Macronutrients: organic nitrogen (Kg/ha), mineral nitrogen (Kg/ha) and total nitrogen (organic plus mineral); organic potassium (Kg/ha), mineral potassium (Kg/ha) and total potassium (organic plus mineral); organic phosphorus (Kg/ha), mineral phosphorus (Kg/ha) and total phosphorus (organic plus mineral);

j) Applied SO3 (Kg/ha) and Organic Matter (Kg/ha);

k) Nitrates in the water (Kg/ha).

For these variables, the same transformations and procedures done to the climatic variables were computed. Most have not been shown to be significant in explaining the behavior of maize production, mostly because they are constants between the spatial units, and, as a result, were dropped from the estimation. Table III provides a summary of all the variables used in the empirical analysis after a preliminary selection.

Table III - Regression variables

VARIABLE	DESCRIPTION
Yield	The dependent variable for the analysis, maize yield (tonnes/ha)
Density	Sowing density (seeds/ha)
Elevation	The height above the sea level (km)
MOApplied	Organic matter applied (Kg/ha)
SO3Applied	Applied sulfur trioxide (SO3) (Kg/ha)
Tot_Irrig	Total Irrigation (mm/ha)
Tot_Nit	Total Nitrogen applied (Kg/ha)
Tot_Phos	Total Phosphorus applied (Kg/ha)
Tot_Pot	Total Potassium applied (Kg/ha)
HMeanMin1	Minimum average Relative Humidity on stage 1 (%RH)
RMeanFQ1	First quantile of mean Solar Radiation on stage 1 (W/m²)
HMeanMean2	Mean of the mean Relative Humidity on stage 2 (%RH)
TSup35_2	Number of days with temperatures above 35ºC on stage 2 (days)
TMinFQ3	First quartile of minimum temperatures on stage 3 (ºC)
RMeanMean3	Mean of the mean solar radiation on stage 3 (W/m²)
VMeanMax3	Maximum average wind speed on stage 3 (km/h)
Hminmax4	Range between maximum and minimum relative humidity on stage 4 (%RH)
TS	Dummy variable equal to 1 if the soil is clayey and equal to 0 when the soil is sandy
V_DKC6492	Dummy variable equal to 1 if the maize seed variety is DKC6492
V_DKC6808	Dummy variable equal to 1 if the maize seed variety is DKC6808
V_P0937	Dummy variable equal to 1 if the maize seed variety is P0937
V_P1551	Dummy variable equal to 1 if the maize seed variety is P1551
V_P1574	Dummy variable equal to 1 if the maize seed variety is P1574
V_P1772	Dummy variable equal to 1 if the maize seed variety is P1772

Source: Own elaboration; Software: Stata.

It should be noted that the base group of seed variety dummies is represented by the variety P0937 and, as such, will be omitted from the estimation of the models in order to protect from the dummy variable trap.

Furthermore, together with the variables described in the table above, the variables *SO3ApplSQ*, *TotIrrigSQ*, *TotNitSQ*, *TotPhosSQ* and *TotPotSQ* were also introduced which represent, respectively, the square transformation of *SO3Applied*, *Tot_Irrig*, *Tot_Nit*, *Tot_Phos* and *Tot_Pot*. The motivation for the creation of the listed variables is due to the suspicion of its non-linearity relationship with the dependent variable under study.

In Table VI, in Appendix B, are presented the basic summary statistics for the data where it's possible to conclude that, in the year 2020, the maize yield (*Yield*) assumed values between 6.46 and 24.88 tonnes/ha by id, with a mean of approximately 16.76 tonnes/ha.

6. ANALYSIS OF RESULTS

This chapter presents the results of the estimation of the proposed models. Due to the high number of observations and the computation requirments, the econometric estimation was performed with GeoDaSpace *software*. This software has the advantage of the possibility of estimating spatial models that control for both spatial autocorrelation and heteroscedasticity, allowing for robust estimation. In order to start the analysis on the variables that might explain the maize yield in the Azinhaga exploitation, in the year 2020, it was first estimated the OLS Model as a diagnostic model, followed by the SLM, SEM and SARAR(1,1) Model. Below, in Table IV are the coefficients, and respective significances, and standard errors of the OLS model for all the variables under study.

Table IV – *Linear OLS Model* for the dependent variable maize yield

	OLS (Non-spatial Model)	
Variables	**Coefficients**	**Std. Error**
Constant	-140.924457***	9.264613
Density	0.000049***	0.000003
Elevation	0.071802***	0.010045
HMeanMean2	0.141577**	0.060173
HMeanMin1	0.061728***	0.007041
Hminmax4	-0.077222***	0.006003
MOApplied	0.000611	0.000484
RMeanFQ1	0.012033***	0.001324
RMeanMean3	0.399709***	0.02171
SO3Applied	0.109899***	0.012606
SO3ApplSQ	-0.000496***	0.000111
TMinFQ3	1.86542***	0.092077
TS	1.162607***	0.106856
TSup35_2	-0.331708***	0.035633
Tot_Irrig	-0.047609***	0.002403
TotIrrigSQ	0.000035***	0.000002
Tot_Nit	-0.087637***	0.003208
TotNitSQ	0.000335***	0.000012
Tot_Phos	-0.060328***	0.005973
TotPhosSQ	-0.000111***	0.000024
Tot_Pot	-0.034638***	0.003915
Tot_PotSQ	0.000419***	0.000049
VMeanMax3	1.1965***	0.055723
V_DKC6492	-6.16626***	0.210507
V_DKC6808	-0.236221***	0.062065
V_P1551	-6.399422***	0.218422
V_P1574	-7.5359***	0.202598
V_P1772	-8.034694***	0.187707
Pseudo R^2	0.4028	
Nº of Observations	53891	

*Significance levels: ***1%, **5%, and *10%.*

Source: Own Elaboration; Software: GeoDaSpace.

On standard errors, the OLS model and the SLM were based on the White standard errors, and the SEM and SARAR(1,1) models were based on the KP HET standard errors (Kelejian & Prucha, 2010).

After the initial estimation of the OLS Model, a Breusch-Pagan (BP) test was performed to verify the presence of heteroscedasticity (Table VII, Appendix B). Since the *p*-value is lower than the significance level of 1%, the null hypothesis of homoscedasticity is strongly rejected and, therefore, suggests the presence of heteroscedasticity. However, the result of this test should be viewed with caution because it is distorted by the presence of spatial dependence.

Regarding the diagnostics for spatial dependence, Moran's I test was executed on the residuals (Table VIII, Appendix B). The rejection of the null hypothesis of the absence of autocorrelation indicates strong evidence supporting the existence of spatial autocorrelation in the residuals of the model. In addition, the Robust Lagrange and the Lagrange Multiplier (Robust LM and LM) were also realized to test the presence of this spatial effect. The simple LM (lag), tests for a missing spatially lagged dependent variable, whereas the LM (error), tests for error dependence. In terms of its robust approaches, the first one tests for a missing lagged variable in the possible presence of error dependence, while the second one tests for the reverse (Anselin, 1988; Anselin et al., 1996). Still in Table VIII, in Appendix B, it's possible to note that both null hypothesis of the simple LM tests are significant, resulting in the rejection of the null hypothesis of no autocorrelation suggesting, therefore, the presence of spatial dependence. The same happens with the robust tests of no spatially lagged dependent variable (LM (lag)) and the hypothesis of no spatially autocorrelated error term (LM (error)). All must be rejected at a 1% significance level.

Undoubtedly, a spatial model is needed to accommodate the two spatial effects. In Output 1 and Output 2 (Appendix C) can be found the Spatial lag and Spatial Error Models for all variables, respectively.

First analyzing Output 1 (Appendix C), of the *Spatial Lag Model*, it was introduced the Spatial Weights Matrix *W*, computed and analyzed in the previous chapter, resulting in the new spatial lag term of W_Yield. Its coefficient, Rho, measures the degree of spatial dependence between observations in the sample data (Anselin, 1988). This parameter appears in this model as positive and highly significant. As a result, the general fit of the model is greatly improved (Pseudo R-Square equal to 0.9082). However, despite this large improvement, resorting to the Anselin-Kelejian Test (Table IX, Appendix B), there's strong evidence of spatial autocorrelational in the residuals still in the model (Anselin & Kelejian, 1997). This way, although the new specification with a spatial lag term, it didn't filter all the spatial dependence in the data.

With regard to the *Spatial Error Model* (Output 2, Appendix C), the spatially correlated error variable was generated and its coefficient, *Lambda*, appeared as positive and also highly significant. However, the Pseudo R-Square of this model decreased drastically and the *Lambda* parameter is equal to 1 when it should be less than 1, to ensure a stability behavior. This corroborates the hypothesis that spatial effects have not yet been entirely modelled.

A solution to incorporate both mechanisms of spatial dependence mentioned before, is to consider the *SARAR(1,1) Model*, suggested by Kelejian and Prucha (2010), estimated through robust inference with the generalized two-step least squares estimator. In Output 3, Appendix C, it's possible to confirm the validity of the model, with the Pseudo R-Square being very high and the coefficient estimates for the *Rho* and *Lambda* parameters high and significant. On this basis, non-significant variables were successively removed from the elected SARAR(1,1) model until a restricted model was reached in which all variables were statistically significant at a 1% level (except for the intercept). The output of the elected restricted model is illustrated in Table V.

Table V - Restricted *SARAR(1,1) Model*

	Restricted SARAR(1,1) Model	
Variables	**Coefficients**	**Std. Error**
Constant	1.363689**	0.667142
Density	0.000012***	0.000002
Hminmax4	-0.035190***	0.005733
SO3Applied	0.009744***	0.001770
TSup35_2	-0.047774***	0.006197
Tot_Nit	0.003269***	0.000990
TotNitSQ	-0.000008***	0.000003
Tot_Irrig	0.006443***	0.002006
TotIrrigSQ	-0.000004***	0.000001
Tot_Pot	-0.009229***	0.002053
Tot_PotSQ	0.000040***	0.000010
VMeanMax3	0.036827***	0.014257
V_DKC6808	0.359631***	0.040937
V_P1574	-0.319528***	0.049419
Pseudo R^2	0.8959	
Nº of Observations	53891	

*Significance levels: ***1%, **5%, and *10%.*

Source: Own Elaboration; Software: GeoDaSpace.

This model presents a Pseudo R-Square of 0.8959, meaning that about 90% of the variation in maize yield is explained by the independent variables and the spatially lagged variables included in the model. It's also worth noting that, in variables such as *HMeanMean2*, *TMinFQ3* and *RMeanMean3*, due to the fact that they vary so little in the sample, there is not enough information to conclude their non-significance by the absence of causality in the year 2020. All variables are statistically significant at a 1% level and, both the coefficient of the spatially lagged dependent variable (*W_Yield*), *Rho*, and, the coefficient of the spatially correlated error, *Lambda*, are positive and highly significant.

Starting the analysis with the variable *Density*, on average, spatial units with a 100 seed per hectare increase with regards to sowing density, increases maize yield by 12 kg/ha, in the same spatial unit, *ceteris paribus*. As far as the total irrigation, total nitrogen and potassium applied, one must also take into account its squares. About the nutrients, on average, an increase of 1 Kg/ha of SO3 applied in a spatial unit, increases maize yield by 9,7 kg/ha regarding the same spatial unit, all other variables being equal.

Regarding *TotIrrigSQ*, its negative coefficient indicates a maximum (turning point) of the parabola. This maximum is given by the symmetric of the ratio between *Tot_Irrig* coefficient and the double of the *TotIrrigSQ* coefficient. As a result, the turning point is 870.6351 mm, meaning that before the turning point, the maize yield rises when the total irrigation increases and, after that point, as the total irrigation increases the maize yield decreases, *ceteris paribus*. Given that this variable ranges between 556.3 and 911.5 mm/ha and presents a mean of 641.1 mm/ha (Table IV, Appendix B), the value at which the yield starts to decrease is well above the mean of the *Tot_Irrig* variable, being for the most part positively correlated with the yield.

The same happens to *TotNitSQ* that presents a maximum at 196.9458 kg/ha. Initially, an increase of total nitrogen applied causes an increase the yield and, after the dosage of the turning point, as the total nitrogen applied increases the maize yield decreases. Ranging between 32.4 and 318.9 Kg /ha with a mean of 260.16 Kg/ha, the turning point is below the average.

Finally, due to the positive coefficient associated with the variable *TotPotSQ*, the parabola presents a minimum turning point. This value stands at 115.6466 kg/ha and now, as the total potassium applied increases, the maize yield decreases. After the turning point, an increase of total potassium causes the maize yield to rise, *ceteris paribus*. Also looking at Table VI, in Appendix B, the total potassium applied goes from 0 to 180.88 Kg /ha, placing the turning point slightly under the mean of the variable.

For the varieties of maize planted in 2020, only two of them were significant in the SARAR(1,1) restricted model. On average, spatial units where the variety DKC6808 was planted, increased maize yield by almost 360 kg/ha tonnes/ha comparing to when the variety P0937 was applied, *ceteris paribus*. Also, spatial units where the variety P1574 was planted, decreased maize yield by almost 320 kg/ha then when the variety P0937 is applied, *ceteris paribus*.

Regarding the climatic variables, it was only possible to assess statistically significance in three variables. On average, a one unit increase of the range between maximum and minimum relative humidity, that is, an increase of 1 percentage point in the range between the maximum and minimum relative humidity, on stage 4, decreases maize yields by 35.2 kg/ha regarding a spatial unit, all other variables being equal. In the same way, on average, spatial units where there´s an increase of 1 more day with temperatures above 35ºC, on stage 2, decreases the maize yield by almost 48 kg/ha, in that spatial unit, with all other variables remaining constant. And finally, on average, an increase of 1 km/h of the maximum average wind speed, on stage 3, in one spatial unit, increases the maize yield by almost 37 kg/ha, in that spatial unit, *ceteris paribus*.

7. CONCLUSIONS, LIMITATIONS AND FUTURE WORKS

This research sought to apply different spatial regression models to estimate a function for the maize yield with the aim to identify the relevant determining factors, and their effect, on maize yield on the exploitation of Quinta da Cholda firm, in Azinhaga, Golegã, district of Santarém, Portugal for the year 2020.

To this end, data analysis and cross-section estimations were performed with different estimation methods (OLS and GS2SLS) in order to find the most suitable spatial specification to describe the data. Firstly, by means of the global Moran's I, the existence of spatial autocorrelation of the dependent variable, in the year 2020, was verified employing the spatial weights matrix construction criterion queen. With this statistic it was possible to conclude that regions with high maize yield are neighbors of regions also with high values of yield. In contrast lower yield regions, present neighbors with lowers yields as well.

Following this, the OLS diagnostic and spatial econometric models were estimated, were the *SARAR(1,1) Model* was chosen since it is the only one that allows the incorporation of both mechanisms of spatial dependence. Considering the cross-sectional data analysis done earlier, after the elimination of non-significant variables, the majority of the results produced by the restricted *SARAR(1,1) Model*, shown to be in agreement with the existing literature on maize crop.

For the variables considered, *Density* is one of the variables that has a positive effect on maize yield. As reviewed on the literature, and modeled in several studies, an increase in the sowing density reflects on an increase of maize yield. At low densities many modern hybrid maize varieties do not grow as effectively as they do with a larger sowing population. Similarly, although small, the applied SO3 nutrient also presents a positive effect on the dependent variable. Just as essential as the other macronutrients discussed, sulfur trioxide deficiencies are reflected in suboptimal yield.

With respect to total irrigation, this variable influence was well captured by a quadratic function whose parabola has a maximum. This means that, from a certain point, high values of irrigation have a negative impact in the maize yield. This value is very close to the maximum quantity of irrigation applied by the firm and, allows to conclude, as in the literature, that too much water can be harmful to the crop, creating situations of waterlogged land. This means that the quantity of irrigation chosen, depends on the terrain topography as well as other factors such as density and precipitation, as stated by the literature reviewed. The same nonlinearity with the dependent variable happens to the macronutrient's nitrogen and potassium. The first one is also well described by a parabola, whose maximum is now not so close to the maximum applied value. Nitrogen is a very soluble compound that is easily washed away by the excessive irrigation or precipitation, being very difficult to be properly measured, an armful at very high quantities. As for the potassium, this presents a minimum in the respective quadratic function. Being the most absorbed macronutrient after nitrogen, in line with nonlinear relation with maize yield, only after a certain value of applied potassium (minimum) does the yield start to increase alongside with the nutrient. Since potassium contributes to the improvement of the quality of the maize and hence its yield, higher values of this nutrient make the maize yield rise.

And finally, each stage of corn, except for stage 1, presented a statistically significant independent variable. For stage 2, the number of days with temperatures above

35ºC has a negative implication on the maize yield. According to the literature reviewed, in this stage, the early vegetative growth stage, a large number of days with temperatures above the indicated threshold, cause the aerial part of the plant to die. For stage 3, the maximum wind speed presents a positive effect on the yield. This is due to the fact that during this stage, the flowering stage, wind can indeed help spreading the pollen shed to improve pollination, resulting in a more optimal final maize yield. And finally, an increase in the range between the minimum and maximum relative humidity, on stage 4, negatively influenced the maize yield. Relative humidity variables are poorly documented in the literature. However, this result is expected because, in the last stage of the maize life cycle, kernels moisture contents at maturity must be in average 30% in order to be harvested and the maize yield calculated. This way, an increase in the air relative humidity is likely to degenerate this process.

Regarding the seeds used, their influence on maize yield depends on all the factors observed in this study such as the soil characteristics, fertilization applied and the climatic conditions, in the year 2020, in the exploitation in question. No literature has been found on the inclusion of seed variety in estimation models. This research concluded that, for the Dekalb brand, the DKC6808 variety was shown to have a more positive effect on maize yield then when the P0937 variety, of the Pionner brand, is seeded. However, when the P1574 variety is planted, the yield is lower than when the P0937 variety is seeded.

One limitation encountered throughout the development of this essay was the lack of information in some spatial units. In particular, variables such as soil pH and amounts of pesticides and herbicides administered, could not be included in the analysis since these variables only had data for some spatial units.

In future development of this research, it would be interesting to include in the analysis the application of herbicides and pesticides measured by development stages and see their effect on the yield. In the same line of thinking, it could be performed a more economic analysis with variables such as costs and corn profitability.

REFERENCES

Afyuni, M. M., Cassel, D. K., & Robarge, W. P. (1993). Effect of landscape position on soil water and corn silage yield. *Soil Science Society of America Journal 57*, 1573–1580.

AGRO ANALÍTICA (2021). *Conceito* [In line]. Available at: https://www.agro-analitica.com/conceito/ [Access in: 03/05/2021].

Almeida, E. (2012). *Econometria Espacial Aplicada*, 1ª Ed. Guanabara: Editora Alínea.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht, Boston: Kluwer Academic Publishers.

Anselin, L., & Florax, R. (1995). Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results. In L. Anselin & R. Florax (Eds.), *New Directions in Spatial Econometrics* (pp. 21-74). Berlin, Heidelberg: Springer.

Anselin, L. (1996). The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In M. Fischer, H. Scholten & D. Unwin (Eds.), *Spatial Analytical Perspectives on Gis* (pp. 111-125). London: Taylor Francis.

Anselin, L., Bera, A. K., Florax, R. & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26 (1), 77-104.

Anselin, L., & Kelejian, H. (1997). Testing for spatial error autocorrelation in the presence of endogenous regressors. *International Regional Science Review, 20,* 153-182.

Anselin, L., & A. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D.E.A. Giles (Eds.), *Handbook of Applied Economic Statistics (*pp. 237–289). New York: Marcel Dekker.

Anselin, L. (2003). Spatial Econometrics. In Badi H. Baltagi (Eds.), *A Companion to Theoretical Econometrics* (pp. 310-316). Wiley Blackwell Publishing.

Anselin, L. (2005). *Exploring Spatial Data with GeoDa: A Workbook*. Illinois: Center for Spatially Integrated Social Science.

Anselin, Luc, Ibnu Syabri and Youngihn Kho (2006). GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis 38 (1), 5-22.

ANPROMIS (2017). *O milho* [In line]. Available at: https://www.anpromis.pt/o-milho.html [Access in: 03/04/2021].

Bellido, L.L. (1991). *Cultivos herbáceos – Cereales* (Vol.1). Madrid: Mundi-Prensa.

Barros, J., & Calado, J. (2014). *A cultura do milho*. Texto apoio, Escola de ciências e tecnologia - Universidade de Évora, Portugal, 4-33.

Bockstael, N. E. (1996). Modeling Economics and Ecology: The Importance of a Spatial Perspective. *American Journal of Agricultural Economics*, *78*(5), 1168.

Bongiovanni, R., & Lowenberg-Deboer, J. (2002). Economics of nitrogen response variability over space and time: results from 1999-2001 field trials in Argentina. In *Proceedings of the 6th International Precision Agriculture Conference*, Minneapolis, MN, 1826-1841.

Chhetri, S., Poudel, A., & Kandel, B. P. (2018). Effect of Sowing Dates on Grain Yield and Other Agronomic Traits of Different Maize Inbred Lines. *International Journal of Scientific & Engineering Research*, *9* (5), 428-434.

Cliff, A.D., & Ord, J.K. (1973). *Spatial autocorrelation*. London: Pion.

Cliff, A. D., & Ord, J. K. (1981). *Spatial processes: Models and applications*. London: Pion Limited.

Cruz, J. C., Pereira Filho, I. A., Alvarenga, R. C., Gontijo Neto, M. M., Oliveira, M. F.(2008). Manejo da Cultura do Milho. Ministério da Agricultura, Pecuária e Abastecimento. In J. C. Cruz (Ed.), *Cultivo do milho* (4.ª ed) (pp. 1-5). Sete Lagoas: Embrapa Milho e Sorgo.

Daughtry, C. S. T., Gallo, K. P., & Bauer, M. E. (1983). Spectral Estimates of Solar Radiation Interception by Corn Canopies. *Agronomy* Journal *74*, 527-531.

Dekalb Asgrow Deltapine (2020). Determining corn growth stages. [In line]. Available at: https://www.dekalbasgrowdeltapine.com/en-us/agronomy/corn-growth-stages-and-gdu-requirements.html [Acess in: 05/04/2021].

Dixon, B. L., Hollinger, S. E., Garcia, P., & Tirupattur, V. (1994). Estimating Corn Yield Response Models to Predict Impacts of Climate Change. *Journal of Agricultural and Resource Economics*, *19* (1), 58-68.

FAOSTAT (2021). Crops and livestock products [Data Base], September 2021. Available at: http://www.fao.org/faostat/en/#data/QCL/visualize .

FAO, 2020. *Crop Yields* [Data Base], September 2021. Available at: https://ourworldindata.org/crop-yields .

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician, 5* (3), 115-145.

Huff, F. A., & J. C. Neill (1980). Assessment of Effects and Predictability of Climate Fluctuations as Related to Agricultural Production. Final Rep., Phase I, State Water Survey Div., University of Illinois, Champaign.

IBM Corp. Released 2021. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp.

ISPA (2021). *Precision Ag Definition* [In line]. Available at: https://www.ispag.org/about/definition [Access in: 27/06/2021].

Kaspar, C. T., Colvin, S. T., Jaynes, D. B., Karlen, D. L., James, D. E., Meek, D. W., Pulido, D., & Butler, H. (2003). Relationship Between Six Years of Corn Yields and Terrain Attributes. *Precision Agriculture*, *4*, 87–101.

Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, *17*(1), 99-121.

Kelejian, H. H., & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, *157*, 53–67.

Kravchenko, A. N., & Bullock, D. G. (2000). Correlation of corn and soybean grain yield with topography and soil properties. *Agronomy Journal*, *92*, 75-83.

Lambert, D., Lowenberg-DeBoer, J., & Malzer, G. (2006). Economic analysis of spatial-temporal patterns in corn and soybean response to nitrogen and phosphorus. *Agronomy Journal, 98* (1), 43-54.

Lark, R. M., Stafford, J. V., & Bolam, H. C. (1997). Limitations on the spatial resolution of yield mapping for combinable crops. *Journal of Agricultural Engineering Research, 66*, 183–193.

Liu, Y., Swinton, S. M., & Miller, N. (2006). Is site-specific yield response consistent over time? Does it pay? *American Journal of Agricultural Economics, 88* (2), 471-483.

Moran, P. A. P. (1948). The interpretation of Statistical Maps. *Journal of the Royal Statistical Society, Series B*, *10* (2), 243-251.

Offutt, S. E., P. Garcia, & M. Pinar (1987). Technological Advance, Weather, and Crop Yield Behavior. *North Central Journal of Agricultural Economics*, *9* (1), 49-63.

Pionner (2021). *Staging Corn Growth* [In line]. Available at: https://www.pioneer.com/us/agronomy/staging_corn_growth.html [Access in: 05/04/2021].

QGIS.org, 2021. QGIS Geographic Information System. QGIS Association.

Quinta da Cholda, 2021. Sustentabilidade Sustentável [In line]. Available at: https://quintadacholda.pt/ [Access in: 03/05/2021].

Sárvári, M., & Futó, Z. (2001). Correlation between the sowing date, yield and grain content of maize hybrids on chernozem soil. *Acta Agraria Debreceniensis J., 1,* 32-41.

Seffrin, R. (2017). *Análise Exploratória de Dados Espaciais Aplicada a Produtividade de Milho no Estado do Paraná* (Dissertação). Universidade Tecnológica Federal do Paraná.

Spomer, R. G., & Piest, R. F. (1982). Soil productivity and erosion of Iowa loess soils. *Transactions of the American Society of Agricultural Engineers, 25*, 1295-1299.

Stafford, J.V., & Bolam, H.C. (1998). Near-ground and aerial radiometry imaging for assessing spatial variability in crop condition. In P. C. Robert et al. (Eds.), *Proceedings of the 4th International Conference on Precision Agriculture* (pp. 291-302). ASA, CSSA, and SSSA, Madison, WI.

StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC.

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

Sudduth, K. A., Drummond, S. T., Birrell, S. J., & Kitchen, N. R. (1997). Spatial modeling of crop yield using soil and topographic data. In J. V. Stafford (Ed.), *Precision Agriculture '97: Proceedings of the 1st European Conference on Precision Agriculture* (pp. 439-447). Oxford, UK: BIOS Scientific Publishers.

Appendix A

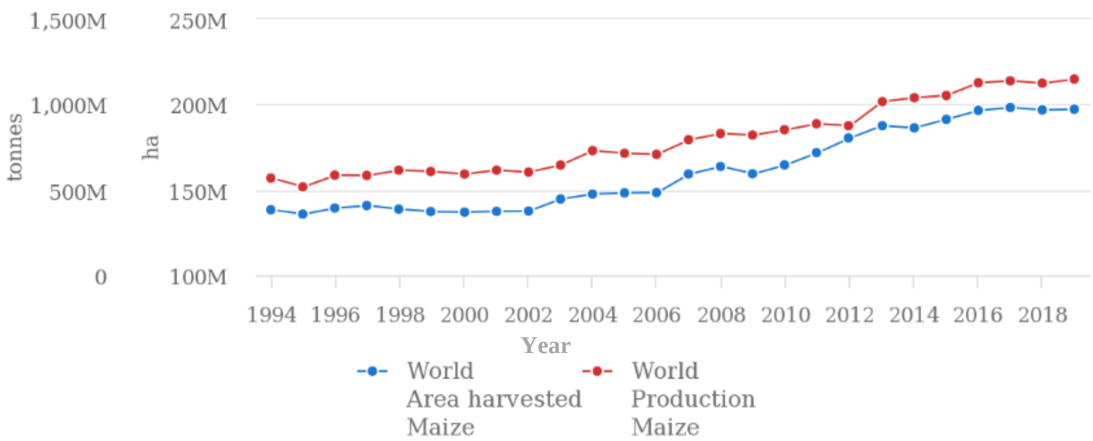

Figure 5 - Production quantities of maize in world, 1994-2019

Source: FAOSTAT (2021).

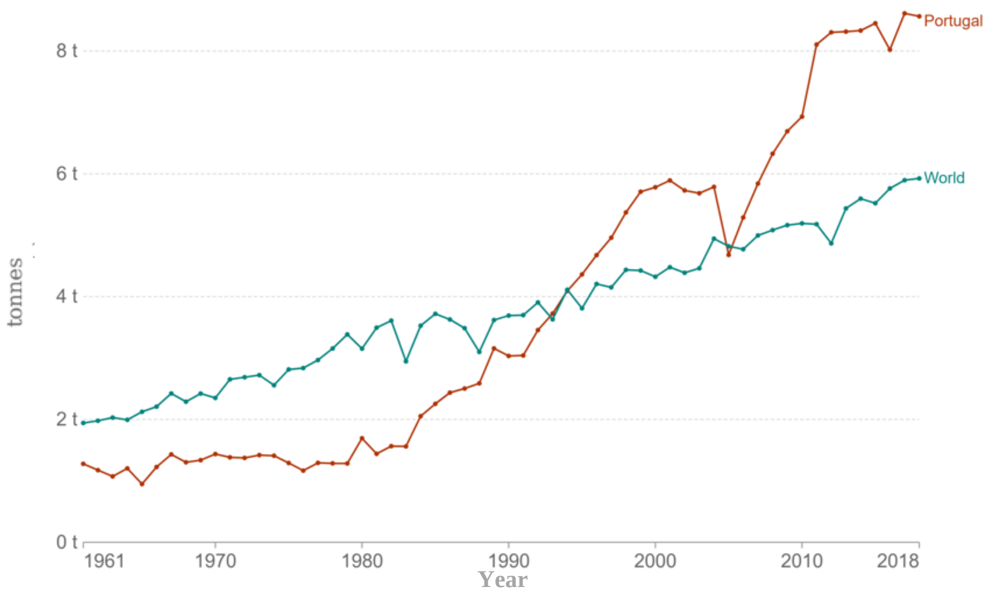

Figure 6 - Average maize yield in Portugal and the world, 1961-2018

Source: UN Food and Agriculture Organization (FAO, 2020)

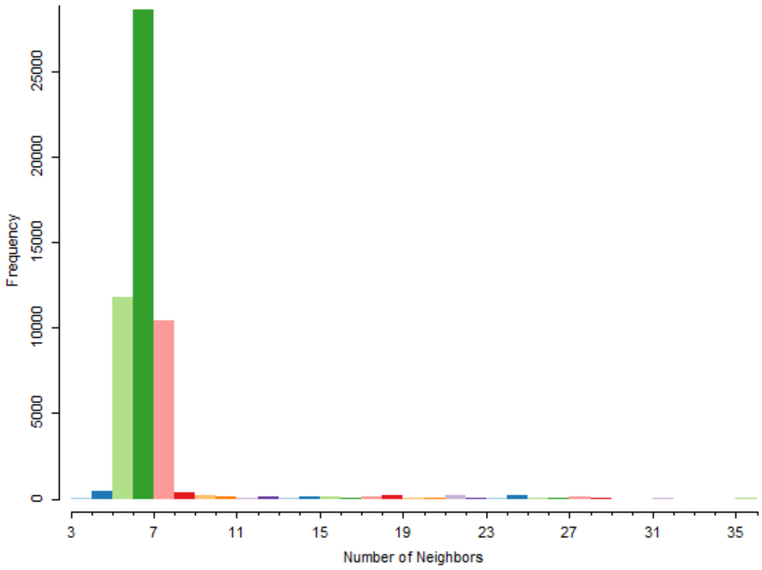

Figure 7 - Histogram of the number of neighbors of the queen contiguity matrix

Fonte: Own elaboration; Software: GeoDa.

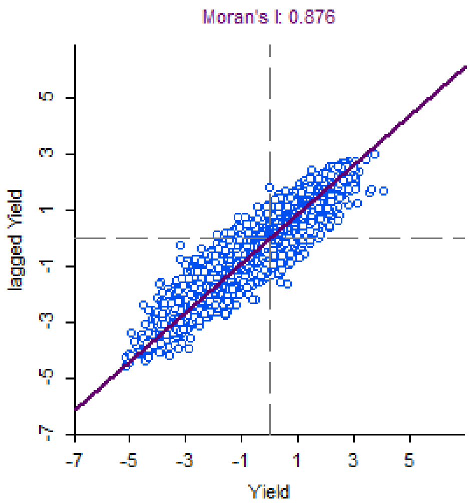

Figure 8 - Moran dispersion diagram of the variable *Yield*

Source: Own elaboration; Software: GeoDa

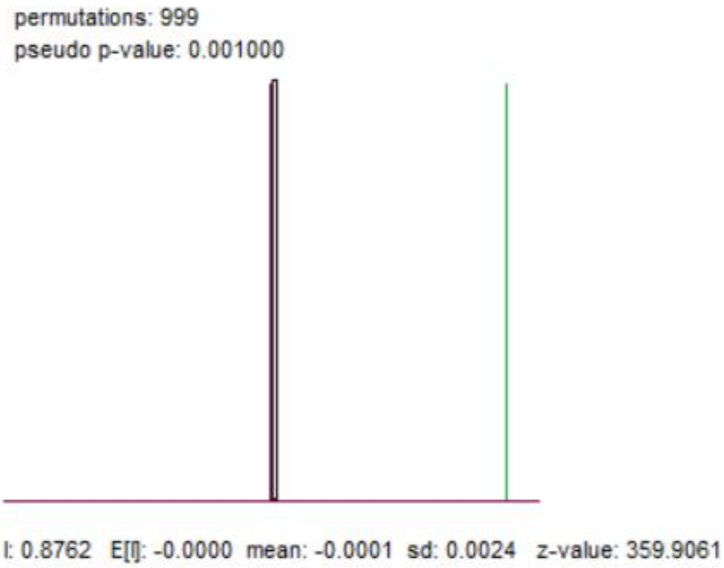

Figure 9 - Significance test for the global moran index of the variable *Yield* under the assumption of absence of spatial autocorrelation

Source: Own elaboration; Software: GeoDa

APPENDIX B

Table VI - Summary statistics of the regression variables

Variable	Obs	Mean	Std. dev.	Min	Max
Yield	53,891	16.75507	2.002042	6.46	24.88
Density	53,891	88414.34	2899.101	70000	125000
TS	53,891	.5304225	.4990782	0	1
Tot_Irrig	53,891	641.0956	58.60684	556.3	911.5
Elevation	53,891	55.80501	26.06193	1.1	73.18
Tot_Nit	53,891	260.1625	71.02801	32.4	318.9
Tot_Phos	53,891	106.0019	42.12934	0	143.6
Tot_Pot	53,891	124.4212	54.75174	0	180.88
SO3Applied	53,891	41.81926	11.19382	29.4	77
MOApplied	53,891	1251.686	1220.018	0	3351.6
RMeanFQ1	53,891	177.9588	100.5307	61.73	332
TSup35_2	53,891	7.572693	5.485655	2	18
TMinFQ3	53,891	12.3364	.9533009	10.8	14
RMeanMean3	53,891	316.5332	7.654035	305.11	324.93
VMeanMax3	53,891	3.989184	1.541798	1.73	5.6
HMeanMin1	53,891	66.12295	6.316151	57.4	83.92
HMeanMean2	53,891	75.05453	2.545814	71.55	78.64
Hminmax4	53,891	33.16561	2.242951	32.44	41.13
V_DKC6492	53,891	.0793825	.2703373	0	1
V_DKC6808	53,891	.0240485	.1532013	0	1
V_P0937	53,891	.4345995	.4957089	0	1
V_P1551	53,891	.2147112	.4106256	0	1
V_P1574	53,891	.135904	.3426897	0	1
V_P1772	53,891	.0214322	.1448213	0	1
TotNitSQ	53,891	72729.39	26414.49	1049.76	101697.2
TotPhosSQ	53,891	13011.25	7633.742	0	20620.96
TotIrrigSQ	53,891	414438.3	80201.27	309469.7	830832.3
TotPotSQ	53,891	18478.34	12245.19	0	32717.58
SO3ApplSQ	53,891	1874.15	1102.724	864.36	5929

Source: Own Elaboration; Software: Stata.

Table VII - Breusch-Pagan Test to the OLS Model

	Statistic Value	*P*-value
Breusch-Pagan Test	10322.191	0.0000

Source: Own Elaboration; Software: GeodaSpace.

Table VIII - Spatial dependence diagnostic tests for OLS Model

	Statistic Value	*P*-value
Moran's I (error)	328.653	0.0000
LM (lag)	106331.325	0.0000
LM (error)	107636.573	0.0000
Robust LM (lag)	101.228	0.0000
Robust LM (error)	1406.476	0.0000

Source: Own Elaboration; Software: GeodaSpace.

Table IX - Anselin-Kelejian test to the Spatial Lag Model

	Statistic Value	*P*-value
Anselin-Kelejian Test	57.805	0.0000

Source: Own Elaboration; Software: GeodaSpace.

APPENDIX C

Output 1 - *Spatial Lag Model* (SLM) for all variables

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :DataBase.dbf
Weights matrix      : File: W.gal
Dependent Variable  :      Yield                Number of Observations:       53891
Mean dependent var  :    16.7551                Number of Variables   :          29
S.D. dependent var  :     2.0020                Degrees of Freedom    :       53862
Pseudo R-squared    :     0.9101
Spatial Pseudo R-squared:  0.3943

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT     -19.0269092       5.9100538      -3.2194139       0.0012845
             Density       0.0000099       0.0000019       5.1609197       0.0000002
           Elevation      -0.0071705       0.0050511      -1.4195782       0.1557305
         HMeanMean2        0.0353815       0.0274451       1.2891748       0.1973373
          HMeanMin1        0.0128914       0.0038520       3.3467199       0.0008177
            Hminmax4      -0.0113692       0.0034243      -3.3201398       0.0008997
           MOApplied      -0.0007223       0.0001991      -3.6271970       0.0002865
            RMeanFQ1       0.0018003       0.0007366       2.4442487       0.0145154
         RMeanMean3        0.0533001       0.0159815       3.3351122       0.0008526
          SO3ApplSQ       -0.0001301       0.0000463      -2.8111890       0.0049359
          SO3Applied       0.0225427       0.0061656       3.6562153       0.0002560
             TMinFQ3       0.2542673       0.0757189       3.3580433       0.0007850
                  TS       0.0879092       0.0627125       1.4017801       0.1609809
           TSup35_2       -0.0437929       0.0194551      -2.2509724       0.0243873
         TotIrrigSQ        0.0000061       0.0000013       4.5758617       0.0000047
           TotNitSQ        0.0000487       0.0000124       3.9393589       0.0000817
          TotPhosSQ       -0.0000358       0.0000112      -3.2150151       0.0013044
           TotPotSQ        0.0001378       0.0000226       6.0917366       0.0000000
           Tot_Irrig      -0.0082439       0.0018315      -4.5012591       0.0000068
             Tot_Nit      -0.0130581       0.0031898      -4.0936754       0.0000425
            Tot_Phos      -0.0046444       0.0035532      -1.3071015       0.1911782
             Tot_Pot      -0.0108174       0.0019441      -5.5643327       0.0000000
           VMeanMax3       0.1521188       0.0491607       3.0943182       0.0019727
          V_DKC6492       -0.8699546       0.2266263      -3.8387189       0.0001237
          V_DKC6808        0.0596599       0.0233426       2.5558375       0.0105933
             V_P1551      -0.9173751       0.2375286      -3.8621676       0.0001124
             V_P1574      -1.0801094       0.2693890      -4.0094783       0.0000609
             V_P1772      -1.1971934       0.2811800      -4.2577470       0.0000206
             W_Yield       0.8669735       0.0331499      26.1531487       0.0000000
------------------------------------------------------------------------------------
Instrumented: W_Yield
Instruments: W_Density, W_Elevation, W_HMeanMean2, W_HMeanMin1, W_Hminmax4,
             W_MOApplied, W_RMeanFQ1, W_RMeanMean3, W_SO3ApplSQ,
             W_SO3Applied, W_TMinFQ3, W_TS, W_TSup35_2, W_TotIrrigSQ,
             W_TotNitSQ, W_TotPhosSQ, W_TotPotSQ, W_Tot_Irrig, W_Tot_Nit,
             W_Tot_Phos, W_Tot_Pot, W_VMeanMax3, W_V_DKC6492, W_V_DKC6808,
             W_V_P1551, W_V_P1574, W_V_P1772
```

Source: Own Elaboration; Software: GeoDaSpace

Output 2 - *Spatial Error Model* (SEM) for all variables

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HET)
---------------------------------------------------------
Data set            :DataBase.dbf
Weights matrix      : File: W.gal
Dependent Variable  :      Yield               Number of Observations:       53891
Mean dependent var  :    16.7551               Number of Variables   :          28
S.D. dependent var  :     2.0020               Degrees of Freedom    :       53863
Pseudo R-squared    :     0.1629
N. of iterations    :          1               Step1c computed       :         Yes
```

Variable	Coefficient	Std.Error	z-Statistic	Probability
CONSTANT	30058918664930.7578125	15660590286773.8906250	1.9193988	0.0549339
Density	0.0000064	0.0000020	3.2178249	0.0012917
Elevation	-0.8155639	0.1012221	-8.0571693	0.0000000
HMeanMean2	0.2090032	0.6192039	0.3375354	0.7357133
HMeanMin1	0.1595815	0.0392875	4.0618911	0.0000487
Hminmax4	-0.0395500	0.0256666	-1.5409134	0.1233379
MOApplied	-0.0409051	0.0044414	-9.2098885	0.0000000
RMeanFQ1	0.0296245	0.0111117	2.6660770	0.0076742
RMeanMean3	0.4051226	0.2697185	1.5020198	0.1330920
SO3ApplSQ	-0.0049054	0.0013973	-3.5107604	0.0004468
SO3Applied	0.6463073	0.1570899	4.1142514	0.0000388
TMinFQ3	2.7647548	0.8606186	3.2125202	0.0013158
TS	-2.4522154	0.9209229	-2.6627804	0.0077498
TSup35_2	-0.8090102	0.3396518	-2.3818811	0.0172245
TotIrrigSQ	0.0000016	0.0000022	0.7447719	0.4564096
TotNitSQ	0.0006221	0.0001499	4.1512131	0.0000331
TotPhosSQ	-0.0003510	0.0001572	-2.2325492	0.0255787
TotPotSQ	0.0043876	0.0004253	10.3171079	0.0000000
Tot_Irrig	-0.0018453	0.0032430	-0.5690058	0.5693522
Tot_Nit	-0.1614152	0.0421721	-3.8275314	0.0001294
Tot_Phos	-0.0658848	0.0419384	-1.5709923	0.1161844
Tot_Pot	-0.2879117	0.0292522	-9.8424081	0.0000000
VMeanMax3	1.7753383	0.3896337	4.5564284	0.0000052
V_DKC6492	-12.2579281	2.0249731	-6.0533783	0.0000000
V_DKC6808	0.5818470	0.0568955	10.2265837	0.0000000
V_P1551	-12.3379613	2.0476527	-6.0254170	0.0000000
V_P1574	-12.8451656	2.0510217	-6.2628131	0.0000000
V_P1772	-13.2441322	2.0651181	-6.4132566	0.0000000
lambda	1.0000000	0.0000000	171689943514012800.0000000	0.0000000

Source: Own Elaboration; Software: GeoDaSpac

Output 3 - *SARAR(1,1) Model* for all variables

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
-------------------------------------------------------------------
Data set            :DataBase.dbf
Weights matrix      : File: W.gal
Dependent Variable  :     Yield                Number of Observations:       53891
Mean dependent var  :   16.7551                Number of Variables   :          29
S.D. dependent var  :    2.0020                Degrees of Freedom    :       53862
Pseudo R-squared    :    0.7822
Spatial Pseudo R-squared:  0.3866
N. of iterations    :         1                Step1c computed       :         Yes
```

Variable	Coefficient	Std.Error	z-Statistic	Probability
CONSTANT	-79.4405460	20571792324831.4140625	-0.0000000	1.0000000
Density	0.0000125	0.0000020	6.1934004	0.0000000
Elevation	-0.0340928	0.1144671	-0.2978391	0.7658260
HMeanMean2	0.0533121	0.7036056	0.0757699	0.9396022
HMeanMin1	0.0547498	0.0415438	1.3178799	0.1875439
Hminmax4	-0.0387034	0.0283402	-1.3656734	0.1720415
MOApplied	-0.0034940	0.0048983	-0.7133160	0.4756502
RMeanFQ1	0.0072276	0.0130661	0.5531555	0.5801570
RMeanMean3	0.2297653	0.3079264	0.7461697	0.4555649
SO3ApplSQ	-0.0007528	0.0015516	-0.4851985	0.6275356
SO3Applied	0.1183950	0.1751961	0.6757854	0.4991769
TMinFQ3	1.0510301	0.9748126	1.0781868	0.2809504
TS	0.3386527	1.0546635	0.3211002	0.7481344
TSup35_2	-0.2167490	0.4126324	-0.5252836	0.5993861
TotIrrigSQ	0.0000113	0.0000024	4.7612607	0.0000019
TotNitSQ	0.0002189	0.0001780	1.2299771	0.2187057
TotPhosSQ	-0.0001473	0.0001790	-0.8230573	0.4104754
TotPotSQ	0.0006228	0.0004562	1.3650166	0.1722478
Tot_Irrig	-0.0151525	0.0034841	-4.3490802	0.0000137
Tot_Nit	-0.0587254	0.0501514	-1.1709626	0.2416138
Tot_Phos	-0.0213201	0.0496753	-0.4291904	0.6677847
Tot_Pot	-0.0467439	0.0324718	-1.4395205	0.1500031
VMeanMax3	0.7034122	0.4573610	1.5379802	0.1240534
V_DKC6492	-4.0420090	2.3969900	-1.6862853	0.0917409
V_DKC6808	0.2614922	0.0646582	4.0442216	0.0000525
V_P1551	-4.2175247	2.4197933	-1.7429276	0.0813463
V_P1574	-4.7822285	2.4265271	-1.9708119	0.0487454
V_P1772	-5.2207570	2.4537042	-2.1277043	0.0333616
W_Yield	0.4785093	0.1535741	3.1158204	0.0018343
lambda	1.0000000	0.0183432	54.5162487	0.0000000

```
Instrumented: W_Yield
Instruments: W_Density, W_Elevation, W_HMeanMean2, W_HMeanMin1, W_Hminmax4,
             W_MOApplied, W_RMeanFQ1, W_RMeanMean3, W_SO3ApplSQ,
             W_SO3Applied, W_TMinFQ3, W_TS, W_TSup35_2, W_TotIrrigSQ,
             W_TotNitSQ, W_TotPhosSQ, W_TotPotSQ, W_Tot_Irrig, W_Tot_Nit,
             W_Tot_Phos, W_Tot_Pot, W_VMeanMax3, W_V_DKC6492, W_V_DKC6808,
             W_V_P1551, W_V_P1574, W_V_P1772
```

Source: Own Elaboration; Software: GeoDaSpace