Analysis of spatial and temporal variability of maize productivity based on physical parameters

Manuel Barreiros Martins Moura Neves

Dissertation for obtaining the master's degree in
Engenharia Agronómica

Supervisors: Dr.Joost Van Heerwaarden (Wageningen University & Research- Plant Productions Systems)
Dr. José Paulo Pimentel de Castro Coelho (Instituto Superior de Agronomia-Universidade de Lisboa)

Jury:

President: Doctor Maria do Rosário da Conceição Cameira, Associate Professor at Instituto Superior de Agronomia, Universidade de Lisboa

Members: Doctor José Paulo Pimentel de Castro Coelho, Associate Professor with Habilitation at Instituto Superior de Agronomia, Universidade de Lisboa

Doctor Gonçalo Pereira Fernandes Caleia Rodrigues, Assistant Professor at Instituto Superior de Agronomia, Universidade de Lisboa

2021

Acknowledgments

Since this master's thesis is the final phase of my course, I could not fail to thank all those who accompanied me during my academic journey at the Instituto Superior de Agronomia and for my short stint at Wageningen University.

To my advisors, especially to Dr. Joost Van Heerwaarden, who so kindly welcomed me at Wageningen University, who encouraged me to learn more and more and who had the greatest patience and attention with me, even after having returned to Portugal due to the pandemic.

To my ISA advisor, Professor José Paulo Pimentel, who has always been available throughout these months, with all the help and advice I needed.

To Dr.Maria João Paulo Silva, for facilitating access to Wageningen University and for being willing to solve problems and doubts that arose during this work.

To Eng. João Coimbra, who so promptly provided the data, thus entrusting me with years of work developed in Quinta da Cholda, Sa and who helped me whenever I needed.

To Agro Analítica, in the person of António Coimbra, who has always been available to clarify all doubts regarding the data.

To all my friends who accompanied me at this stage, especially Simão Coimbra and Júlia Gonçalves, who were pioneers in the way, in this way, to the University of Wageningen. To all of you, thank you for the unity and companionship over these years

To my parents and my brother, because, in addition to having had unlimited availability to support me in this adventure it is thanks to them that I have gained the taste for the agricultural world and they are the ones who inspire me and teach me the most.

To Mariana, for her patience and unconditional support along throughout this period.

Resumo alargado

Ao longo da história, a agricultura tem sofrido bastantes evoluções, nomeadamente nos períodos pós-revolução industrial, com o acesso à maquinaria agrícola e com a automatização de vários processos, e após a segunda guerra mundial, com o aparecimento dos produtos de síntese, em particular dos fitofarmacêuticos e dos fertilizantes. O seu crescimento em larga escala e a elevada intensidade de utilização daqueles fatores tem tido consequências produtivas, económicas e ambientais muito significativas e positivas, sendo que também tenham comportado acréscimos de riscos para os agricultores, uma vez que a custo-eficácia da sua aplicação é afetada pela variação espacial e temporal dos recursos do solo, das necessidades das culturas, entre outros.

Atualmente, a maioria dos agricultores não tem em conta a variação existente nos seus campos, acabando por tratar grandes áreas como elementos uniformes e, em consequência, aplicando a mesma quantidade de fatores de produção em todo o campo, ignorando fontes de variabilidade como a topografia, o tipo e as caraterísticas do solo, a incidência de pragas e doenças, etc., que afetam a produtividade tanto a nível quantitativo como qualitativo.

O aparecimento do conceito de agricultura de precisão, e das ferramentas que o integram, veio contrariar as práticas uniformes e menos sustentáveis, reduzir o risco e permitir uma abordagem agronómica adaptada à heterogeneidade local. Estes avanços, têm permitido fornecer ao agricultor o acesso a várias camadas de informação, tanto a nível de dados de rendimento/produtividade das culturas, fatores edáficos, topográficos, etc., que após uma análise e avaliação cuidada permitem a definição de zonas diversas, mas homogéneas dentro do mesmo campo, permitindo-lhe assim uma gestão diferenciada e eficiente dos recursos.

O rendimento de um campo é o resultado da interação de vários fatores como o clima, o tipo de solo, a aplicação de fatores de produção e da sua gestão. De acordo com vários autores, a utilização de mapas de rendimento/produtividade de vários anos, permite-nos monitorizar o comportamento dos recursos e as suas interações, dando-nos não só uma visão anual, mas plurianual, uma vez que as produtividades de um campo variam de forma espacial e temporal. Os mesmos, referem que a gestão de áreas diferenciadas, deve ser feita de acordo, não apenas com a variabilidade espacial, mas combinando ambos os tipos de variabilidade, temporal e espacial.

Em Portugal, o milho é a cultura arvense mais importante, e a Quinta da Cholda S.A. é um dos maiores produtores. Dada a importância que a cultura tem em Portugal e a prevalência do clima mediterrânico no país e em particular na região do Ribatejo, que deveras aumenta a dificuldade do desafio que é produzir de forma eficiente e sustentável, é de maior interesse investir em casos de estudo que analisem aquela região e que permitam estudar e ser o ponto

de partida para um maior conhecimento dos fatores agrícolas envolvidos e das interações entre eles. Dito isto, avaliar a variabilidade espacial e temporal dos campos da Quinta da Cholda S.A. e estudar os fatores que a impulsionam, juntamente com o mapeamento das suas características espaciais e temporais, dará ao agricultor/empresário mais informação para otimizar as suas estratégias e aumentar os rendimentos. Para cumprir o que foi mencionado, esta investigação terá os seguintes objetivos:

1) Identificação de campos de produtividade variável e não variável;
2) Estudar que fatores físicos impulsionam a variabilidade temporal dos rendimentos médios;
3) Estudar que fatores físicos impulsionam a variabilidade espacial dos rendimentos médios e em que medida as zonas de variabilidade têm diferentes rendimentos e diferentes respostas aos fatores em estudo;
4) Desenvolvimento de mapas de tendência espacial com base na combinação da variabilidade espacial e temporal em dois dos campos de estudo.

Para estudar o que foi proposto, foram utilizados dados da exploração Quinta da Cholda S.A., localizada na freguesia da Azinhaga, concelho da Golegã, Portugal. A exploração tem 530 ha de milho, regados por pivot central e aspersores de cobertura total. A exploração tem investido muito na agricultura de precisão nos últimos anos, e apesar de ser composta por campos bastante produtivos, apresentam ainda uma significativa variabilidade entre campos e dentro do mesmo campo. Por recolherem informações georreferenciadas há mais de 5 anos, a cedência dos dados referentes ao intervalo de 2015-2019 foi possível.

Neste estudo foram considerados, para a análise entre campos, referente aos três primeiros objetivos, 13 campos, o equivalente a 192 ha.

Para ir de encontro ao primeiro objetivo, calculou-se os valores de rendimentos médios para cada campo ao longo dos 5 anos e, posteriormente, padronizaram-se os resultados de forma a estarem todos na mesma escala, sendo transformados noutros com a média de 0 e desvio padrão de 1 para facilitar comparações. Calcularam-se os desvios padrões e as médias das produtividades para cada campo e classificaram-se, de acordo com as condições pré-definidas, os campos como variáveis e não variáveis com base na variabilidade temporal, sendo o desvio padrão dos valores padronizados a forma objetiva de classificação adotada.

Os resultados foram bastante resolutivos visto que, dos 13 campos, 5 foram considerados variáveis, ou seja, apenas 5 apresentaram grandes níveis de inconsistência no que toca às produtividades ao longo dos 5 anos estudados. De forma a conhecê-los melhor e para perceber possíveis padrões existentes, criaram-se histogramas e boxplots para fazer uma análise em relação a 4 variáveis, a saber: condutividade elétrica aparente do solo (ECa);

índice topográfico de humidade (TWI); altimetria, e; o tipo de solo predominante. Estas variáveis foram escolhidas pelo facto de estarem todas relacionadas com o solo, da sua informação estar disponível, ou ser facilmente calculada, e por constituírem métodos não destrutivos de reconhecimento dos campos.

Para o estudo do segundo e terceiro objetivos, foram criados e analisados 3 modelos lineares múltiplos, para tentar estimar e explicar a variação temporal e espacial entre campos.

O primeiro, juntando as interações entre a condutividade elétrica aparente do solo, o índice topográfico de humidade, a altimetria, juntamente com a adição do variável tipo de solo foi considerado o mais completo e complexo. Como segundo, relativo à variação espacial entre campos, foi considerado um modelo aditivo, constituído pela adição das quatro variáveis escolhidas. Finalmente, foi ainda considerado um terceiro modelo, fruto da redução do segundo modelo.

Dos três modelos ensaiados, o que obteve melhor resultado foi o primeiro.

Para o terceiro objetivo, duas abordagens foram definidas: a primeira baseou-se em estudar quais os fatores que promoviam a existência da variabilidade espacial, medida através da média dos rendimentos. O método foi o mesmo que para o segundo objetivo no que toca à escolha dos três modelos, sendo que as interações entre a condutividade elétrica aparente do solo, o índice topográfico de humidade, a altimetria, juntamente com a adição do variável tipo de solo voltaram a obter os melhores resultados.

Quanto à segunda abordagem, basou-se na previsão da variação espacial, interagindo com as classes de variabilidade temporal definidas no primeiro objetivo. No geral, o objetivo era perceber se era possível distinguir os campos variáveis e não variáveis, perceber se estes campos eram os que tinham maior ou menor produtividade, e se tinham diferentes interações com as quatro variáveis escolhidas, condutividade elétrica aparente do solo, índice topográfico de humidade, altimetria e tipo de solo.

O quarto e último objetivo proposto baseou-se na criação de um mapa, através do método de interpolação, que combinasse a variabilidade temporal e espacial para os dois campos escolhidos, com diferentes características e a fim de visualizar o comportamento do campo ao longo dos 5 anos e de delinear potenciais zonas de intervenção dentro do mesmo campo. Para isso, calculou-se o desvio padrão padronizado e a média dos rendimentos para cada ponto georreferenciado. Os pontos foram classificados de acordo com 4 classes de variabilidade. As referidas 4 classes foram definidas consoante o nível de variação temporal existente e os rendimentos médios ao longo dos 5 anos. Criaram-se ainda mapas de desvio padrão e de rendimentos médios para comparação.

O resultado foram mapas com padrões bastante distintos que poderão ser usados para o delineamento de zonas de estudo e posterior ação, tendo em conta a variabilidade espacial e temporal existente.

Abstract

The evaluation of a field´s spatial and temporal variability, together with the mapping of its characteristics, is an important tool to help decision-making. That said, this research will have the following objectives:

1) Identification of variable and non-variable fields.
2) Study which physical parameters drive time variations in average yields.
3) Study which physical parameters drive spatial variations in average yields and to what extent the variability zones have different yields and different responses to the factors under study.
4) Development of spatial trend maps based on the combination of spatial and temporal variability in two fields of study.

To meet the proposed objectives, data, referring to the years 2015-2019, were used, provided by Quinta da Cholda, Azinhaga, Portugal.

The identification of variable and non-variable fields was made according to the temporal variability measured by the standard deviation. The results showed that only 5 out of 13 fields presented great income inconsistency in the defined period.

For the study of temporal and spatial variation´s driving factors, multiple linear models were created.

Regarding the temporal variation, the best model revealed to be the model integrating the interactions between the apparent soil electrical conductivity, the topographic wetness index, altimetry, together with the addition of the variable soil type, obtaining a model capable of explaining 81% of the existing variability.

As for spatial variation, measured by average yields, two approaches were used. The first resulted in a model with the same structure as the previous one mentioned, obtaining a model capable of explaining 54% of the existing variability and the second resulted in an additive model capable of explaining 96% of the existing variability.

The development of a spatial trend map allowed to create a map, which combines spatial and temporal variability and may be used in the future for the design of study and intervention zones.

Keywords: Precision agriculture; Maize; Temporal variability; Spatial variability; Spatial trend maps.

Resumo

A avaliação da variabilidade espacial e temporal de um campo, juntamente com o mapeamento das suas características, consistem numa importante ferramenta de ajuda tomada à decisão. Dito isto, esta investigação terá os seguintes objetivos:

1) Identificação de campos variáveis e não variáveis.
2) Estudar que fatores físicos impulsionam as variações temporais nos rendimentos médios.
3) Estudar que fatores físicos impulsionam a variação espacial nos rendimentos médios e em que medida as zonas de variabilidade têm diferentes rendimentos e diferentes respostas aos fatores em estudo.
4) Desenvolvimento de mapas de tendência espacial com base na combinação da variabilidade espacial e temporal em dois campos de estudo.

Para ir de encontro aos objetivos propostos, utilizou-se dados, referentes aos anos 2015-2019, fornecidos pela Quinta da Cholda, Azinhaga, Portugal

A identificação dos campos variáveis e não variáveis foi feita de acordo com a variabilidade temporal medida pelo desvio padrão. Os resultados demonstraram que 5 em 13 campos apresentavam grande inconsistência de rendimentos no período definido.

Para o estudo dos fatores impulsionadores da variação temporal e espacial criaram-se modelos lineares múltiplos.

Relativamente à variação temporal, o melhor modelo revelou-se o modelo integrador das interações entre a condutividade elétrica aparente do solo, o índice topográfico de humidade, a altimetria, juntamente com a adição do variável tipo de solo, obtendo-se um modelo capaz de explicar 81% da variabilidade existente.

Quanto à variação espacial, medida pelos rendimentos médios, utilizaram-se duas abordagens. A primeira resultou num modelo com a mesma estrutura que o anterior referido, obtendo-se um modelo capaz de explicar 54% da variabilidade existente e a segunda resultou num modelo aditivo capaz de explicar 96% da variabilidade existente.

O desenvolvimento de um mapa de tendência espacial permitiu criar um mapa, que combina a variabilidade espacial e temporal podendo ser no futuro, utilizado para delineamento de zonas de estudo e intervenção.

Palavras-chave: Agricultura de precisão; Milho; Variabilidade temporal; Variabilidade espacial; Mapas de tendência espacial.

Index

Figure's Index

Table’s Index

List of acronyms

AIC- Akaike information criterion

DEM- Digital Elevation Model

DGPS- Differential Global Position System

ECa- Apparent soil electrical conductivity

GIS- Geographic Information System

GNSS- Global Navigation Satellite System

GPS – Global Position System

IDW- Inverse Distance Weight

TWI- Topographic wetness index

NDVI- Normalized Difference Vegetation Index

RS- Remote Sensing

RTK- Real-time Kinematic

SCA- Specific Catchment Area

VRT-Variable Rate Technology

1. Introduction

Before the 19th century, agriculture was mostly a small-scale subsistence activity based on the intensive use of human and animal labour (Drummond *et al.* 2003). With the advent of the industrial revolution, the primary sector underwent a clear evolution, as a result of access to agricultural machinery and the consequent automation of various processes, allowing the reduction of the old intensive labour force and the increase of competitiveness (O´Brien, 2016).

This period was followed by two major world wars, the second being, as a result of the availability of chemical means, the main reason for the appearance of synthetic products, such as phytopharmaceuticals and particularly fertilizers, the use of which was quickly popularized, and continued on a large scale, with many changes and developments to this day (Ganzel and Reinhardt, 2018). High input use intensities have obvious environmental consequences and constitutes a risk to farmers since cost-effectiveness is affected by spatial and temporal variation in soil-resources and crop demand. Reducing such risks, as well as environmental impact, requires understanding and predicting this variability and to derive subsequent management interventions.

It is for this reason that there has been growing interest in approaches where input use is adapted to local heterogeneity (Coelho and Marques da Silva, 2009). Such localised agronomic practices actually have a long history. Contrary to popular belief, heterogenous treatment of fields was common in many agricultural systems, even before the introduction of modern production factors. In the past centuries, fields were delimited according to several factors, namely, by the types of soil, by the existence of water, and treated, manually according to their characteristics (Stafford, 2000), on the contrary of what happens today, in which most farmers treat large areas as uniform elements, that is, they apply the same amount of inputs to the entire field, ignoring sources of variability such as topography, soil characteristics among others, ending up obtaining heterogeneous productions, both in quantity and quality. More reverently, the introduction of modern yield monitoring, positioning and sensor technologies has given a new impulse to the study and application of localised agronomic management under the name of Precision Agriculture. This concept emerged in the United States in the 1980s with the main purpose of reducing production costs by improving input recommendations, applying them in the right place, at the right time and in the right amount, thus increasing productivity, reducing losses from excess applications, obtaining significant savings at economic levels and thus preserving the environment (Kienzle, 2003). Based on differentiated action among plots, this approach relies on the availability of modern technologies such as the Global Position System (GPS), Remote Sensing (RS), Geographic Information Systems, (GIS), the variable rate technologies (VRT), which made it possible to identify variability, process data and subsequently proceed with a differentiated performance (Andreo, 2013).

Advances in technology related to the collection and processing of yield data, data from satellites, Unmanned Aerial Vehicles or sensors mounted on planes, as well as the development of sensors for soil moisture measurement, apparent electrical conductivity of the soil, measurement of topographic factors and the improvement of weather data forecasting have provided various layers of information, which after careful analysis and evaluation, has allowed the development of new strategies thus giving great support to farmers in decision making at field and within field scales. According to Kharel *et al.*(2019), the classification of fields and the definition of manageable zones within field is usually based on historic yield data, elevation data, soil type maps, apparent electrical conductivity of the soil, and analyses of soil in grid. This design has, as main objective, the development of a number of manageable zones, in which within each zone the variability is reduced, but that between zones is maximized. This will allow a differentiated and efficient management of resources.

The yields of a field or areas within a field are the result of the interaction between various factors such as climate, soil type, production inputs and agricultural management. According to Diker *et al.* (2004), the use of yield maps of several years allows us to monitor the behaviour of resources and their interactions thus giving a more assertive view on the determination of potential management zones and the possibility , not only of studying the behaviour of income annually but multi-yearly, since fields and within fields areas can vary spatially and temporally.

Blackmore (2000), used 6 years of yield data (1993-1998) to develop a spatial trend map in order to establish and characterize temporal and spatial variation patterns by averaging the yield in each grid cell over the years. The data was categorized in stability classes according to coefficient of variation at each point achieved. A different approach was taken by Diker, *et al.* (2004), which used data from two center pivot irrigated maize fields in the years 1997-2000, and which applied a two-state frequency analysis based on the reclassification of the yield grid maps, by assigning the state 0 to yield below the within-year-mean and 1 to yields at and above the mean. Consequently, it was analyzed the frequency that the states occurred annually for all fields and maps were created to delineate yield response zones. Similar conclusions were drawn by both authors, stating that the creation of these spatial and temporal response maps is useful in identifying yield-limiting factors and that they can help in a potential design of zones, which together with other technologies and information, can be areas of research and improvement in a long term.

In more recent studies, in addition to the development of spatial trend maps, the authors decided to compare temporal and spatial variability quantitatively. Maestrini and Basso (2018) used yield monitor datasets from Midwest maize fields to compare the magnitude of temporal and spatial variability. The quantification of spatial variability was made through the standard deviation of the distribution of the yield observed in each yield map, while for temporal variability, they used the standard deviation of the averages across the years. The study showed higher temporal variability than spatial variability, and that these findings can help to look at new forms of management. The same was studied by Kharel *et al.* (2019), who used silage maize data collected by yield monitoring systems, between 2015 and 2017, to come to the same conclusion, that the handling of areas only based on spatial variability may not be useful and sufficient. It should be used combined with temporal variability.

In Portugal, maize is the most important arable crop, and Quinta da Cholda S.A is one of the largest producers. Given the importance that the crop has in Portugal and the special and inconsistent Mediterranean climate that covers the country, thus increasing the difficulty of the challenge that is to produce efficiently and sustainably, it is of greater interest to invest in case studies that analyze this region and that allow studying and being the starting point for a greater knowledge of the agricultural factors involved and interactions between them. That said, evaluating the spatial and temporal fields´ variability and studying the factors that drives it, together with the mapping of spatial and temporal characteristics, will give the farmer more information to optimize agronomic strategies in order to increase yields. To meet what has been mentioned, this research will have the following objectives:

1) Identifying variable and non-variable fields.
2) Study what drives the temporal variations in yields.
3) Study what drives the spatial variation in mean yields and to what extent do variability zones have different yields and different responses to driving factors.
4) Develop within field spatial trend maps based on spatial and temporal variability in two fields of study.

The research will focus on an irrigated maize farm located in Golegã, Portugal, for which yield data have been recorded over the last ten years. The study period will be more precisely from 2015 to 2019. Furthermore, the collection of data such as altimetry or electrical conductivity of the soil were measured as well as the knowledge of type of soils was acquired.

The methodology will emphasize the creation of multiple linear regression to try to explain explain the existing variability and the relationships between the explanatory variables and the response variable.

The development of spatial trend maps will allow the visualization of the consistency of the chosen fields over the five years and the design of potential differentiated management zones.

2. Literature review

2.1 Maize crop

2.1.1 Maize production in the world

In 2018/2019, the United States (32,61%), China (22,91%), Brazil (9,42%), European Union (8,41%) and Argentina (5,41 %) made up the list of the five biggest maize producers in the world (Figure 4).

Compared to the largest maize producers in the world, Portugal's contribution is negligible.

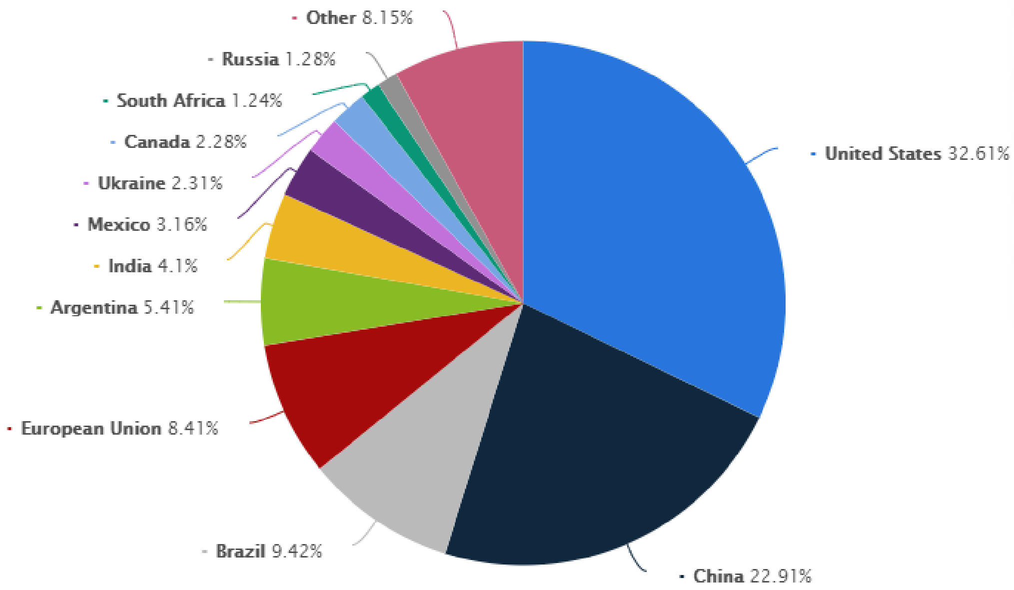


Figure 1- Biggest maize producers in the world.
Source: Statista,2020.

Although productivity continues to increase, Portugal continues to show weaknesses with regard to its production system. The fact that maize is a commodity, means that faces a lot of competition and is low valued, requires producers to become more competitive. In addition, changes are needed in order to face climate changes.

2.1.2 Maize production in Portugal

As mentioned initially, maize is the most important arable crop in Portugal occupying 83356 ha and producing almost 714000 tons in 2018 (INE), (Table 1), however the current scenario is not positive.

The arable crop's, in this case the maize, have suffered major changes in the last decades. In the late 80's, the surface occupied was more than 200 thousand ha. Currently has been drastically reduced.

Portugal is dependent on maize imports. It happens because the consumption keeps increasing and the production decreasing.

Quinta da Cholda is in the centre region which means that belongs to the region that dedicates more area to maize, it is one of the most productive areas in Portugal and it is the region where the place of study is located.

As Torres Coimbra (2019) says it is known for having very appropriated soils to the crop, with water close to the surface and at low cost, which is a benefit.

Table 1-Arable crops' surface and production in Portugal in 2018. Source: Adapted from INE .

	Surface(ha)	Production (t)
Maize	83356	713860
Common wheat	22872	56571
Durum wheat	4153	11178
Rye	15761	16706
Oat	37332	55779
Barley	20526	60238
Triticale	16378	28244
Rice	29350	160794

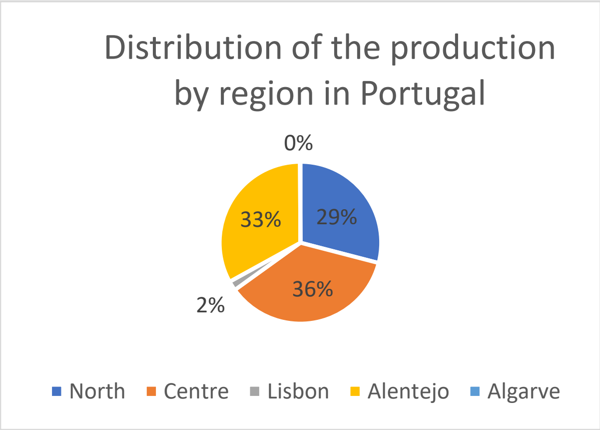


Figure 2- Distribution of the production by region in Portugal in 2018. Source: Adapted from INE.

Although the maize production area is decreasing (Figure 3), due to constant decreases in market prices, productivity has been increasing, as shown in Figure 4. This is due to developments concerning factors of production and its use efficiency, since there is more technical qualification, there are producer's organizations that are organized in order to support farmers, the high health quality of production in Portugal and investment in precision agriculture related technologies.

These precision agriculture tools' adoptions might be the solution to improve productivity, increase production areas and consequently decreasing imports, increasing the degree of self-provisioning and decrease the carbon footprint.

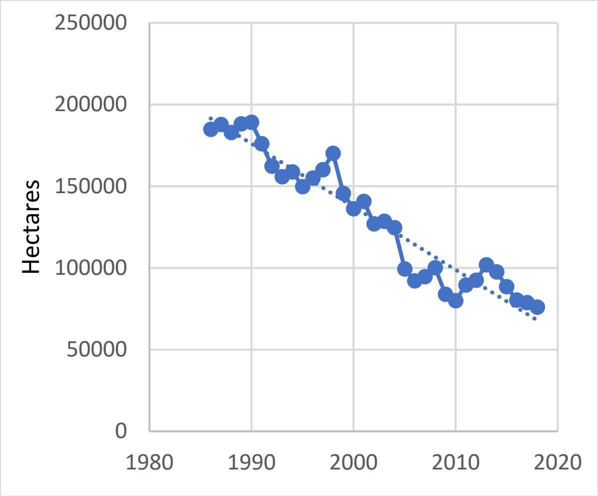


Figure 3-Evolution of irrigated maize surface, in hectares, in Portugal between 1986 and 2018.
Source: Adapted from INE.

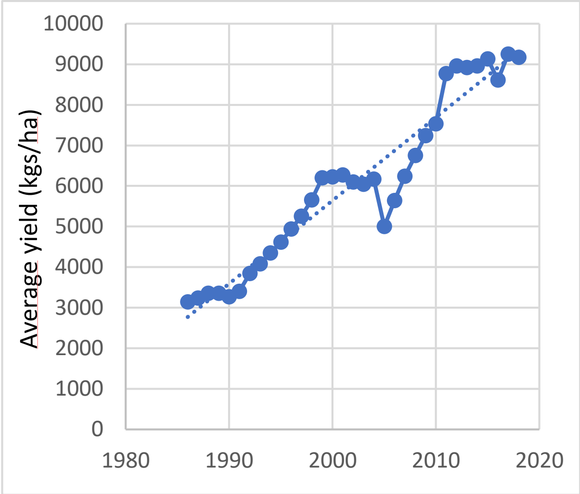


Figure 4-Average yield, in kgs per hectare, between 1986 and 2018.
Source: Adapted from INE.

2.2 Precision agriculture

The concept of precision Agriculture is based on the use of technologies that can be used to improve profitability while reducing the impact of agriculture on the environment (Shannon, *et al.* 2018). As Gebbers and Adamchuk (2010) says, it comprises a set of technologies that combines sensors, information systems, enhanced machinery, and informed management to optimize production by accounting for variability and uncertainties within agricultural systems.

This practice usually appears associated with two main goals: increasing farmers' incomes and a reduction of the environmental impact caused by the agricultural activity. The first of these objectives can, in turn, be achieved by two distinct and complementary ways: reducing production costs and increasing crop productivity and quality (Coelho and Marques da Silva, 2009).

There are five major components of technology used by precision agriculture management practices. They are Geographical Information Systems (GIS), Global Positioning Systems (GPS), sensors, variable rate technology (VRT) and yield monitoring (YM) (Rains and Thomas, 2015).

Although the domain of geo-spatial technologies is important, the focus of the application of precision agriculture is the management of agronomic information and knowledge.

To use it properly , these technologies should be regarded as mere diagnostic tools (remote sensing, mapping) or means of action (VRT, GPS), and agronomic knowledge in this field is the weakest link in the chain (Braga and Pinto, 2015).

To Braga and Pinto (2015), most precision agriculture applications involve large volumes of data that needs to be managed and converted into useful information that can be applied for day-to-day decision-making on farms.

Precision agriculture is distinguished from traditional agriculture mainly by its level of management. While traditional agriculture manages whole fields as single units, precision agriculture makes a customized management for small areas within fields focusing on the benefits at environmental and economic levels (Davis *et al.*1998).

According to Robertson *et al.* (2007) the operations implemented are mostly, but not only, based on the use of vehicle guidance that allows the farmer to reduce overlap in application of agricultural chemicals, reduced traffic associated with tramlining reducing this way, soil compaction and operator fatigue, yield monitoring and variable rate technology (VRT) for application of agricultural chemicals and fertilizers. This can be summarized in reductions in costs in order of 5-8% and a return on investment around 2-3 years (Braga, 2017).

However, precision agriculture might not be for everyone. It requires a substantial investment in technology and knowledge, which could be a barrier for those who want to achieve it (Shannon *et al.* 2018),(Figure 5).

According to Davis *et al.* (1998) before switching from conventional management to precision agriculture it is required a careful reflection.

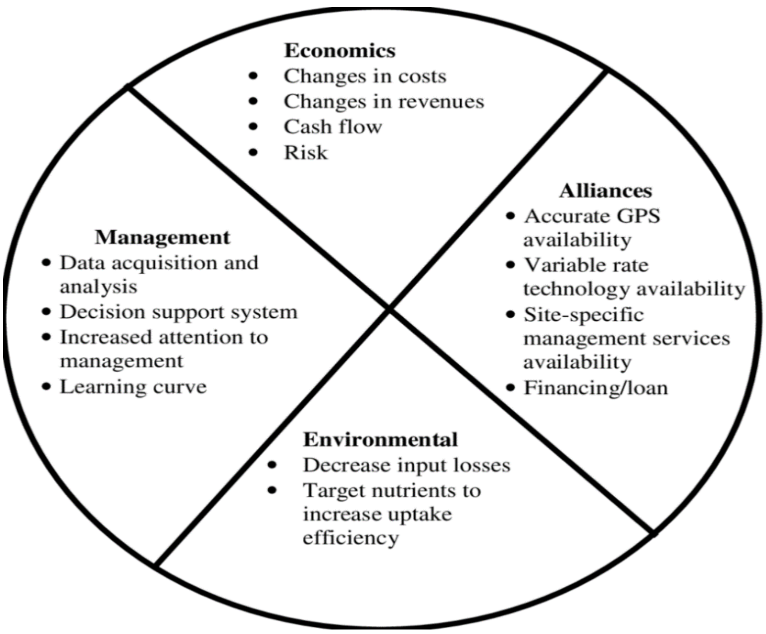


Figure 5 - Issues affecting adoption of precision agriculture management. Source: Shannon et.al (2018).

2.2.1 Precision Agriculture Tools

2.2.1.1 Global Positioning System

GPS (Global Positioning System) is a satellite-based system that indicates to a moving receiver its position, its speed and the time of any point on, or near, the earth's surface in a three-dimensional reference (Tzafestas, 2014).

GPS is part of a group of satellite positioning systems called Global Navigation Satellite Systems (GNSS).

Most of the systems were developed by military forces. According to Stombaugh *et al.* (2018) ,some systems have global coverage and others target specific geographic regions. Some systems may be used to determine base position and others are intended to increase accuracy and / or use in other systems, however they all work by the same principles.

Among them, are Global Positioning System (GPS), GLONASS and Galilean, which are respectively developed by the United States of America, Russia and Europe.

Basically, the system is composed by two main components: a satellite system and a user signal receiver. The first is composed by 24 satellites ,that have an orbital period of 12 hours, and the second on is composed by three components: a radio receiver, a clock and the software needed to perform all calculations to determine the position (Coelho and Marques da Silva, 2009).

For the position to be determined, the GPS receiver requires at least four satellites (Rains and Thomas, 2015).

GPS provides location information at any time however its accuracy of 10 meters might not be enough for some precision agriculture analysis and operations.

Its accuracy is affected by numerous reasons. As explained by Schmidt (2018), the accuracy obtained depends on the proper installation of the system, the degree of technology, the number and location of satellites, atmospheric conditions, among others.

Very high GPS accuracy can be achieved by using a differential GPS (DGPS). To minimize the error it uses two GPS receivers (a rover and a base) that , according to Perez-Ruiz and K.(2012), track same satellites, so that many of the error can be minimized and higher accuracy can be achieved in real time.

When the base station's position is known, the error in estimation the location of the base station can be determined, and the correction is made and transmitted to the field GPS. The disadvantage of this method, is that obtaining two GPS devices, might be too expensive, which lead us to another available differential correction services like RTK.

Real-time kinematic (RTK) positioning is a satellite navigation technique established as the most accurate solution for GNSS applications, producing, according to Perez-Ruiz and K., (2012) typical errors of less than 2 cm,(Figure 6).

Analysing the system, we can say that GPS is the base of almost all precision farming systems since, to determine a spatial variability of a crop or a soil characteristic, is it necessary to know the exact location of the used sampling point (Coelho and Marques da Silva, 2009).

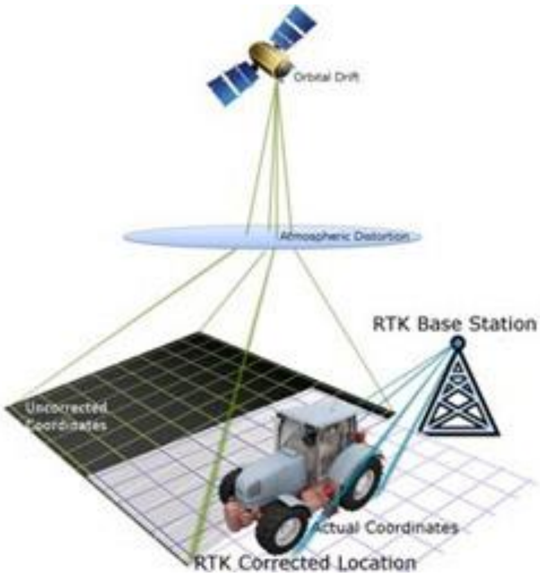


Figure 6- Use of RTK in precision agriculture.
Source: https://precision.agwired.com.

2.2.1.2 Geographic information systems

GIS can be described as a collection of tools that captures, stores, analyses, manages, and presents data that are linked to geographical locations (Bhat and Ahmad, 2011). The data is runned by computer hardware and software systems that uses feature attributes and location data to produce maps (Swain and Singha, 2018).

In agriculture, according to Swain and Singha (2018), GIS is used to store layers of information, such as yields, soil survey maps, remotely sensed data, crop scouting reports and soil nutrient levels. Geographically referenced data can be displayed in the GIS, adding a visual perspective for interpretation.

Furthermore, its use in Precision Agriculture systems is fundamental, since most of the technologies that serve as base to these systems, need georeferenced information such as GPS. It is this combination of technologies that makes it possible to create the complex data structure that underlies most precision agriculture’s systems(Coelho and Marques da Silva, 2009) ,(Figure 7).

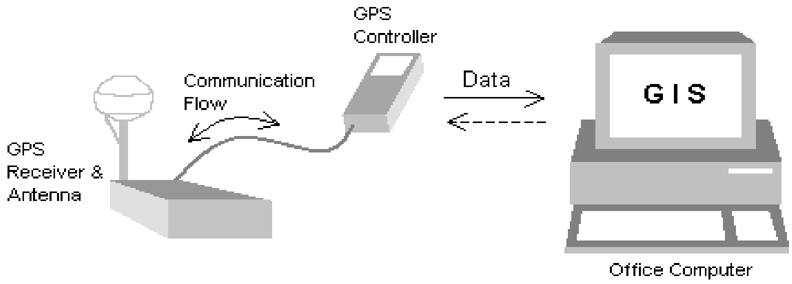


Figure 7- Conceptual view of data-focused integration; Source: https://www.esri.com.

Once analysed, this information is used to understand the relationships between the various elements affecting a crop on a specific site (Meena, 2019).In addition to data storage and display, the GIS can be used to evaluate present and alternative management by combining and manipulating data layers to produce an analysis of management scenarios (Hakkim *et al.* 2016).

2.2.1.3 Yield monitor

Yield Monitoring is considered the most direct method to assess the field production and how it should be better managed. It has the ability of measuring yields as it is harvested (Rains and Thomas, 2009).

Over the years, yield monitoring has been the most used technology in precision agriculture. This happens because due to the fact of being a simple way to understand the within variability marked by the field production.

There are some methods of measuring crop yields, however, the most important is on-the-go yield monitoring because collects georeferenced data on crop yield and characteristics, such as moisture content, while the crop is being harvested.

According to Coelho and Marques da Silva (2009), it is mostly used in arable crops like maize, wheat, soybeans, etc.

Yield monitors are a combination of several components. A yield monitoring system for a combine harvester consists of flow sensor, moisture sensor, ground speed sensor, header position switch, DGPS unit, cutting width sensor, grain loss sensor, grain density sensor (if the flow sensor is a volumetric type) and a computer/display console (Han and Dodd, 1999) (Figure 10).

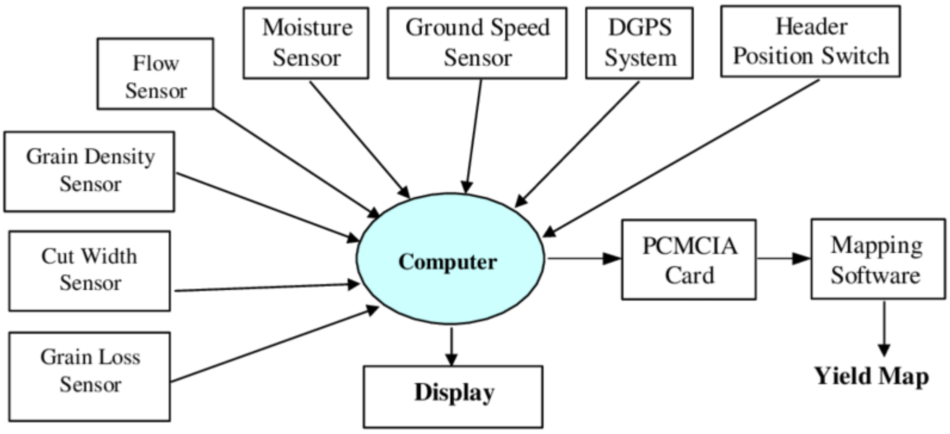


Figure 8- Yield monitor components.
Source: Han and Dodd (1999).

- Flow sensor: Determines grain volume harvested;
- Moisture sensor: Records grain moisture content at the time of harvest so that all yield data can be converted to a standard moisture value;
- Cutting width sensor: The cutting width sensor measures the distance between the crop side and the cutting table side at both sides of the cutting table;
- DGPS: The data is recorded and georeferenced by a Global Positioning System (GPS) receiver and corrected by a differential correction signal receiver that obtains the correction signal from a base station;
- Ground speed sensor: It measures the ground speed of the harvester (sometimes travel speed is measured by GPS radar or ultrasonic sensor);
- Computer/console/monitor: The computer calculates the instantaneous yield and the moisture. The yield calculated at each field location can be displayed on a map using a Geographic Information System (GIS) software package;
- Grain loss sensor: Due to the harvester discharging an amount of grain, it must be accounted by adding to the current yield value. The grain loss sensor measures the amount of grain that the harvester discharges;
- Grain density sensor: It is used to convert yield production expressed in volume, in mass units;
- Header position (on/off) sensor: It is used in turning at row ends and other non-crop areas. The sensor sends a signal to the computer stopping the area and yield calculation, while it raises its head, and reverses the procedure when it is time to harvest again (Han and Dodd, 1999) .

2.2.1.4 Soil electrical conductivity

Measurements of the apparent soil electrical conductivity (ECa) have long been used to characterize the spatial distribution of a variety of soil properties, including salinity, water content, texture, organic matter, bulk density, and cation exchange capacity.

The main utility of the electrical conductivity of the soil, consists on the fact that sandy soils have low conductivity, silts have a medium conductivity and clays have a high conductivity. Consequently, conductivity correlates strongly to soil grain size and texture (Corwin and Scudiero, 2019).

ECa started out as a measure of the amount of salts in the soil in arid zones where an irrigated agriculture was being practised, and in areas having shallow water tables.

The accumulation of salts causes loss of water through evapotranspiration, increasing concentrations of salts in the remaining water. It reduces plant growth and consequently yields. Salinity reduces the osmotic potential making it more difficult for the plant to extract water and may also cause specific-ion toxicity or upset the nutritional balance of plants (Corwin and Lesch, 2005).

In addition, apparent electrical conductivity has been used at field scales to determine some properties like leaching fraction, to help irrigation planning, or identifying drainage patterns, and compaction patterns due to farm machinery.

Seen has a way of studying the spatial variability of several soils' physico-chemical properties, we can say that is a quick, reliable, easy-to-take tool to study the properties that influence crop yield. This, because spatial variation in crops is the result of an interaction of biological, edaphic, anthropogenic, topographic factors (Corwin and Lesch, 2005).

Measurement of ECa can be done by direct contact, usually using at least four electrodes thar are in contact with the soil to inject a current ad measure the voltage that results (Figure 12) or, as BORBA *et al.* (2015) says, without contact with the ground, through electromagnetic induction(Figure 11). It uses a transmitter coil to induce a field into the soil and a receiver coil to measure the response (Zimmermann *et al.* 1972).

After the geo-referenced soil ECa data collection and analysis, in order to better study the soil properties and its variability, soil sampling and profile opening locations can be strategically chosen. It is in this regard that the mapping of the apparent electrical conductivity of the soil, proves to be an essential basic tool for planning the implementation of crops as well as their management and definition of cultural operations such as irrigation management, sowing density, soil mobilization, among others.

Figure 9- Electromagnetic induction ECa measuring method.
Source: Own source

Figure 10-Soil direct contact ECa measuring method.
Source: *https://www.agriexpo.online*.

2.3 Spatial interpolation

Spatial interpolation is a method used to estimate the value for a query point (or a raster cell) with an unknown value from a set of known sample point values that are distributed across an area (Leah and Goulden, 2020).

According to (Karydas *et al.* 2009), interpolation is based on the assumption that, values at points close together in space tend to have similar characteristics than points further apart, in other words, they are spatially correlated.

This procedure is widely used in precision agriculture, either to create yield maps or to create altimetry maps among others (Torres Coimbra, 2019).

Among spatial interpolation methods, the main ones are:

- Kriging;
- Inverse Distance Weight (IDW).

2.3.1 Kriging

According to (Grego C. R., Oliveira R. P., 2014) kriging is a geostatistical technique used to estimate property values for locations where this property was not measured. For this method to be used it is necessary that there is spatial dependence (autocorrelation), defined by the semivariogram.

A semivariogram summarises the way that properties vary from place to place and allows to understand the relationships between observations separated by different lag distances (Oliver, 2013).

The usual method to estimate it from data is with the equation:

$$y(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [X(i) - X(i+h)]^2 \quad \text{(Eq.1)}$$

Where:

- h- is the lag (or spatial sampling interval) and is a vector in both distance and direction because of the existence of anisotropy;
- $X(i)$- is a sample value at location ;
- $X(i+h)$- is another sample value separated from $X(i)$ with the sample interval h;
- $N(h)$ is the total number of pairs of $X(i)$ and $X(i+h)$.

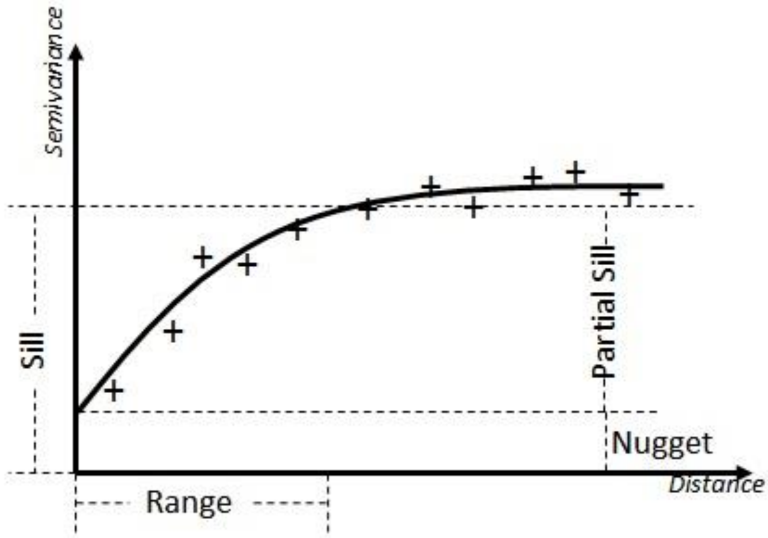


Figure 11- Illustration of a regularized semivariogram.
Source: https://www.polyu.edu.hk.

Figure 11 represents a semivariogram in which:

- Sill – is the distance within which the samples are spatially correlated;
- Nugget – Refers to the variability in the field data that cannot be explained by distance between the observations, in other words, represents the amount of non-spatially autocorrelated error;
- Range – Represents the distance limit beyond which the data are no longer correlated. Up to this value it is considered that the points closest to each other are probably more similar than the points further away, and from this value onwards, it is considered that there is no longer spatial dependence between the samples, that is, they have no influence on each other.

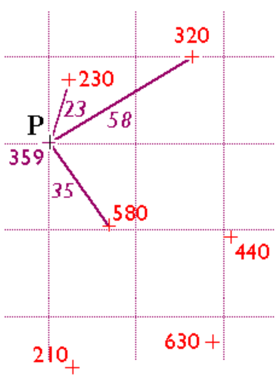
2.3.2 Inverse Distance Weight

Inverse Distance Weighted (IDW) assumes that the nearer a sample point is to the cell whose value is to be estimated, the more closely the cell's value will resemble the sample point's value (Handayani *et al.* 2019), i.e., the measured values closest to the prediction location have more influence on the predicted value than those farther away. In order to predict a value for any unmeasured location, IDW uses the points of the neighbourhood by pondering them as a function of the inverse of the distance(Babak and Deutsch, 2009), (Figure 12).

This method is very dependent on the amount and dispersion of samples in the portion.

A power parameter can also be added. It allows to give more or less influence to known points based on the distance from the output point. The higher the power value, the more influence is given to the closer points and vice versa.

The method used in this work was Kriging because it is considered more robust comparing to IDW, and because it derives from a statistical model capable of estimating prediction errors and saying how correlated variables are at varying distances, which leads to higher accuracy.



$$z_P = \frac{\sum_{i=1}^{n}\left(\frac{z_i}{d_i}\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i}\right)}$$

$$= \frac{\frac{230}{23} + \frac{320}{58} + \frac{580}{35}}{\frac{1}{23} + \frac{1}{58} + \frac{1}{35}}$$

Figure 12-Simplified IDW procedure explanation. Source: https://www.e-education.psu.edu.

3. Materials and Methods

3.1 Site characterization

The plots under study are located in Quinta da Cholda, belonging to the parish of Azinhaga, in the municipality of Golegã, district of Santarém, Portugal (39º N, 8º O). At the moment, this company operates about 530ha where maize for grain is produced. It is located in an area characterized by fertile soils, low slope fields and abundant water since they are next to the Tagus River, the main river Portuguese.

All fields explored are irrigated by center pivots and solid set systems.

Figure 13-Farm location on Portugal's map.
Source:Adapted from Google Earth.

Figure 14-Illustration of fields under study.
Source:Own source.

The plots are in two very close regions even if separated by about 40 kms. On the left side of Figure 14, we can observe the fields that are in the parish of Valada, while on the right side, we observe those in the parish of Azinhaga.

The company operates 19 fields in total, which are subdivided into sub fields because they have different forms and consequently have different irrigation systems.

For these reasons, for this study, the subfields were considered as different fields.

The fields are formed by consisting only clayey and sandy soils, and their areas can vary between 0.5 and 72.00 hectares.

To this study, only 13 fields were considered, equivalent to 192 hectares.The description of the fields is presented in Appendix A.

3.1.1 Climate characterization

For the realization of the climatic characterization of the fields and the region in question, the climate data for 30 years (1984-2014) were taken into account. It was considered a period of 30 years, according to the World Meteorological Organization (WMO) and according to the (*IPMA - Clima Normais*, 2020), for corresponding to the number of years long enough to admit that the climatic value represents the predominant value of that element in the place considered.

Due to the lack of data available by the weather station closest to the fields, the Station of Santarém about 20 km away, it was decided to complete the missing data using the station of Abrantes, 30 km away. This method was adopted since, comparing the available values between stations, it was concluded that they were quite similar. These data were consulted through the portal of the Sistema Nacional de Informação de Recursos hídricos **(SNIRH)** and the weather station of Abrantes, (**MeteoAbrantes**).

Regarding the data from the study years, 2015-2019, were obtained by the meteorological station of the farm, since it has been operating since 2015.By Koppen classification (1928), the region is marked by a mediterranean temperate climate, characterized by being rainy and moderately warm, with intense rains in winter (type Cs). Because it has an average temperature of the coldest month of the year, below 18 °C and above -3 °C and because the average temperature of the hottest month is higher than 22°C, it is included in the sub-type Csa, being thus classified as a clime tempered with rainy winter and dry and hot summer.

Gaussen´s ombrothermic diagrams are used to summarize trends in temperature and precipitation during a defined period. They allow to establish the relationship between temperature and precipitation and to determine the dry period in the study zone, which corresponds to the spot in where the average monthly temperature line is twice of the precipitation line.

As can be seen in Figure 15, the dry season corresponds to the months from June to July, however, from 2015 to 2019, the trend has been for an increase in temperatures during the month of September, thus leading to it being included in the dry period as we can see in Figure 16.

The study period has very similar average monthly temperature values compared to the period 1985-2014. It is noted, when comparing both curves, that the same hasn´t occurred regarding the precipitation of the years mentioned.

Of all the months of the year, highlight the months of sowing, that is, between April and May.

When analysing the diagrams, Figure 15 and 16, we conclude that the average monthly precipitation of the study period was much higher, contrary to the most ancestral period,(Figure 15). Excess precipitation may be a factor that makes it impossible for the soil to reach the appropriate state for sowing operations.

During the dry season, precipitation is not a problem since maize production is done using irrigation. Comparing the study period, there was on average a lower precipitation compared to the average of 30 years, which implied the application of water in the form of irrigation according to the needs of the crop.

September and October usually coincide with the harvest season. In these months low precipitation is not a problem. It is a producer-friendly factor that reduces moisture in the grain and avoids the use of the dryer allowing it to save a lot economically. Compared to the 30-year average diagram, the average precipitation between 2015 and 2019 was much lower and was therefore an advantage for the producer.

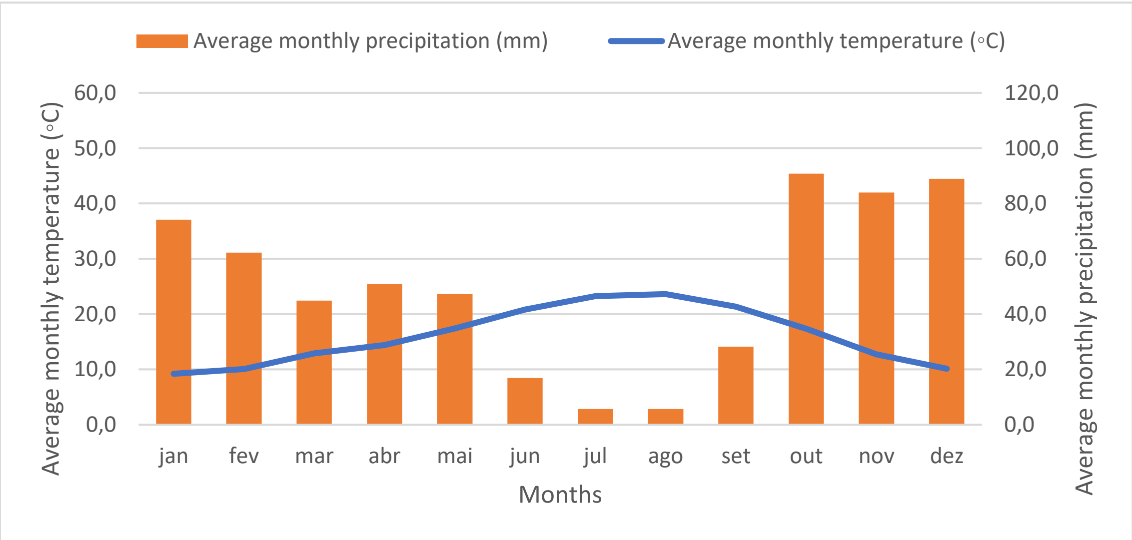


Figure 15-Gaussen's ombrothermic diagram from 1985-2014.

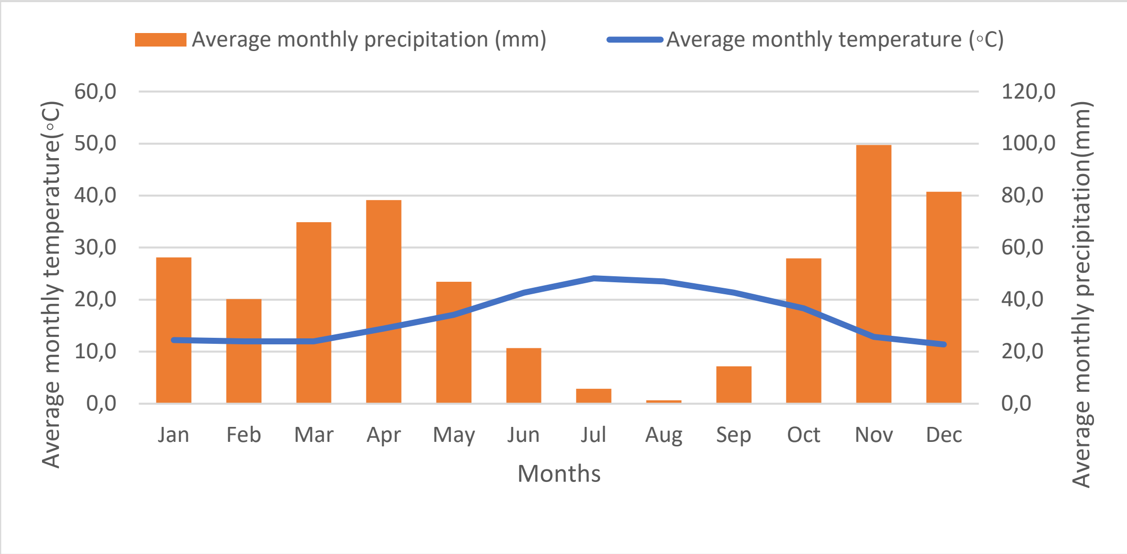


Figure 16-Gaussen's ombrothermic diagram from 2015-2019.

3.1.2 Edaphic Characterization

According to the soil chart of the SROA (1972), elaborated according to the soil classification of Portugal (Cardoso, 1965), the soils that predominate in the fields of the farm are incipient soils.

In addition to these, halomorphic soils, hydromorphic soils and poorly unsaturated clay soils are present on a smaller scale.

According to the soil classification of Portugal, **incipient soils** are divided into suborders, with the predominant in this study corresponding to alluviosols.

<u>Alluviosols</u> are characterized by being very little evolved soils, in which soil formation factors did not act long enough to develop well differentiated pedogenetic horizons. What can happen in many cases is a certain accumulation of organic matter on the surface, which is never very large, since mineralization takes place quickly due to the good aeration of the top layer.

Alluviosols are divided into two subgroups:

<u>Modern alluviosols</u>- which are characterized by receiving, in general from time to time, additions of alluvial sediments.

<u>Ancient alluviosols</u>- represent soils that no longer receive, as a rule, additions of alluvial sediments. They usually constitute river terraces and usually present the water table at a greater depth compared to modern alluviosols.

According to the FAO classification (FAO, 2014),alluviosols correspond to the <u>Fluvisols</u> in the World Reference Base for Soil Resources (WRB) classification.

Halomorphic soils- These are soils that have a large amount of salts and/or relatively high sodium content of exchange in the absorption complex. They are formed by the process called salinization, with a superficial horizon, where soluble salts of sodium, calcium, magnesium, and others accumulate. Its pH rarely rises above 8.5

In this case, the halomorphic soils correspond to the Solonchaks in the WRB classification.

Hydromorphic soils- are soils subject to temporary or permanent soaking. Water causes intense reduction phenomena in all or part of its profile, especially iron oxides, which are quite soluble and move a lot throughout its profile. Its formation is always related to flat or concave reliefs, often appearing in almost all alluvial formations.

Hydromorphic soils correspond to Planosols, in the WRB classification.

Poorly unsaturated clay soils- are soils characterized by being little evolved, of ABtC profile, in which the saturation degree of the Bt(clayey) horizon is greater than 35% and does not decrease with depth or in the underlying horizons.

These poorly unsaturated clay soils, according to the FAO classification are considered Luvisols.

3.2 Data description

3.2.1 Topographic wetness index- TWI

The topographic wetness index is a relative measure of the long-term availability of soil moisture at a given location in the landscape (Kopecký and Čížková, 2010) and (Beven and Kirkby,1979). This index characterizes the spatial distribution of the saturated surface zones. It demonstrates the relief effects on the location and extension of the accumulation areas, which are the most propitious to reach the state of saturation (11º Sinageo - Utilização do Índice Topográfico de Umidade como suporte ao planejamento e gestão ambiental de Unidades de Conservação de Uso Sustentável, 2016).

It is defined as follows:

$$TWI = \ln\left(\frac{SCA}{Tan\ \alpha}\right) \quad \text{(Eq.2)}$$

Where SCA is the specific catchment area and α is the slope angle.

SCA is a parameter of the tendency to receive water, while the local slope and the draining contour length, implicit in the SCA, describe the tendency to evacuate water (Figure 17).

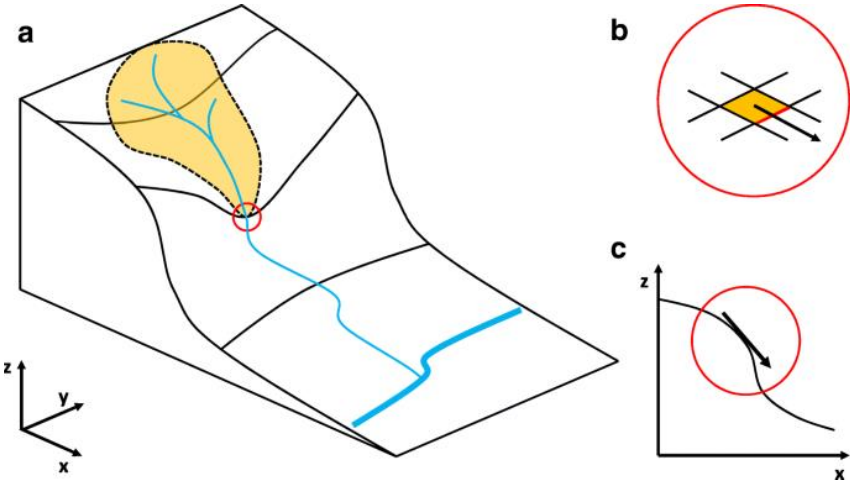


Figure 17- TWI scheme.
Source: (Mattivi et al., 2019).

Figure 17 represents very well the study variables of the topographic wetness index, where **a** is the Flow accumulation area, **b** is the Flow direction and the corresponding flow width for a DEM cell and **c** is the Tangent of the slope angle α.

The easiest way to interpret it is thinking that TWI is one the most important factors that indicates the potential of runoff generation. In other words, the high values of TWI means having greater propensity of reaching the saturation´s state and vice versa.

The main source of data to conduct this kind of study is represented by the Digital Elevation Model (DEM), from which it is possible to obtain several topographic indexes.

According to (*Digital Elevation Model | National Land Survey of Finland*, indate),the digital elevation model is based on a numerical representation of the earth's surface that contains actual height points representing topography. It also represents a method for calculating elevations between height points.

The elevation of the terrain´s data was obtained through the tractors, with built-in GPS RTK, with an accuracy of 2.00 cm horizontally and 4.00 cm vertically, during the sowing operations. Collecting elevation data during slower and smoother operations such as sowing allows obtaining higher-quality data (Torres Coimbra, 2019).

The **QGIS 3.10** software was used to create a grid- based DEM raster with a spatial resolution of 2.00m by kriging. Subsequently, the algorithm required for the calculation of the SCA and slope angle α was defined individually. Finally, the calculation function, Raster Calculator, was used to calculate the TWI based on the Eq.2, presented earlier.

In so that the data from this study were all on the same scale and for easy viewing, the TWI data´s spatial resolution was reduced to 7.00 m by kriging.

3.2.2 Altimetry measurements

Method of measuring the elevation of surface points.

Altimetry's data was obtained through the tractors, with built-in GPS RTK, with an accuracy of 2.00 cm horizontally and 4.00 cm vertically, during smaller spacing operations, notably mobilizations with power harrows or sowing operations.

After data collection, by kriging, the data´s initial resolution 2.00m, was converted into a 7.00m by 7.00m grid in the **QGIS 3.10 's** software.

Figure 18- Tractor with a built-in GPS RTK and a power harrow. Source: Adapted from Milho Amarelo.

3.2.3 ECa measurements

The soil apparent electrical conductivity (ECa) data collection was made gradually over the years, being the first reading made in 2015 and the last in 2019.

Measurements were made through the electromagnetic induction measuring method, by service providers, at 1.00m depth.

Its treatment was based, as well as on the other study variables, on the use of the kriging method, for the calculation of missing values and for disposing of the values in a grid with the final spatial resolution of 7.00 by 7.00 m for further analysis.

3.2.4 Maize yield measurements treatment

The result of monitoring the productivity for the five years of harvest, is a file with georeferenced points that are associated with several variables such as the moisture content of the grain, the speed of displacement of the harvester, the amount of grain harvested, the calculated productivity, among others. However, it is necessary to filter the data in order to eliminate unreliable points, thus avoiding errors that can lead to wrong decision making.

To filter and delete points considered untrusted, the following conditions have been applied:

- Points that are not in the plot and that were accounted for due to GNSS signal errors;
- Points, whose working width does not correspond to reality;
- The Flow below 1.9 kg/s. Below this value the sensor tends to fail leading to lack of confidence in values;
- Removal of bedside turns and lines in which the harvester traverses the field instead of following the normal direction of cutting maize;
- The Velocity is below 1.6 Km/h or above 10 Km/h;
- Points where yields exceed maximum biological limits, in this case 26 tons;
- The Moisture is three times more or less than the standard deviation.

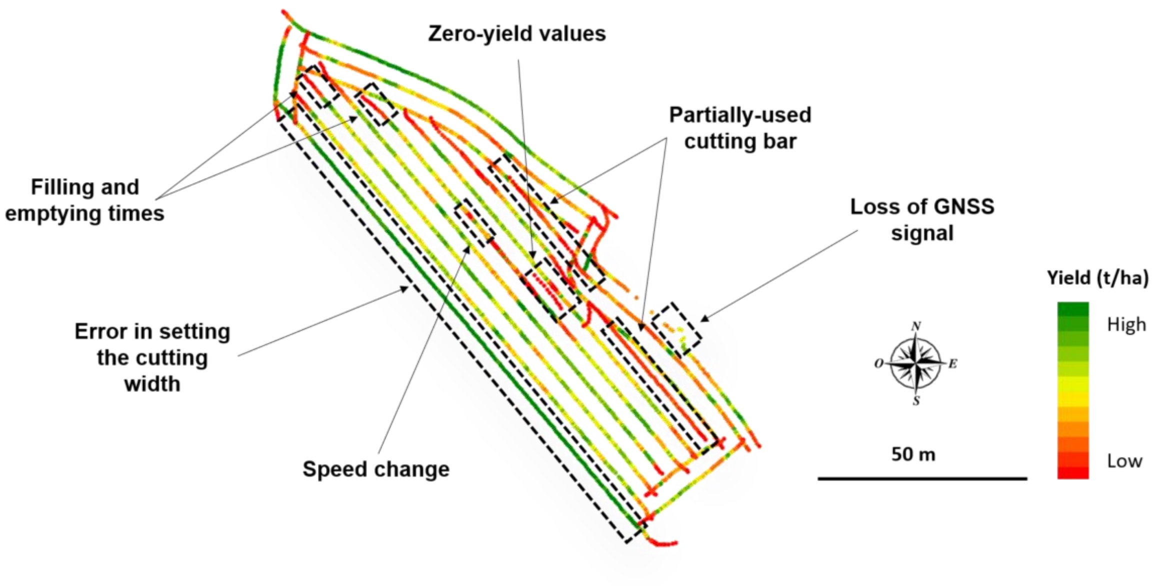


Figure 19-Visual example of existing errors in monitoring yield data. Source: https://www.aspexit.com/.

After filtering the points, the correct points were interpolated by kriging to generate grid-based raster with a spatial resolution of 2.00 by 2.00 m and subsequently reduced to 7.00 by 7.00 for easy viewing and so that the data was all at the same resolution.

3.3 Methods

3.3.1 Field and within field variability classification

The classification was made taking into account the level of yield obtained in the fields, in five years, in order to understand the level of variability over the time considered.

A data processing was previously done according to the following steps:

1) Average yield calculation per field and per year since the data was presented in georeferenced values. The average was calculated through the equation:

$$\bar{X} = \frac{\sum Xn}{n} \qquad \text{(Eq.3)}$$

Where $\bar{X}$ is the year average yield, Xn is the yield value from each point of the field and *n*, the number of georeferenced points.

2) The values calculated in step 1) were centered and scaled. This process was done through the scale() function of the R program, which centers and scales the columns and their values, so that they were all on the same scale, in other words, the values were standardized. This was done because different years have different values, and consequently, the values have different variations. To standardize yields, the following formula was followed:

$$Z = \frac{X - \mu}{\sigma} \qquad \text{(Eq.4)}$$

Where Z are the standardized yield values, X is the average yield over the points calculated in step 1), μ is the overall average yield per year and σ, the overall standard deviation per year.

3) Calculate of standard deviation and averages of scaled yield.

For the standard deviation (σ), the following formula was followed:

$$\sigma = \frac{\sqrt{\sum(X - \bar{X})^2}}{n} \qquad \text{(Eq.5)}$$

where σ is the standard deviation, X is the standardized yield for one year of the field, $\bar{X}$ is the five years´ yield average and n is the number of years.

a) Generating temporal variability classes for the between field analysis.

Temporal variability was estimated by calculating the standard deviation across the years of the standardized yield. We standardized the yields for every field-year according to the method presented in the previous topic.

After these calculations, the variability of the fields was defined based on the following criteria:

1. Variable if σ > standard deviation median;
2. Non-variable if σ ≤ standard deviation median.

b) Generating variability classes for the within field analysis according to temporal and spatial variability.

The within field analysis considered two variables: the standard deviation across the years of the standardized yield, corresponding to temporal variability, and the standardized average yields, corresponding to spatial variability. The calculation method was the same previous mentioned for between fields analysis, excepted average yield calculation per field, in step 1), because georeferenced information was required.

For this part of the study, the classes were defined based on the following criteria:

1. Variable and high yield if σ > standard deviation median and μ ≥ average yield median;

2. Variable and low yield if σ > standard deviation median and μ < average yield median;

3. Non-variable and high yield if σ ≤ standard deviation median and μ ≥ average yield median;

4. Non-variable and low yield if σ ≤ standard deviation median and μ < average yield median.

3.3.2 Modelling

3.3.2.1 Multiple linear regression models

Multiple linear regression models, in general, are a statistical technique that uses two or more explanatory variables to predict the outcome of a response variable, in this case, TWI, ECa, Altimetry, Soil type and yield standard deviation and yield averages respectively(Kenton, 2020).

It is often used to know how strong the relationship is between two or more independent variables and one dependent variable or to know the value of the dependent variable at a certain value of the independent variables.

Formula and calculation of multiple linear regression:

, where, for *i=n* observations:

$$Yi = \beta_0 + \beta_{1xi1} + \beta_{2xi2} + \ldots + \beta_{pxip} + \epsilon \qquad \text{(Eq.6)}$$

Yi-dependent variable;

xi -explanatory variables;

β_0-intercept (constant term)- Average value of *Yi* when xi are set equal to zero ;

β_p-slope coefficients for each explanatory variable- The coefficients β_1 and β_2 are also called partial regression coefficients. It measures the change in the mean value of *Yi*, per unit change in $x1$ holding the value of $x2$ constant;

ϵ-the model's error term (also known as the residuals).

3.3.2.2 Explanatory variables

For this analysis we will use, from the available data, the types of soils existing in these areas, measurements of electrical conductivity (ECa), at the topographic level, the altimetry and the Topographic wetness index (TWI).

What these four factors have in common is the fact that they all related to the soil, mainly and more specifically with its structure. In addition, the variables are obtained through non-destructive methods, except for the type of soil whose methods for its knowledge is irrelevant with regard to soil mobilizations. Another factor is the fact that they remain constant over several years and that they are relatively easy and inexpensive in the long term.

Soil is one of the main elements of agriculture, and its knowledge allows us to make a differentiated and sustainable management. By knowing the type of soil, we know the ecological basis where plants will develop.

The knowledge of its physical and chemical properties gives us information about its structure, its nutritional availability, the capacity of water and mineral retention, among others.

One way to know the properties of each type of soil, is through the apparent electrical conductivity of the soil (ECa). This measurement serves as a fast, easy and reliable means of setting spatial patterns. According to Rabello, (2009), these patterns are fruit of the variation of electrical conductance, influenced by salinity, moisture, texture and resistivity of the soil, all factors contribute directly to the success of the crop.

The choice of the topographic wetness index (TWI) was made because maize is an irrigated crop and the fields being considered quite flat, that is, with very low slopes, which consequently affects water drainage and potentially allows soil saturation. Since the studied areas are of high dimension, the possibility of existing fields and/or zones within the same field where water saturation can occur is not to be ruled out.

These zones may reflect various factors such as compaction, due to the transitability of agricultural vehicles, or even technical errors, whether in pivots and solid set systems, or in moisture measurement sensors. Areas where water saturation may exist can lead to root asphyxia and possibly to the death of plants, which is synonymous of yield losses. Since this is an index related to water availability or lack of water in the long period, it makes perfect sense to relate it to soil properties.

Regards altimetry, it is a topographic component that represents the vertical distance between points and from which studies on surface drainage can be carried out. It is also related directly to the TWI´s calculation.

Its choice was simply to examine whether the small differences that are expected, since the fields are all considered flat and well drained, have any effect on incomes.

The junction and/or interaction between these variables was thus considered a possible added value to study the temporal and spatial yield variability between fields.

Other variables such as the cultivars used, irrigation supplies, the amount of inputs applied, among others, were not included in order to simplify the study.

3.3.2.3 Selection of models

The models for the dependent variables yield standard deviation and average yield were both based on 4 variables, TWI, ECa, Altimetry and Soil type. For each dependent variable, three models were created.

<u>The first model</u>, to predict temporal variability and for the first spatial variability analysis´approach, was composed by the interaction between TWI, ECa and Altimetry, plus

the addition of the categorical variable Soil type. It was defined this way to be seen as the most complex and complete model. The full model formula is presented as:

Temporal analysis:

- **Yield standard deviation= ECa*Altimetry*TWI +Soil Type**

Spatial analysis:

- **First approach: Average yield= ECa*Altimetry*TWI +Soil Type**

The second model, to predict temporal variability and for the first spatial variability analysis´approach was an addictive model, that is, composed by the addiction of all variables without interactions. The full model formula is presented as:

Temporal analysis:

- **Yield standard deviation= ECa+Altimetry +Soil Type+TWI**

Spatial analysis:

- **First approach: Average yield= ECa+Altimetry +Soil Type+TWI**

The third model, to predict temporal variability and for the first spatial variability analysis ‘approach, was defined as the result of model reduction.

Regarding the second approach of spatial analysis, to meet the third study objective in the thesis, and to meet the reliability of the prediction, only one model was defined, an addictive model interacting to the pre-defined between field variability classes. The full model formula is presented as:

Spatial analysis:

- **Second approach: Average yield= (ECa+Altimetry +Soil Type+TWI)*Variability classes**

An additional model, with the reduced model structure in the first approach interacting with variability classes, was included as term of comparison.

It was used the stepwise regression backward method, by the step() function of the program R to reduce the model if possible and if it made sense in order to create the most fitted model.

Summary and ANOVA analysis were performed to all models to confirm and see the variables´ contribution to the results.

3.3.2.4 Model analysis

To analyse the significance of the explanatory variables and their interactions in the models, ANOVA variance analyses were performed.

Analysis of variance (ANOVA)

Variance analysis (ANOVA) is a statistical analysis tool that allows estimating how the quantitative dependent variable changes according to the levels of one or more categorical independent variables ((Bevans, 2020). Besides being the appropriate procedure for testing the equality of several means it helps estimating and testing hypotheses about the treatment effect parameters.

To obtain the most appropriate model to answer the questions in question, multiple regression models were created, and constant variance analyses (ANOVA) were performed to arrive at a model that was statistically significant.

The F test was used to analyse the significance of regression coefficients (β_p) together and individually. To test its significance through the analysis of variance, the hypotheses were tested:

- Ho: β_1=β_2=…=$\beta_p = 0$;
- H1: At least one β_p is not equal to zero.

The null hypothesis (Ho) implies that all the regression coefficients are equal to zero, that is, that the model, only with interception (β_0), better explains the variance, while the alternative hypothesis(H1) tells us that at least one regression coefficient other than zero explains the variance better than the model only with the interception.

The results analysis was performed through the p-value for F-statistics, presented as "Pr(>F)" in the ANOVA tables (Appendix B).

For this, a significance level of α=0.05 was previously established, which means that the probability of rejecting the null hypothesis when it is true is 5%.

Therefore, the p-value for each variable tests the null hypothesis, in which the coefficient is equal to zero, that is, without effect. A low p-value indicates that we can reject the null hypothesis, in other words, a variable with a p-value <0.05 is likely to make a significant contribution to the model. On the other hand, a variable with a p-value higher than the significance level, suggests that changes in the explanatory variable are not associated with changes in the response variable.

To perform the ANOVA analyses, the summary() and anova() functions of the R program were used.

3.3.2.5 Model evaluation

The models were evaluated according to two methods, using the Akaike information criterion, AIC() function of the R. stats package, and the r-squared criterion from the analysis of the summary table from the summary() function.

Coefficient of correlation (r) and Coefficient of determination, R-Squared (R^2)

Pearson's correlation coefficient and coefficient of determination indicate the degree of collinearity, that is, how accurately a straight line can be drawn capable of describing the relationship between the points described by the observed and estimated values (Yuemei *et al.* 2008). The correlation coefficient is used as an indicator of linearity between the observed and estimated values, and this value is contained in a range between -1 and 1. If r has a null value (equal to 0), it indicates that there is no linearity, and that all values close to it have low linearity. If r is -1 or 1, the linearity of the system is perfect, which is negative or positive, respectively (Yuemei *et al.* 2008).

According to (Hahs-Vaughn *et al.* 2020), the coefficient of determination R^2, is the square of the sample correlation coefficient between the predictors, independent variables), and response variables, the dependent ones. It measures the model´s quality and tell us the fraction of the variance of the dependent variable that is explained by the regression model. The variance that is not explained by the model is explained by other factors (i.e., unknown variables or sampling variability), (Long and Teetor, 2019).

Simply put, the higher R^2 is, the more, of the total variation, is explained by the model´s predictors, being this value contained in an interval between 0 and 1, and the perfect regression line for the model would have a value of $R^2 = 1$.

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad \text{(Eq.7)}$$

Where:

R^2- coefficient of determination;

yi - value measured by the reference method for the i-th sample and for the variable under study;

$\hat{y}$-value estimated by the model for the i-th sample and for the variable under study;

$\bar{y}$- mean of the values observed in the reference method.

Adjusted R-Squared ($\mathbf{Adj.}\,\boldsymbol{R^2}$)

Adjusted R-Squared is an alternative measure to simple R-square.

What happens is that the R^2 improves whenever predictors are added, even if they are not related to the response variable, and thus cannot help identifying the predictors that should be included or those that should be excluded.

Thus, the adjusted r-squared ($\mathrm{Adj.}\,R^2$) consists of a modified version of R^2 that was adjusted to the number of predictors of the model and that penalizes the inclusion of insignificant independent variables. It improves only if adding a new predictor variable, improves the model more than expected and worsens when a predictor improves a model less than expected.

The adjustment is done by the equation:

$$Adj.R^2 = 1 - ((1 - R^2)\frac{N-1}{N-k-1} \qquad \text{(Eq.8)}$$

Where:
k- number of predictors excluding the intercept;
N-total sample size.

To access the r-squared and adjusted R-squared results, the summary() function of the R program was used.

Akaike´s information criterion (AIC)

The Akaike information criterion (AIC) is a mathematical method used to evaluate how well a model fits the data it was generated from (Bevans, 2020). It compares the quality of a set of statistical models to each other.

Akaike’s Information Criterion´s basic formula is defined as:

$$AIC = -2(log - likelihood) + 2K \quad \text{(Eq.9)}$$

Where:

K- number of model parameters (the number of variables in the model plus the intercept);

Log-likelihood – is a measure of model fit. The higher the number, the better the fit. This is usually obtained from statistical output.

A good model is the one that has minimum AIC among all the other models.

To calculate the AIC´s values, the AIC() function from the R program was used.

4. Results

4.1 Variable´s analysis by variability class

To analyse possible trends regarding fields and associated variables, histograms and boxplots were developed. A histogram is a graphic version of a frequency distribution while a boxplot is a graphic that summarizes a great deal of information about the distribution of data around the median. Horizontal lines show the median of the data set, the bottom and top of the box show the 25th and 75th percentiles, that is, the location of the middle 50% of the data; the vertical lines are called the “whiskers”. The upper and lower whiskers either presents, respectively, the maximum and minimum value, or outliers.

The results were as follows:

As can be seen in Figure 20, most fields are considered non-variable over the five years of production. From the 13 fields studied, 8 were classified as non-variable which means they were the more consistent, presenting least variations comparing with the remaining 5 that were considered variables presenting greater variations during the study time.

With this frequency table we were able to have a more general picture of the fields´ behaviour over the 5 years.

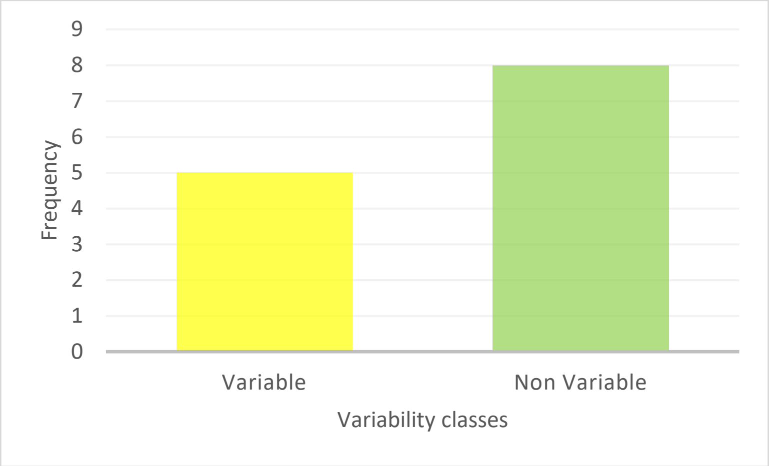


Figure 20-Fields frequency per variability class.

Regarding to altimetry, the average altimetry´s values vary between 66.6 meters and 69.95. Due to the small difference between the maximum mean value and the minimum mean value (about 3 meters), it was expected that there were no large oscillations between the characteristic values of each variability class, however the variable class is characterized by a shorter range of values. Since the fields are considered all flat, and as was seen earlier, there are only 5 of 13 fields that have been classified as variable, it makes sense that the values vary relatively little.

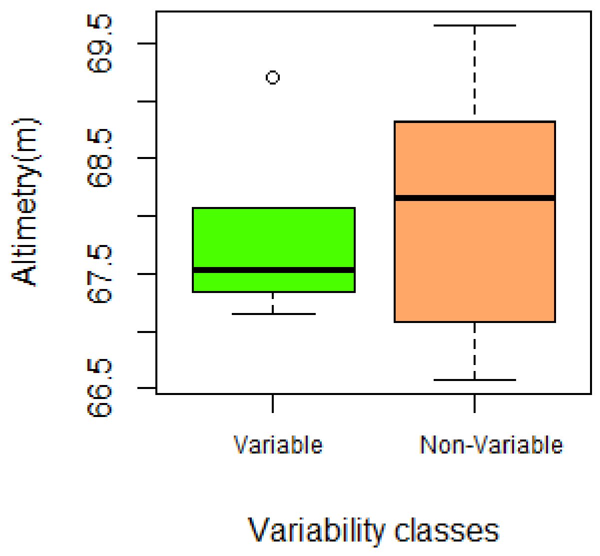


Figure 21-Altimetry's values oscillations per variability class.

The average topographic wetness index (TWI) value´s vary from a minimum of 7.68 to 8.58, as we can see in Figure 22.

The class that covers a wider range of values is undoubtedly the variable class with values around 7.9 and 8.58. Although TWI is a relative index, the variable class presents a fairly large range of values that, compared to the non-variable class, can be considered high. It should also be noted that the highest values´ areas are the most propitious to reach saturation which might lead to root asphyxia and possibly death.

When looking at non-variable class, it presents a small range covered by mainly mean values of TWI.

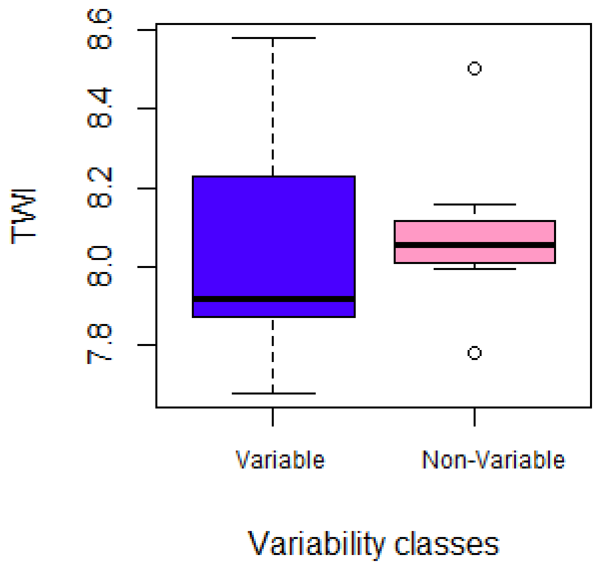


Figure 22-TWI's values oscillations per variability class.

Concerning the electrical soil conductivity (ECa), its values go from 4.60 to a maximum of 24.2.

Analysing the Figure 23, we can observe that the non-variable class have ECa average values, ranging from the medium to the highest, while the variable class is the opposite. It presents values that go from the minimum to intermediate ECa values.

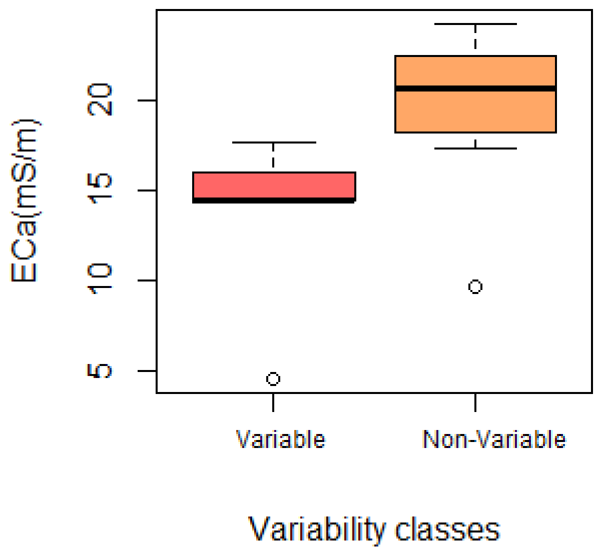


Figure 23-ECa's values oscillations per variability class.

As we can see in Figure 24, variable class is composed by sandy and clayey soils in percentages of 80% and 20% respectively. The non-variable class is composed, by sandy and clayey soils´ composed fields in the same amount.

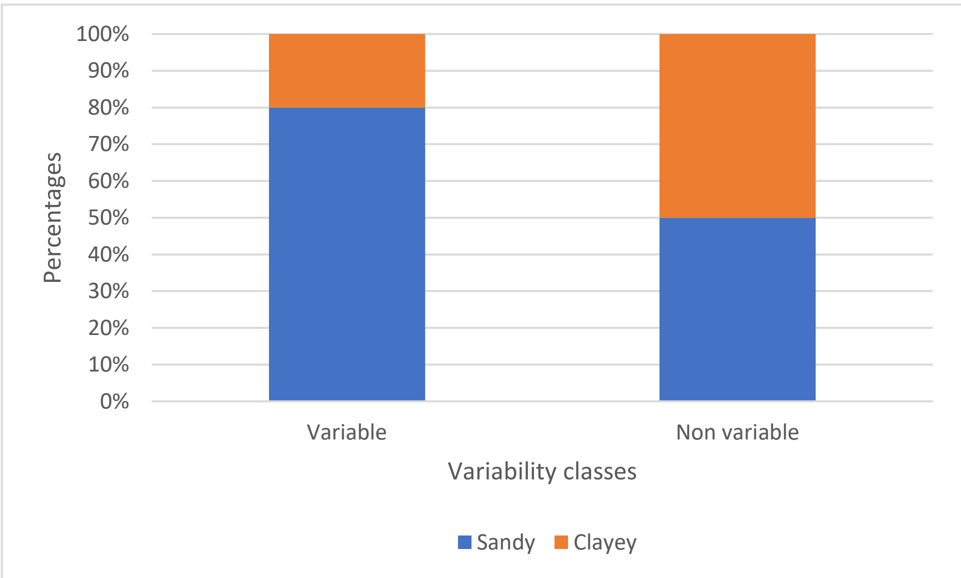


Figure 24-Soil presentation by variability class.

4.2 Correlation analysis

The correlation analysis allows us to verify the relationship between each variable under study. That is, the correlation coefficient indicates the linearity that may exist between the variables, which can be positive or negative (between -1 and 1).

Analysing Table 2, in relation to the correlations between the response variables, yield standard deviation and average yield, it is observed that there is a negative correlation with a reasonable value of -49%.

As for the direct correlations between the explanatory variables, we observed only the existence of significant correlations between Altimetry and the Topographic Wetness Index (TWI), with a value of 56%. For the other variables and their correlations, they have little influence on each other.

Regarding the correlation between the variable Yield standard deviation and the explanatory variables, we observed that there are some significant correlations such as the negative correlation with ECa, of -53%. The rest have uninfluential values.

Still in the same table we also find the correlations between the variable Average Yield and the others. The results show significant correlations of about 68% with ECa. Regarding Altimetry and TWI, they present reasonable values of 39% and 30%, respectively.

Soil type wasn´t included in the correlation table since it is a categorical variable.

Table 2-Correlation analysis between response and explanatory variables.

	Yield standard deviation	Average Yield	ECa	Altimetry	TWI
Yield standard deviation	1				
Average Yield	-0.49	1			
ECa	-0.53	0.68	1		
Altimetry	-0.10	0.39	0.05	1	
TWI	0.24	0.30	0.28	0.56	1

4.3 Modelling temporal variability

To study the variation in yields observed in each field across the five years of study and to test the hypothesis that these variables can be temporal yield variability drivers, the following models were developed:

Table 3-Model 1 summary table.

Model	Observed vs Predicted	R-squared (R^2)	Adjusted R-squared (Adj.R^2)	AIC
Full model (1.1): **Yield standard deviation= ECa*Altimetry*TWI +Soil Type**	Figure 25	0.94	0.81	2.04
Additive model (1.2): **Yield standard deviation= ECa+Altimetry +Soil Type+TWI**	Figure 26	0.59	0.38	18.75
Model (1.3), reduced from 1.2 **Yield standard deviation= ECa+Altimetry +TWI**	Figure 27	0.57	0.43	17.16

*-Interaction between variables.

Observing Table 3, it is verified that, the best model to predict and explain the temporal variability among the fields under study is model 1.1, the full model, because it presents the following characteristics:

- It has an adjusted coefficient of determination (Adj.R^2) able to explain about 81% of the existing variability.
- Of the three models it presents the lower AIC value, 2.04, thus confirming that it is the most fitted.

The 1.1 model was not reduced because, by the stepwise backward method, the reduction of this would not present improvements, thus leaving the full model intact.

Of the other models, model 1.2, the additive model, even with reasonable results, was considered the least capable because it can only explain about 38% of the existing variability and because it has the highest value of AIC, 18.75.

Regarding model 1.3, the result of the application of stepwise backward method, proved to be a model capable of explaining 5% more than the additive model, com um an adjusted R^2 of 43% and with an AIC value of 17.16.

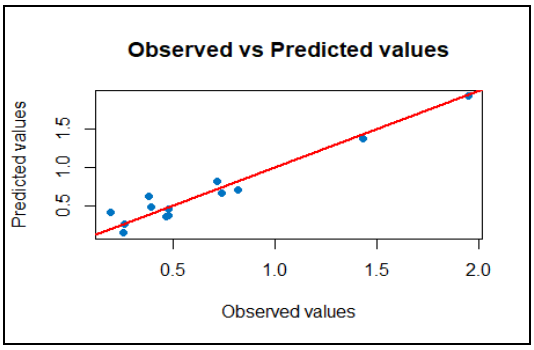


Figure 25-Graphical representation of the observed and predicted values by adjusting model 1.1.

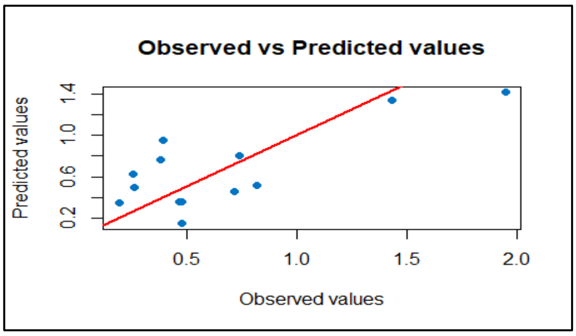


Figure 26-Graphical representation of the observed and predicted values by adjusting model 1.2.

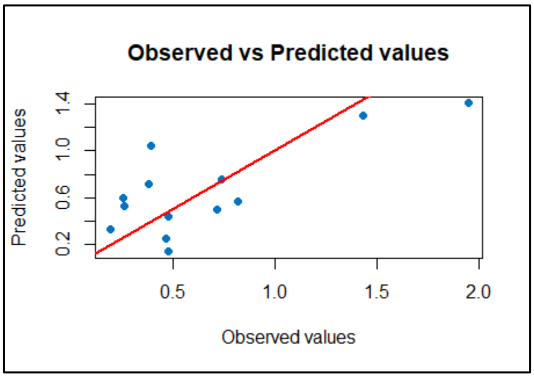


Figure 27-Graphical representation of the observed and predicted values by adjusting model 1.3.

4.4 Modelling spatial variability

To study the spatial variability between fields, the standardize average yields were calculated and, the same hypothesis was formulated, whether or not we could explain the special variability between fields using only the explanatory variables concerned. As already mentioned, we tried to develop models through two approaches, one that included the variability classes and the other that did not. The results were as follows:

1st approach models' summary table

Table 4-Model 2 summary table.

Model	Observed vs Predicted	R-squared (R^2)	Adjusted R-squared (Adj.R^2)	AIC
Full model (2.1): Yield average= ECa*Altimetry*TWI +Soil Type	Figure 28	0.85	0.54	17.59
Additive model (2.2): Yield average= ECa+Altimetry +Soil Type+TWI	Figure 29	0.63	0.44	21.04
Model (2.3), reduced from 2.2. Yield average= ECa+Altimetry	Figure 30	0.59	0.51	18.41

*-Interaction between variables.

Regarding Table 4, the results show us that the best model to predict and explain the spatial variability between fields is model 2.1, because it presents the following characteristics:

- It has an adjusted coefficient of determination (Adj.R^2) able to explain about 54% of the existing variability;
- Of the three models it presents the lower AIC value, 17.59, thus confirming that it is the most fitted.

The 2.1 model was not reduced because, by the stepwise backward method, the reduction of this would not present improvements, thus leaving the full model intact.

Of the three models, model 2.1, even with reasonable results, was considered the least capable because it can only explain about 44% of the existing variability and because it has the highest value of AIC, 21.04.

Regarding model 2.3, the result of the reduction of model 2.2 by the stepwise backward method, it obtained better results than 2.2, explaining 7% more than the worst model, with an adjusted R^2 of 51% and with an AIC value of 18.41.

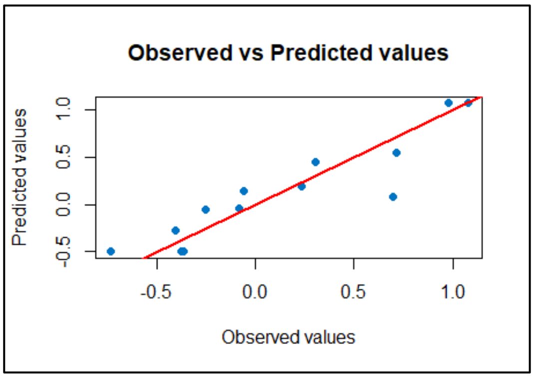


Figure 28-Graphical representation of the observed and predicted values by adjusting model 2.1.

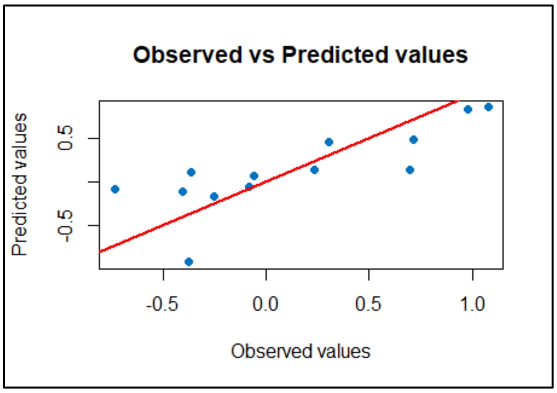


Figure 29-Graphical representation of the observed and predicted values by adjusting model 2.2.

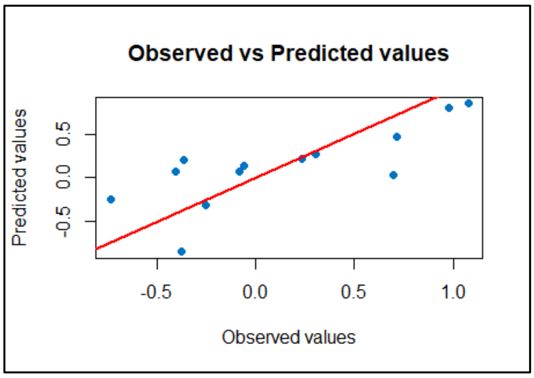


Figure 30-Graphical representation of the observed and predicted values by adjusting model 2.3.

2nd approach models' summary table

Table 5-Model 3 summary table.

Model	Observed vs Predicted	R-squared (R^2)	Adjusted R-squared (Adj.R^2)	AIC
Additive model (3.1): Yield average= (ECa+Altimetry +Soil Type+TWI)*Variability classes	Figure 31	0.98	0.96	-15.03
Model (3.2): same structure as 2.3, plus interaction with classes. Yield average= (ECa+Altimetry)*Variability classes	Figure 32	0.74	0.56	18.36

Regarding Table 5, the results show us model 3.1 is capable of predicting and explaining the spatial variability between fields, because it presents the following characteristics:

- It has an adjusted coefficient of determination (Adj.R^2) able to explain about 44% of the existing variability;
- It presents the AIC value of -15.03.

The additive model was not reduced because, by the stepwise backward method, the reduction of this would not present improvements, thus leaving the additive model intact.

It should be noted that a so-called complete model, with interactions between all variables, was not presented, because the model would obtain unreliable results.

Regarding model 3.2, obtained using the reduced model 2.3´s structure interacting with variability classes, it achieved worse results than 3.1, with an adjusted R^2 of 56% and with an AIC value of 18.36.

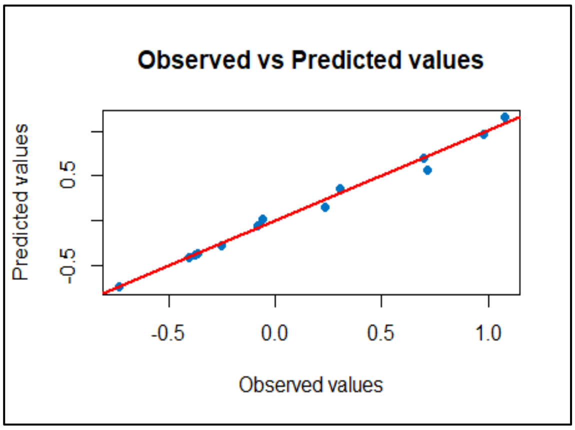


Figure 31-Graphical representation of the observed and predicted values by adjusting model 3.1.

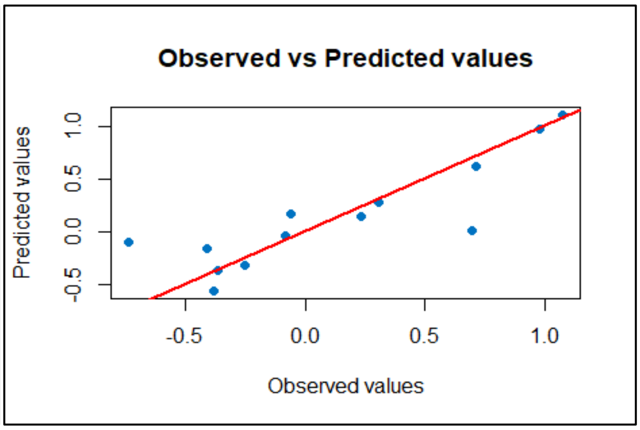


Figure 32-Graphical representation of the observed and predicted values by adjusting model 3.2.

4.5 Within-Field Spatial and Temporal Variability Management

In order to make a within field analysis, from the 13 fields available, two were chosen. The selection criteria were based on the fact that they are fields with different areas, different types of soils, different irrigation systems and because they have complete information regarding the four variables under study. This last factor, despite having been considered in the fields´selection, ended up not being included in this part of the work.

The field Avis Cob Velha, with an area of 6.66 hectares, is characterized by sandy soils and irrigated by a solid set system while the Lameiras´ field has an area of 18.25 hectares and is composed by clayey soils and irrigated by a center pivot.

For the within field analysis, the data available between 2015 and 2019 was used.

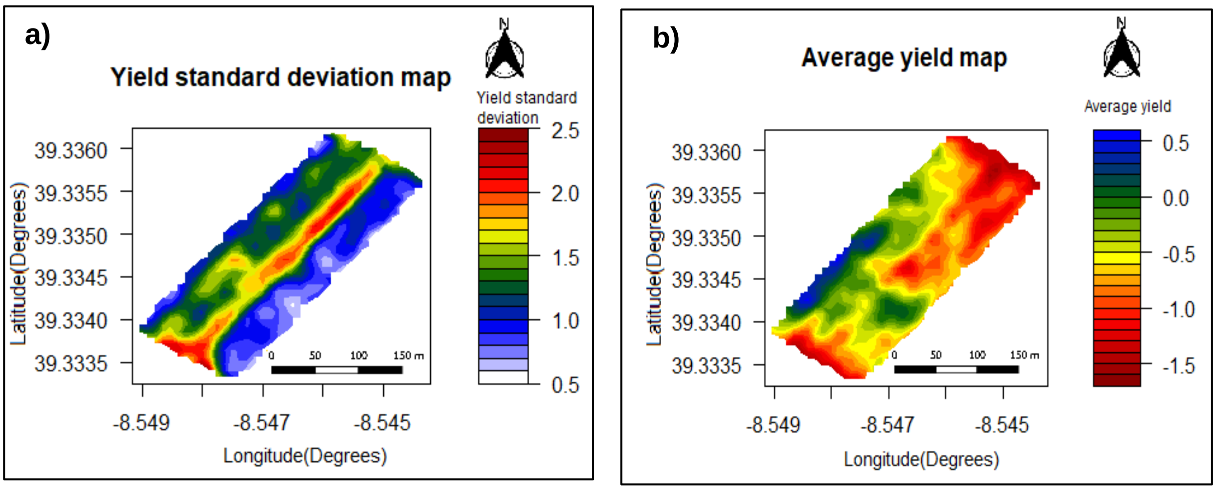
The first step was plotting the raw data points, followed by the creation of semivariograms, both presented in the appendix C for each field. Its analysis allowed us to observe the expected, that is, observations that are closer to each other have a more similar behaviour, reflected in lower semivariances, than those located at a greater distance.

This spatial structure in the data indicates that it is possible to create contiguous management zones based on the georeferenced point data.

To more easily visualize the spatial and temporal characteristics of the selected fields, an interpolation was made through the interp() function of R. followed by an improvement in the boundaries between zones within field using the filled.contour() R. function.

Maps were created to visualize temporal and spatial variability, and to see if there were patterns.

The same methods, previously mentioned in the classification of the variability of the fields, were applied to classify all georeferenced points according to the same criteria. The results are shown in Figure 33, 34 and Table 6.



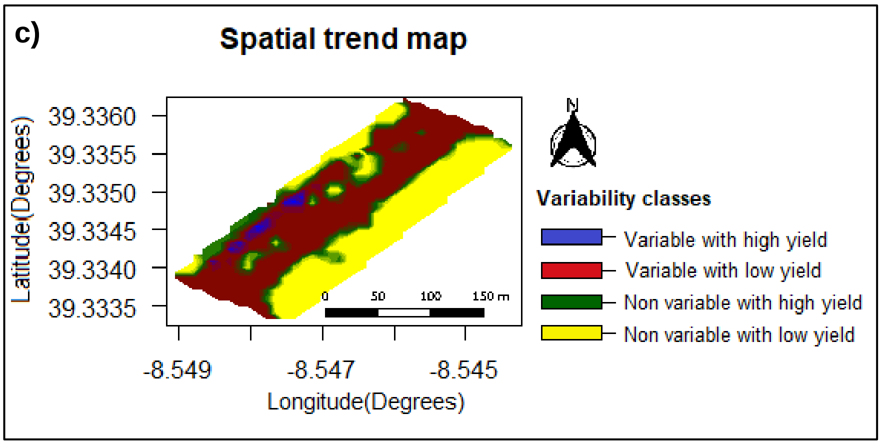


Figure 33-Avis Cob Velha's within field analysis maps: a) Yield standard deviation map, b) Average yield map, c) Spatial trend map.

Figure 33 a) refers to the standard deviation across the years of the standardized yield of Avis Cob Velha´s field. The map indicates that the spatial distribution of yield has varied a lot over the years, with some areas with significant standard deviations values, mainly in the most central area of the field in redder colors, being considered the areas of lower consistency over time. When analysing its average yield map, we can observe that, in general, it is a field with low average yields, with only a few areas with above-average yields.

The combination of maps referring to temporal and spatial variability gives rise to a spatial trend map that combines the spatial and temporal characteristics of the field and, can be a starting point for future management decisions. The maps were classified according to four categories: variable with high yield, variable with low yield, non-variable with high yield and non-variable with low yield.

When analyzing Figure 33 c), we can observe some clear patterns. The predominant areas are colored in red and yellow respectively, corresponding to variable with low yield and non-variable with low yield areas.

These is a two simple management zone example that can be used in the future, and in the long run, as case study zones.

Table 6 shows us the area for each variability class. The field proved to be a field, with 18,9% of the area classified as variable with high yield, 30.2% variable with low yield, 30.5% non-variable with high yield and 20.4% non-variable with low yield.

Regarding the Lameiras field, when analysing the yield standard deviation map (Figure 34 a)), we easily observed that the field has standard deviation values mostly lower than 1, except for some zones leading us to conclude that overall, it has been a consistent field over the years.

When analysing the average yield map (Figure 34 b), we find that it is, in average, a highly productive field, with most of the area having standardized average yield values greater than 0, characterized by greener and bluer areas. The map also shows small less productive zones colored in red.

The combination of these two maps originated the spatial trend map that is represented in Figure 34 c) and show us a variable field.

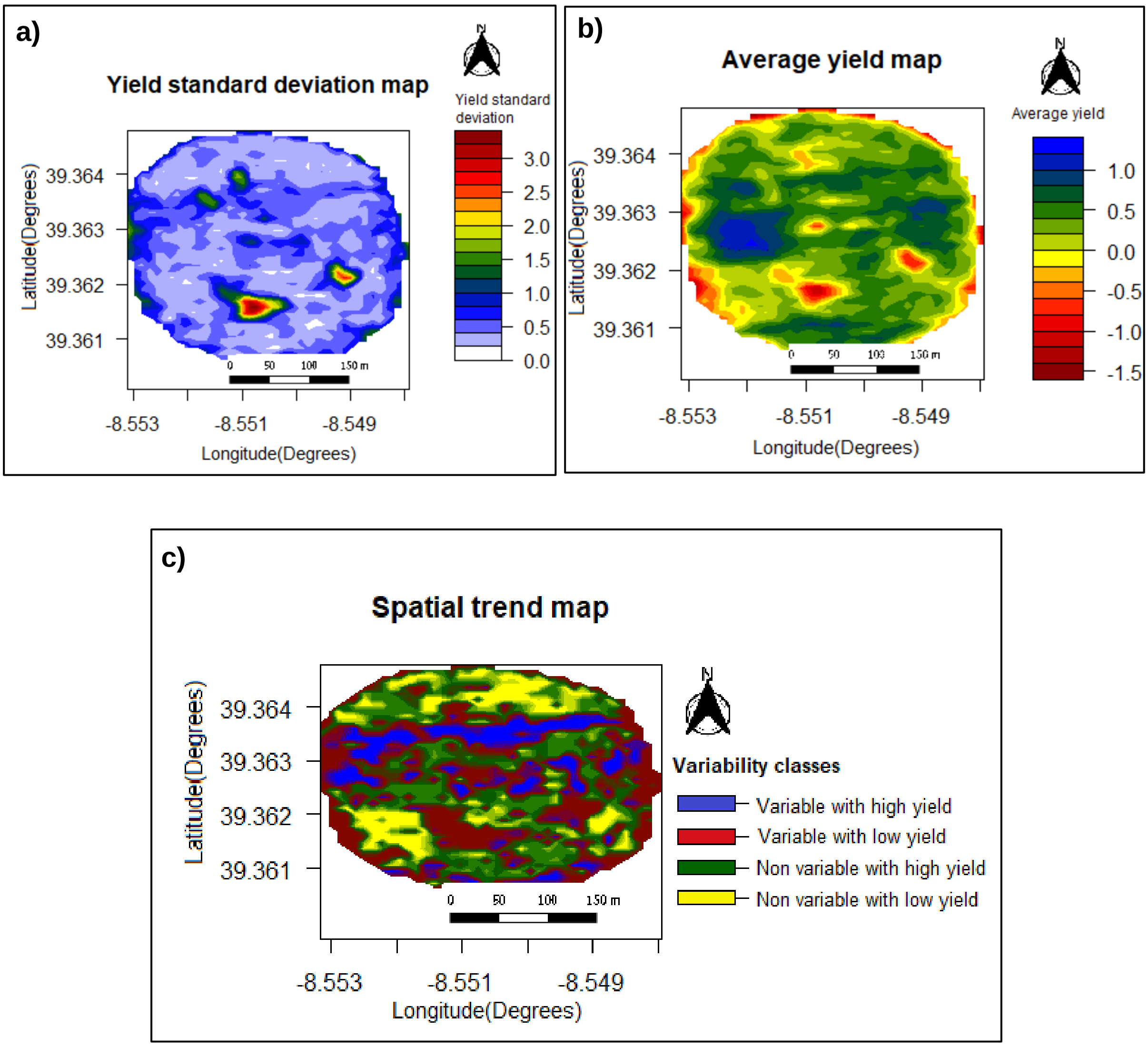


Figure 34-Lameiras' within field analysis maps: a) Yield standard deviation map, b) Average yield map, c) Spatial trend map.

Table 6-Area in each variability zone within fields based on the number of 7 × 7 m grid cells in each field.

Field	Variability Class	Number of Grid cells	% of field	Area (ha)	Total Area (ha)
Avis Cob Velha	Variable and High yield	553	18.9	1.26	6.66
	Variable and Low yield	882	30.2	2.01	
	Non-variable and High yield	892	30.5	2.03	
	Non-variable and Low yield	598	20.4	1.36	
Lameiras	Variable and High yield	27	2.6	0.47	18.25
	Variable and Low yield	499	48.0	8.76	
	Non-variable and High yield	60	5.8	1.05	
	Non-Variable and Low yield	454	43.7	7.97	

5. Discussion

This study had several objectives, among them, the identification of the fields ‘variations based on the existing temporal and spatial variability obtained, respectively, through the analysis of the standard deviation across the years of the standardized yield and the standardized average yields observed between fields, the identification of driving factors that would allow to explain and predict the variations in yields, that is, to study the possibility of predicting the variability of the fields, which drives the variation in the means and to what extent the variability zones have different yields and responses to driving factors, and finally, the study of the design of intervention zones, to an within field level, based on the variability classes of two chosen fields.

The first part of the analysis allowed us to have a general idea of what goes on in the fields explored by Quinta da Cholda SA. In 13 fields analysed, 62% of the fields were considered non-variable throughout the study period, which indicates that farmers' concerns about adapting the fields to reality and rapid technological development have been reflected in their results. The farmer has long gathered a wide variety of information to know the variability in his fields and then improve the decision-making process good based on geospatial information.

For this study, we chose to choose four variables in order to study the interactions with each other and to find and analyse possible patterns related to the fields ‘variability. The results covered a wide range of average values not allowing us to find particular patterns, only getting an idea of the variables’ range values´ variation between variability classes.

The second phase of the analysis allowed us to study what were the factors that led to fluctuations in yields for the defined period. For this, multiple linear models were created in order to understand what drived the variations.

Regarding temporal variability, the model that was able to explain most of the existing variability was the complete and more complex, model 1.1, **Yield standard deviation=ECa*Altimetry*TWI+Soil Type**, Table 3, which being evaluated by adjusted R-squared of 81% and 2.04 AIC criterion value, obtained significantly better results than model 1.2 and 1.3. The most fitted model showed us, according to ANOVA tables presented in Appendix B, that the singular value of the variable ECa and TWI had some significance as well as the result of the interaction between them. The fact that they are, in practice, variables that are related to soil characteristics, may be an important factor to study in the future.

The second part of the modelling process was based on the study of spatial variability and was divided into two approaches, one that studied only the variation in means and the other that studied the variation of the means and that allowed us to understand if we could distinguish variable and non-variable fields, if they were also the higher or lower yield fields or if they had different interaction with, in this case, ECa, TWI, Altimetry or soil type.

Regarding the first approach, the model that obtained the best result was the most complete and complex, model 2.1, **Yield average= ECa*Altimetry*TWI +Soil Type**, Table 4, which included the interaction between TWI, ECa and Altimetry, together with the addition of the effect related to soil type. A reasonable model was taken to explain about 54% of the existing variability, values higher than the other models. According to ANOVA table's presented in Appendix B, the only factor that really had a higher significance, compared to others, was the ECa, proving one more time, being a useful study variable.

The second approach, was different from the first, in the sense that it started from the additive model, **Yield average= (ECa+Altimetry+TWI +Soil Type)*Variability classes**, Table 5 and not a complete one. In addition, to have a comparison term, we chose to explore model´s 2.3 structure, reduced from 2.2, interacting with variability classes, obtaining this way model 3.2, a model with worse results than the first.

A model capable of explaining about 96% of the existing variability was obtained, thus being considered a good model. That said, it can be affirmed that it is possible to distinguish whether, within a variable or non-variable field, its level of productivity, and understand the weight that each variable, individual or interacting with another has in the average yield. In this case, when analyzing the ANOVA table, present in Appendix B, we can see that altimetry and electrical conductivity were the most significant singular variables, however, with regard to the interaction with the variability classes, soil type and altimetry had special significance, highlighting its importance to explain average yields between fields. It should be noted that the predefined variability classes, individually, had no significant interaction with the average yields, which may mean that the criteria defined for each class may require some adjustments.

Thus, it can be said that it is possible to predict with some certainty the temporal and spatial variability through the chosen variables, however, it is emphasized that for a better conclusive analysis, the models should be validated in other places and in other regions.

The establishment of management areas based on the study of the variability of a field and areas within the same field allows the farmer, in the long term, to define strategies and make decisions in a better position based on existing patterns.

The last part of this work aimed to apply a method that would create a map that would combine spatial variation and temporal variation in the defined period and can indicate whether the farmer should focus on his management more spatially or temporally. The variation is the result of the interaction between several factors that have as final product the income obtained, however the study of these factors and their interactions implies time and intensive labour as reported by Diker *et al*. (2004).

Within the factors that can be studied is the soil. If the variability of a zone is significant at the spatial level, soil analyses are usually performed many times on a regular grid, which can be economic and physically expensive.

By applying this method, we can reduce the number of sampling points, and can also use technologies such as the measurement of the electrical conductivity of the soil, which never without soil samples, help to describe spatially and temporally the yields and understand their potential causes, namely the interaction of the soil with the existing nutrients and fertilizers applied. The existence of very mobile nutrients such as nitrogen, or potassium allows easier management during the growing crop season, however the existence of other also important but less mobile, such as phosphorus, sulphur, which even when applied in the liquid state, might show insufficient to meet the needs of the crop, thus being a long-term management a possible solution to the problem, increasing the reserves available on the ground and improving their availability to the plants. In this way, the variability classes can be seen as potential way of identifying nutritional variations and will give the farmer useful information regarding the application of fertilizers.

Still in relation to soils, and although the physical and chemical properties are not known in its entirety due to lack of studies, it is known that these two fields are composed essentially of clayey soils, in the case of Lameiras, and sandy in the case of Avis Cob Velha. It should be noted that there is similarity of studies related to the variability of the fields, as is the case of

Martinez *et al*. (2020) and Diker *et al.* (2004),the most variable fields were essentially sandy fields, which is not surprising, since compared to clayey soils, due to their physical and chemical characteristics, they are mostly nutrient-poor and with low water retention capacity. Nevertheless, as can be seen in Figure 24, of the 13 fields analysed, of the variable fields, only 20% were composed by clayey soils and the remaining 80% were sandy.

Since maize is a crop, in this case irrigated, the study of irrigation techniques can give us extra information. In the case of Avis Cob Velha, the fact that is irrigated through a solid set system, may be the starting point for a future study on the uniformity of water distribution by plot and possible technical corrections, despite being reviewed annually.

Regarding the Lameiras field, it is a field that presents great variability over the 5 years and, as in Avis Cob Velha, the study of the uniformity of irrigation should also be applied. Lameiras is a field watered by center pivot, and when looking at the ends of the fields we can see areas of constant variability with low yield levels. This may be explained by possible pressure losses, which in addition to affecting the uniformity of watering along the field, affects the final sprinkler of the pivot, creating the patterns at the ends as we can see in Figure 33 c).

The analysis of the existing patterns in both spatial trend maps, together with the junction of the information related to ECa, TWI and Altimetry, should be used to study mainly the largest variable zones of low yield.

In general, there are several factors that lead to agronomic uncertainty, however if we identify some, potentially this uncertainty will be reduced. Another useful way to use variability classes in a practical way would be to create a gross margin map Blackmore (2000),something that has not been explored here but which allows us to understand whether it pays, at an economic level, to study and intervene in certain areas and if it does not compensate, seeing that perhaps reducing inputs, or even not sowe in those areas, may be a solution.

This study allowed the study and present possible management zones through precision farming tools, and an increase in the range of offers of action strategies to the farmer, however, it is a continuous work that needs to be deepened.

6. Conclusion

The use of precision agriculture tools, together with an appropriate treatment and analysis allows the study of existing spatial and temporal variability.

This study tells us that it is possible to use variables such as ECa, Altimetry, TWI and soil type to understand the variability of yields over the years at various levels.

In this case, the variable that obtained the best results was ECa, reinforcing its usefulness as a tool to study the physical characteristics of the soil and consequently the existing variability. In relation to TWI, it had special significance in the explanation of temporal variability, and, with respect to spatial variability, it also had some effect. As for altimetry, it was only important in explaining spatial variability.

The variable Soil Type had the worst result in the explanation of variability between field, thus demonstrating the need for replacement by another variable with greater detail.

Regarding the spatial trend maps, from the perspective of practical functionality for the farmer, it is possible to create maps combining both variabilities, allowing us to visualize homogeneous areas more easily for future research and treatment. The combination of spatial and temporal information with a gross margin map will potentially determine in future research, the cost of variability.

In conclusion, more research, data for a greater number of years, combined with, such as climatic data, remote sensing, and an analysis of the chemical physical properties of soil's data is needed. The combination of all these layers of information will allow us to offer new tools and understand what is the best way to intervene and what resources we should allocate.

7. Appendix A – Description of the fields in study.

Table 7-Area, irrigation system and yields´ description by field.

Fields	Area (ha)	Irrigation system	Average yield (t/ha)				
			Minimum	Maximum	Mean	Median	Standard.deviation
Aviz Cob Nova	8.1	Solid set	11.12	17.56	14.98	15.22	1.21
Aviz Cob Velha	6.66	Solid set	12.96	18.29	15.54	16.65	2.65
Aviz Júlia	26.82	Center pivot	11.84	19.26	16.61	16.81	1.22
Aviz Mira	26.61	Center pivot	12.61	19.14	17.19	17.69	1.55
Cerca	23.3	Center pivot	8.78	17.65	14.94	16.02	3.00
Estação	8.13	Center pivot	8.63	18.82	15.98	16.23	0.89
Lameiras	18.25	Center pivot	12.79	19.95	17.48	17.30	0.48
Lourenço	17.85	Center pivot	13.52	19.26	17.67	17.71	0.40
Mendanha	11.77	Center pivot	14.10	20.51	18.29	18.46	0.66
Moitas Meio	18.46	Solid set	8.31	18.78	13.96	17.32	6.43
Onias_Cobertura	4.42	Solid set	12.57	18.24	15.88	15.72	0.29
Onias_Pivot	10.74	Center pivot	12.24	18.63	16.54	16.25	0.57
Pessegueiros	9.63	Solid set	10.57	18.67	15.73	16.13	1.14

Table 8-Study parameters´ description by field.

Fields	Soil type	ECa(mS/m)			Altimetry(m)			TWI		
		Minimum	Maximum	Average	Minimum	Maximum	Average	Minimum	Maximum	Average
Aviz Cob Nova	Sandy	8.07	21.75	14.43	66.28	67.78	67.14	4.05	16.68	7.68
Aviz Cob Velha	Sandy	0.76	11.35	4.60	66.93	67.95	67.53	4.38	15.72	7.87
Aviz Júlia	Sandy	1.46	30.11	20.68	67.01	68.03	67.51	4.29	18.12	8.00
Aviz Mira	Sandy	7.19	29.43	17.62	66.73	67.89	67.35	4.45	18.31	7.92
Cerca	Clayey	7.22	30.43	14.44	68.42	70.03	69.20	3.85	18.22	8.23
Estação	Clayey	7.20	28.55	17.34	67.30	69.14	67.98	4.64	15.93	7.78
Lameiras	Clayey	4.21	44.37	19.11	67.38	70.34	68.89	3.95	15.95	8.08
Lourenço	Clayey	10.91	43.39	24.18	67.65	69.91	68.76	4.33	17.73	8.16
Mendanha	Clayey	10.75	42.04	22.15	68.88	70.44	69.65	4.82	15.49	8.50
Moitas Meio	Sandy	6.87	24.96	15.97	66.10	68.84	68.07	4.15	16.84	8.58
Onias_Cobertura	Sandy	3.98	31.44	20.62	66.20	67.34	66.59	4.76	14.17	8.03
Onias_Pivot	Sandy	11.19	29.52	22.64	66.23	67.43	66.65	4.86	16.01	8.05
Pessegueiros	Sandy	2.88	19.19	9.63	67.21	68.92	68.36	4.61	15.20	8.07

8. Appendix B – Model´s Summary and ANOVA´s output tables.

Linear model output: Yield temporal variability

Model 1.1

Table 9-Yield temporal variability: Model 1.1 output.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	2589.2287	1398.4094	1.852	0.138
ECa	-124.5565	74.0104	-1.683	0.168
Altimetry	-39.1574	20.4833	-1.912	0.128
TWI	-296.7175	174.9690	-1.696	0.165
Soil type. Clayey	0.3865	0.3291	1.174	0.305
ECa:Altimetry	1.8806	1.0800	1.741	0.157
ECa:TWI	14.3680	9.2539	1.553	0.195
Altimetry:TWI	4.5006	2.5613	1.757	0.154
ECa:Altimetry: TWI	-0.2175	0.1350	-1.612	0.182
Residual std. error	0.2187 on 4 DF			
Adjusted R^2	0.8139			
R^2	0.938			
F-statistic	7.56 on 8 and 4 DF			
P-value:	0.03393			

Table 10-Yield temporal variability: Model 1.1 ANOVA results.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ECa	1	0.85102	0.85102	17.7981	0.01349 *
Altimetry	1	0.01803	0.01803	0.3770	0.57242
TWI	1	0.89414	0.89414	18.7000	0.01240 *
Soil type	1	0.04057	0.04057	0.8484	0.40913
ECa: Altimetry	1	0.28953	0.28953	6.0553	0.06964 .
ECa:TWI	1	0.63919	0.63919	13.3679	0.02165 *
Altimetry:TWI	1	0.03513	0.03513	0.7347	0.43967
ECa:Altimetry:TWI	1	0.12417	0.12417	2.5970	0.18236
Residuals	4	0.19126	0.04782		
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**					

Model 1.2

Table 11-Yield temporal variability: Model 1.2 output.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	15.21974	18.23141	0.835	0.4280
TWI	1.60681	0.75054	2.141	0.0647 .
ECa	-0.07424	0.03025	-2.454	0.0397 *
Altimetry	-0.38780	0.32864	-1.180	0.2719
Soil type. Clayey	0.27734	0.55064	0.504	0.6281
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**				
Residual std. error	0.3999 on 8 DF			
R^2	0.5851			
Adjusted R^2	0.3776			
F-statistic	2.83 on 4 and 8 DF			
P-value:	0.09902			

Table 12-Yield temporal variability: Model 1.2 ANOVA results.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TWI	1	0.17150	0.17150	1.0725	0.33067
Altimetry	1	1.16511	1.16511	7.2860	0.02711 *
ECa	1	0.42658	0.42658	2.6676	0.14105
Soil type	1	0.04057	0.04057	0.2537	0.62807
Residuals	8	1.27929	0.15991		
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**					

Model 1.3

Table 13-Yield temporal variability: Model 1.3 output.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	7.0290	7.8927	0.891	0.3963
TWI	1.3653	0.5529	2.469	0.0356 *
Eca	-0.0638	0.0211	-3.024	0.0144 *
Altimetry	-0.2397	0.1405	-1.706	0.1223
Signif. Codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**				
Residual std. error	0.3829 on 9 DF			
R^2	0.5719			
Adjusted R^2	0.4292			
F-statistic	4.008 on 3 and 9 DF			
P-value:	0.04577			

Table 14-Yield temporal variability: Model 1.3 ANOVA results.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TWI	1	0.17150	0.17150	1.1695	0.30763
ECa	1	1.16511	1.16511	7.9448	0.02009 *
Altimetry	1	0.42658	0.42658	2.9088	0.12229
Residuals	9	1.31985	0.14665		
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**					

Linear model output: Yield spatial variability, 1st approach

Model 2.1

Table 15-Yield spatial variability, 1st approach: Model 2.1 output.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-3615.6379	2542.6339	-1.422	0.228
TWI	444.8182	318.1343	1.398	0.235
ECa	145.5748	134.5680	1.082	0.340
Altimetry	53.8321	37.2433	1.445	0.222
Soil type. Clayey	-0.5111	0.5985	-0.854	0.441
TWI:ECa	-18.0892	16.8257	-1.075	0.343
TWI:Altimetry	-6.6253	4.6570	-1.423	0.228
ECa:Altimetry	-2.1735	1.9637	-1.107	0.330
TWI:ECa:Altimetry	0.2702	0.2454	1.101	0.333
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**				
Residual std. error	0.3976 on 4 DF			
R^2	0.8463			
Adjusted R^2	0.5388			
F-statistic	2.752 on 8 and 4 DF			
P-value:	0.1717			

Table 16-Yield spatial variability, 1st approach: Model 2.1 ANOVA results.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TWI	1	0.37087	0.37087	2.3461	0.20035
ECa	1	1.57019	1.57019	9.9331	0.03446 *
Altimetry	1	0.52657	0.52657	3.3312	0.14203
Soil type	1	0.12006	0.12006	0.7595	0.43267
TWI:ECa	1	0.27385	0.27385	1.7324	0.25846
TWI:Altimetry	1	0.42071	0.42071	2.6615	0.17814
ECa:Altimetry	1	0.00683	0.00683	0.0432	0.84548
TWI:ECa:Altimetry	1	0.19171	0.19171	1.2128	0.33260
Residuals	4	0.63230	0.15808		
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**					

Model 2.2

Table 17-Yield spatial variability, 1st approach: Model 2.2 output.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-30.74382	19.90804	-1.544	0.161
TWI	-0.73595	0.81957	-0.898	0.395
ECa	0.09178	0.03303	2.779	0.024 *
Altimetry	0.52110	0.35887	1.452	0.185
Soil type. Clayey	-0.47711	0.60127	-0.794	0.450
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**				
Residual std. error	0.4367 on 8 DF			
R^2	0.6291			
Adjusted R^2	0.4437			
F-statistic	3.393 on 4 and 8 DF			
P-value:	0.06652			

Table 18-Yield spatial variability, 2nd question,1st approach: Model 2.2 ANOVA results.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TWI	1	0.37087	0.37087	1.9450	0.20063
ECa	1	1.57019	1.57019	8.2349	0.02084 *
Altimetry	1	0.52657	0.52657	2.7616	0.13513
Soil type	1	0.12006	0.12006	0.6296	0.45038
Residuals	8	1.52540	0.19068		
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**					

Model 2.3

Table 19-Yield spatial variability, 1st approach: Model 2.3 output.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-15.99551	8.39653	-1.905	0.08591 .
Altimetry	0.21952	0.12366	1.775	0.10626
ECa	0.07012	0.02162	3.243	0.00882 **
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**				
Residual std. error	0.4117 on 10 DF			
$\mathbf{R^2}$	0.588			
Adjusted $\mathbf{R^2}$	0.5056			
F-statistic	7.135 on 2 and 10 DF			
P-value:	0.01188			

Table 20-Yield spatial variability ,1st approach: Model 2.3 ANOVA results.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Altimetry	1	0.63555	0.63555	3.7502	0.081548 .
ECa	1	1.78282	1.78282	10.5199	0.008818 **
Residuals	10	1.69472	0.16947		
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**					

Linear model output: Yield spatial variability, 2nd approach

Model 3.1

Table 21-Yield spatial variability, 2nd approach: Model 3.1 output.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-6116.4675	748.3956	-8.173	0.00383 **
TWI	-109.9584	13.4303	-8.187	0.00381 **
ECa	1.9151	0.2291	8.360	0.00359 **
Altimetry	103.2537	12.6323	8.174	0.00383 **
Soil type. Clayey	-152.0783	18.5847	-8.183	0.00382 **
Non-variable class	6085.6331	748.4277	8.131	0.00389 **
TWI: Non-variable class	109.8931	13.4395	8.177	0.00383 **
ECa: Non-variable class	-1.8243	0.2299	-7.937	0.00417 **
Altimetry: Non-variable class	-102.8116	12.6332	-8.138	0.00388 **
Soil type. Clayey: Non-variable class	151.8198	18.5862	8.168	0.00384 **
Residual std. error	0.1212 on 3 DF			
R^2	0.9893			
Adjusted R^2	0.9571			
F-statistic	30.76 on 9 and 3 DF			
P-value:	0.008423			

Table 22-Yield spatial variability ,2nd approach: Model 3.1 ANOVA results.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TWI	1	0.37087	0.37087	25.2315	0.015199 *
ECa	1	1.57019	1.57019	106.8263	0.001932 **
Altimetry	1	0.52657	0.52657	35.8251	0.009336 **
Soil type	1	0.12006	0.12006	8.1680	0.064684 .
Variability class	1	0.05229	0.05229	3.5573	0.155755
TWI:Variability class	1	0.21790	0.21790	14.8243	0.030935 *
ECa:Variability class	1	0.06965	0.06965	4.7386	0.117698
Altimetry: Variability class	1	0.16075	0.16075	10.9365	0.045493 *
Soil type: Variability class	1	0.98072	0.98072	66.7228	0.003838 **
Residuals	3	0.04410	0.01470		
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**					

Model 3.2

Table 23-Yield spatial variability, 2nd approach: Model 3.2 output.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	8.21165	16.10923	0.510	0.626
Altimetry	-0.13286	0.23813	-0.558	0.594
ECa	0.04208	0.03849	1.093	0.310
Non-variable class	-32.22434	18.60582	-1.732	0.127
ECa: Non-variable class	0.46815	0.27415	1.708	0.131
Altimetry: Non-variable class	0.03772	0.05045	0.748	0.479
Residual std. error	0.39 on 7 DF			
R^2	0.7411			
Adjusted R^2	0.5562			
F-statistic	4.007 on 5 and 37DF			
P-value:	0.04898			

Table 24-Yield spatial variability ,2nd approach: Model 3.2 ANOVA results.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Altimetry	1	0.63555	0.63555	4.1776	0.08025 .
ECa	1	1.78282	1.78282	11.7189	0.01109 *
Variability class	1	0.07845	0.07845	0.5157	0.49596
Altimetry: Variability class	1	0.46629	0.46629	3.0650	0.12346
ECa:Variability class	1	0.08505	0.08505	0.5591	0.47900
Residuals	7	1.06493	0.15213		
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**					

9. Appendix C – Semivariograms and within field point data maps.

Avis Cob Velha

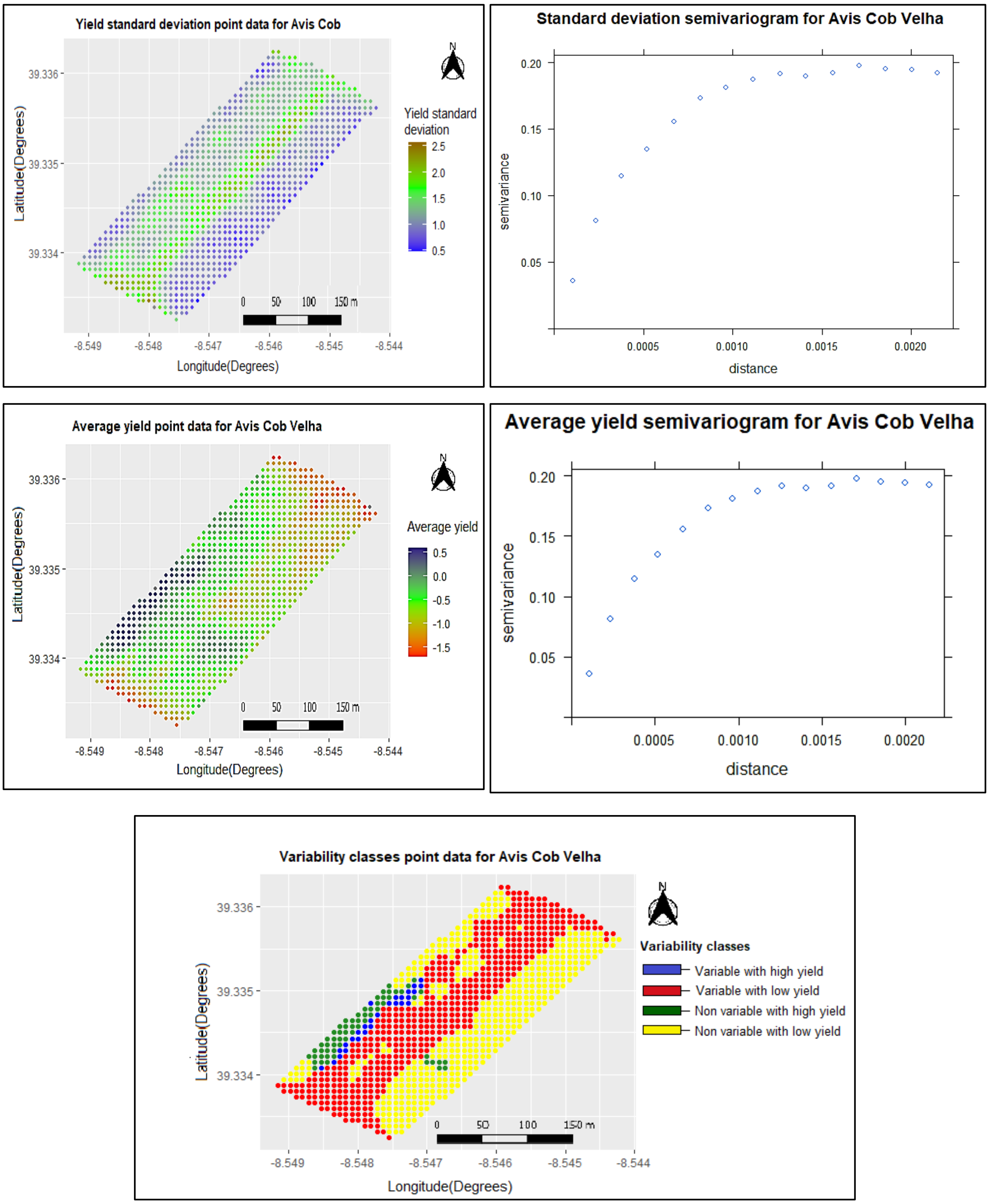


Figure 35- Avis Cob Velha´s within field point data maps.

Lameiras

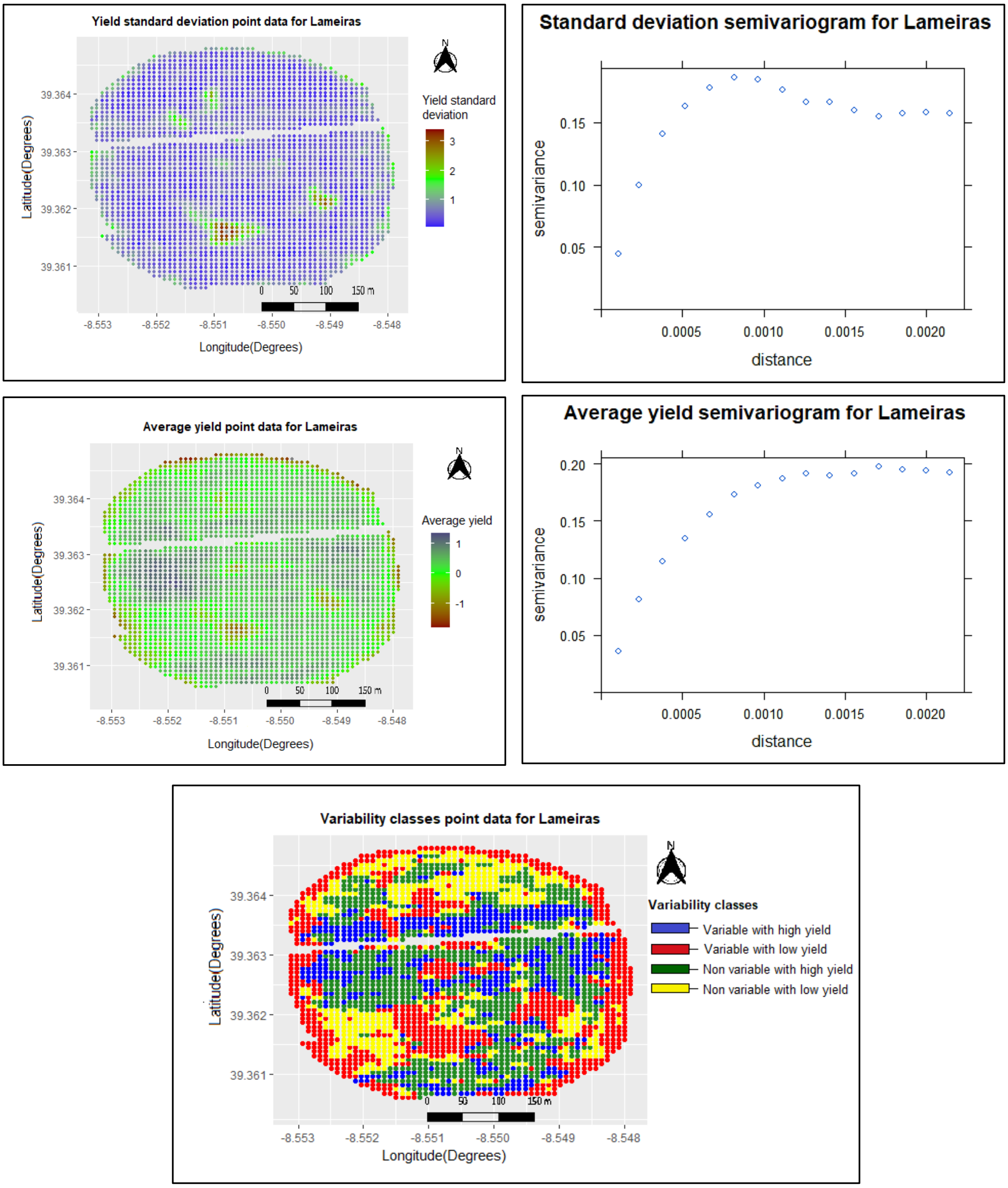


Figure 36-Lameiras´s within field point data maps.

10. Bibliography

Andreo, D. V. (2013) ‘Remote Sensing and Geographic Information Systems in Precision Farming :MEW 9 Seminario’, 1(1), p. 12.

Babak, O. and Deutsch, C. V. (2009) ‘Statistical approach to inverse distance interpolation’, *Stochastic Environmental Research and Risk Assessment*, 23(5), pp. 543–553. doi: 10.1007/s00477-008-0226-6.

Bevans, R. (2020a) *Akaike Information Criterion | When & How to Use It*. Available at: https://www.scribbr.com/statistics/akaike-information-criterion/ (Accessed: 19 October 2020).

Bevans, R. (2020b) *ANOVA em R | Um guia passo a passo completo com exemplos*. Available at: https://www.scribbr.com/statistics/anova-in-r/ (Accessed: 16 October 2020).

Bhat, M. A. and Ahmad, B. (2011) ‘Cloud Computing : A solution to Geographical Information Systems (GIS)’, *International Journal on Computer Science and Engineering (IJCSE)*, 3(2), pp. 594–600.

Blackmore, S. (2000) ‘The interpretation of trends from multiple yield maps’, *Computers and Electronics in Agriculture*, 26(1), pp. 37–51. doi: 10.1016/S0168-1699(99)00075-7.

Borba, M.Z.F., Freddi, O.S., Trevisan, R.G., Eitelwein, M.T. and Molin, J.P (2015) ‘Condutividade elétrica aparente no mapeamento da textura do solo’, *Anais do XXXV Congresso Brasileiro de Ciência do Solo*, (1), pp. 1–4.

Braga, R. (2017) *Agricultura de Precisão: Adoção em Portugal | LinkedIn*. Available at: https://www.linkedin.com/pulse/agricultura-de-precisão-adoção-em-portugal-ricardo-braga/ (Accessed: 7 May 2020).

Braga, R. and Pinto, P. A. (2015) ‘Agricultura de precisão: adopção & principais obstáculos’, *Agrotec-revista técnico-científica agrícola*, pp. 84–89.

Coelho, J. P. C. and Marques da Silva, J. R. (2009) *Agricultura de Precisão*. Lisbon: Associação de Jovens Agricultores de Portugal.

Corwin, D. L. and Lesch, S. M. (2005) ‘Apparent soil electrical conductivity measurements in agriculture’, *Computers and Electronics in Agriculture*, 46(1-3 SPEC. ISS.), pp. 11–43. doi: 10.1016/j.compag.2004.10.005.

Corwin, D. L. and Scudiero, E. (2019) ‘Mapping soil spatial variability with apparent soil electrical conductivity (EC A) directed soil sampling’, *Soil Science Society of America Journal*, 83(1), pp. 3–4. doi: 10.2136/sssaj2018.06.0228.

Davis, G., Casady, W. and Massey, R. (1998) ‘Water Quality Precision Agriculture : An Introduction’, *Precision Agriculture: An Introduction*, pp. 1–7.

Diker, K., Heermann, D. F. and Brodahl, M. K. (2004) ‘Frequency analysis of yield for delineating yield response zones’, *Precision Agriculture*, 5(5), pp. 435–444. doi: 10.1007/s11119-004-5318-9.

Drummond, S.T., Sudduth, K.A., Joshi, A., Birrell and S.J., Kitchen, N.R. (2003) ‘S n m s –s y p’, 46(1), pp. 5–14.

FAO (2014) *World reference base for soil resources 2014. International soil classification system for naming soils and creating legends for soil maps*, *World Soil Resources Reports No. 106*. doi: 10.1017/S0014479706394902.

Ganzel, B. and Reinhardt, C. (2018) *The Postwar Fertilizer Industry Explodes*. Available at: https://livinghistoryfarm.org/farminginthe40s/crops_04.html (Accessed: 4 May 2020).

Gebbers, R. and Adamchuk, V. I. (2010) ‘Precision agriculture and food security’, *Science*, 327(5967), pp. 828–831. doi: 10.1126/science.1183899.

Grego C. R., Oliveira R. P., V. S. R. (2014) ‘Geoestatística aplicada a Agricultura de Precisão’, *Agricultura de precisão: Resultados de um novo olhar*, pp. 74–83.

Hahs-Vaughn, D.L. and Lomax, R.G. (2020) ‘Multiple Linear Regression’, *Statistical Concepts*, pp. 527–599. doi: 10.4324/9780429277825-8.

Hakkim, V., Joseph, E., Gokul, A. and Mufeedha, K. ‘Precision Farming: The Future of Indian Agriculture’, *Journal of Applied Biology and Biotechnology*, 4(06), pp. 068–072. doi: 10.7324/jabb.2016.40609.

Han, Y. J. and Dodd, R. B. (1999) ‘A Review of Yield Monitoring Instrumentation Applied to the Combine Harvesters for Precision Agriculture Purposes’, *Department of Biosystems Engineering, Clemson University*, (November 2016), pp. 426–432.

Handayani, A. P., Deliar, A., Sumarto, I. and Syabri, I. (2019) ‘Inverse Distance Weighting interpolation on the optimum distribution of kernel-Geographically Weighted Regression for land price’, *IOP Conference Series: Earth and Environmental Science*, 389(1), pp. 0–9. doi: 10.1088/1755-1315/389/1/012031.

Hung, M. N. W. B. H. (2016) *11º Sinageo - Utilização do Índice Topográfico de Umidade como suporte ao planejamento e gestão ambiental de Unidades de Conservação de Uso Sustentável*. Available at: http://www.sinageo.org.br/2016/trabalhos/8/8-131-1550.html (Accessed: 5 August 2020).

IPMA-Clima Normais (2020). Available at: https://www.ipma.pt/pt/oclima/normais.clima/?print=true (Accessed: 21 October 2020).

Karydas, C., Gitas, I., Koutsogiannaki, E., Lydakis-Simantiris, N. and Silleos, G. (2009) ‘Evaluation of spatial interpolation techniques for mapping agricultural topsoil properties in Crete’, *EARSeL eProceedings*, pp. 26–39.

Kenton, W. (2020) *Multiple Linear Regression (MLR) Definition*. Available at: https://www.investopedia.com/terms/m/mlr.asp (Accessed: 23 September 2020).

Kharel, T.P., Maresma, A., Czymmek, K. J., Oware, E.K. and Ketterings, Q.M. (2019) ‘Combining spatial and temporal corn silage yield variability for management zone development’, *Agronomy Journal*, 111(6), pp. 2703–2711. doi: 10.2134/agronj2019.02.0079.

Kienzle, J. (2003) ‘Precision Agriculture for Smallholder Farmers.Agriculture for Development’, *TROPICAL AGRICULTURE ASSOCIATION (TAA)*, 20(2), pp. 53–82.

Kopecký, M. and Čížková, Š. (2010) ‘Using topographic wetness index in vegetation ecology: Does the algorithm matter?’, *Applied Vegetation Science*, 13(4), pp. 450–459. doi: 10.1111/j.1654-109X.2010.01083.x.

Leah, W. and Goulden, T. (2020) *Going On The Grid -- An Intro to Gridding & Spatial Interpolation*. Available at: https://www.neonscience.org/spatial-interpolation-basics (Accessed: 23 October 2020).

Long, J. D. and Teetor, P. (2019) *11 Linear Regression and ANOVA | R Cookbook, 2nd Edition*. Available at: https://rc2e.com/linearregressionandanova (Accessed: 19 October 2020).

Maestrini, B. and Basso, B. (2018) ‘Drivers of within-field spatial and temporal variability of crop yield across the US Midwest’, *Scientific Reports*, 8(1), pp. 1–9. doi: 10.1038/s41598-018-32779-3.

Martinez-Feria, R. A. and Basso, B. (2020) ‘Unstable crop yields reveal opportunities for site-specific adaptations to climate variability’, *Scientific Reports*, 10(1), pp. 1–10. doi: 10.1038/s41598-020-59494-2.

Mattivi, P., Franci, F., Lambertini, A. and Bitelli, G. (2019) ‘TWI computation: a comparison of different open source GISs’, *Open Geospatial Data, Software and Standards*. Open Geospatial Data, Software and Standards, 4(1). doi: 10.1186/s40965-019-0066-y.

Meena, A. K. (2019) ‘Use of Precision Agriculture for Sustainability and Environmental Protection’, (July). Available at: https://www.researchgate.net/publication/334251626.

O´Brien, P. . (2016) ‘Agriculture and the Industrial Revolution’, *The Economic History Review*, 30(1), pp. 166–181. Available at: http://www.jstor.org/stable/2595506.

Oksanen, J. (2020) *Digital Elevation Model | National Land Survey of Finland*. Available at: https://www.maanmittauslaitos.fi/en/research/interesting-topics/digital-elevation-model (Accessed: 20 October 2020).

Oliver, M. A. (2013) ‘Precision agriculture and geostatistics: How to manage agriculture more exactly’, *Significance*, 10(2), pp. 17–22. doi: 10.1111/j.1740-9713.2013.00646.x.

Perez-Ruiz, M. and K., S. (2012) ‘GNSS in Precision Agricultural Operations’, *New Approach of Indoor and Outdoor Localization Systems*, (October 2012), pp. 2–26. doi: 10.5772/50448.

QGIS.org, 2020. QGIS Geographic Information System. QGIS Association. http://www.qgis.org

Rabello, L. M. (2009) ‘Condutividade elétrica do solo, tópicos e equipamentos’, *Embrapa*, p. 19.

Rains, G. C. and Thomas, D. L. (2009) ‘Precision Farming: An Introduction’, *The University of Georgia:Cooperative Extension*, pp. 1–10.

Rains, G. and Thomas, D. (2015) ‘Precision Farming: An Introduction’, *Georgia Extension Bulletin, US.*, (March 2009).

Robertson, M., Carberry, P. and Brennan, L. (2007) ‘The economic benefits of precision agriculture: case studies from Australian grain farms’, *Retrieved March*, 12(March), p. 2012.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Shannon, D. K., E.Clay, D. and A. Sudduth, K. (2018) ‘An Introduction to precision agriculture’, pp. 1–12. doi: 10.2134/precisionagbasics.2016.0084.

Stafford, J. V. (2000) ‘Implementing precision agriculture in the 21st century’, *Journal of Agricultural and Engineering Research*, 76(3), pp. 267–275. doi: 10.1006/jaer.2000.0577.

Stombaugh, T. (2018) ‘Satellite-based Positioning Systems for Precision Agriculture’. doi: 10.2134/precisionagbasics.2017.0036.

Swain, K. C. and Singha, C. (2018) ‘Mapping of agriculture farms using GPS and GIS technique for precision farming’, *International Journal of Agricultural Engineering*, 11(2), pp. 269–275. doi: 10.15740/has/ijae/11.2/269-275.

Torres Coimbra, S. (2019) *Analysis of with-in field variability in an Irrigated Maize Farm Simão*

Torres Coimbra Dissertation for obtaining the master ' s degree in Engenharia Agronómica.

Tzafestas, S. G. (2014) *Mobile Robot Sensors*, *Introduction to Mobile Robot Control*. doi: 10.1016/b978-0-12-417049-0.00004-3.

Yuemei, H., Xiaoqin, Z., Jianguo, S. and Jina, N. (2008) 'Conduction between left superior pulmonary vein and left atria and atria fibrillation under cervical vagal trunk stimulation', *Colombia Medica*, 39(3), pp. 227–234. doi: 10.1234/590.

Zimmermann, P. H., Davidov, D., Orbach, R., Tao, L. J. and Zitkova, J. (1972) 'Exchange narrowing of fine structure in dilute magnetic alloys: Mg: Gd', *Physical Review B*, 6(7), pp. 2783–2790. doi: 10.1103/PhysRevB.6.2783.