UNIVERSIDADE DE LISBOA

Instituto de Geografia e Ordenamento do Território

Agricultural land systems: modelling past, present and future regional dynamics

Cláudia Morais Viana

Orientador(es): Prof. Doutor Fernando Jorge Pedro da Silva Pinto da Rocha
Prof.ª Doutora Maria Dulce Alves Freire
Prof.ª Doutora Patrícia Catarina dos Reis Macedo Abrantes

Tese especialmente elaborada para obtenção do grau de Doutor em Geografia, especialidade de Ciências da Informação Geográfica

2022

UNIVERSIDADE DE LISBOA

Instituto de Geografia e Ordenamento do Território

Agricultural land systems: modelling past, present and future regional dynamics

Cláudia Morais Viana

Orientador(es): Prof. Doutor Fernando Jorge Pedro da Silva Pinto da Rocha

Prof.ª Doutora Maria Dulce Alves Freire

Prof.ª Doutora Patrícia Catarina dos Reis Macedo Abrantes

Tese especialmente elaborada para obtenção do grau de Doutor em Geografia, especialidade de Ciências da Informação Geográfica

Júri:
Presidente: Doutor José Manuel Henriques Simões, Professor Catedrático e Presidente do Conselho Científico do Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa;
Vogais:

- Doutora Elisabete Silva, University Reader, Department of Land Economy, University of Cambridge;
- Doutor José António Pereira Tenedório, Professor Associado, Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa;
- Doutora Maria Teresa Amado Pinto Correia, Professora Catedrática, Escola de Ciências e Tecnologias da Universidade de Évora;
- Doutor Eusébio Joaquim Marques dos Reis, Professor Auxiliar, Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa;
- Doutor Fernando Jorge Pedro da Silva Pinto da Rocha, Professor Auxiliar, Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa, orientador;
- Doutor Eduardo Jonas da Costa Gomes, Investigador Júnior, Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa.

Tese financiada pela Fundação para a Ciência e a Tecnologia no âmbito da bolsa de doutoramento SFRH/BD/115497/2016

2022

Preamble

The Chapters 2–8 of the thesis are the full contents of seven scientific article which were published or are being considered for publication in peer-reviewed international scientific journals. Chapters 2–8 maintain the original content of the information published in (or submitted to) scientific journals; however, the structure and formatting is slightly adapted in order to improve consistency with other chapters. The written English (American English or UK English) may vary from one to the other chapter.

The author fully contributed to the entire process of each scientific article from conceiving and designing the analysis, to collecting and preparing data, performing the analysis, writing the article and all the stages of article peer-review and revision.

This thesis was supported by the Portuguese Foundation for Science and Technology (FCT) under Grant [number SFRH/BD/115497/2016] and Centre for Geographical Studies - Universidade de Lisboa and FCT under Grant [number UID/GEO/00295/2013, UIDB/00295/2020 + UIDP/00295/2020].

Acknowledgments

This thesis section is the culmination of a personal and professional extraordinary journey. A four-year period, part of which have been lived on a pandemic context. A few lines are insufficient to properly acknowledge all the incredible support, company, and advice provided by a vast number of people who made the last years so memorable; any acknowledgement I could make would always be small and incomplete.

First and foremost, I'd like to express my deepest and sincere gratitude to Jorge Rocha, my supervisor. Thank you for your constant encouragement and for giving me the freedom to bridge my interests. Your full commitment, enthusiasm, and availability are inspiring. I'd also want to express my appreciation to my co-supervisors, Dulce Freire and Patrícia Abrantes, for their, motivation, collaboration, contributions, and ideas. I'm extremely grateful. I wish to express my cordial appreciation to Profs. Paulo Pereira and Robert Gil Pontius Jr, for their availability, invaluable inputs, and assistance during part of this research. I'm deeply grateful, and I look forward to keeping our fruitful collaboration.

Sincere thanks are extended to the Centre for Geographical Studies and Institute of Geography and Spatial Planning, including the professors, staff and academic services, for their institutional assistance throughout the PhD program. Also, I'm thankful to the Portuguese Foundation for Science and Technology (FCT) for financing this thesis. In addition, I'd like to express my gratitude to the research group MOPT for allowing me to join the group and for funding my research activities. I'm grateful to all MOPT colleagues, particularly Profs. Eduarda Marques da Costa, Nuno Marques da Costa, and Paulo Morgado, for their lectures, recommendations, and support. Also, I'm thankful to the Riskam research group, particularly it former coordinator, Prof. José Luis Zêzere. I would like to give a thanks to the GEOMODLAB.

During my PhD, I met a wonderful group of people and worked in an engaging learning environment. I'd want to express my gratitude to my colleagues and friends Inês, Luis, and Maurício, with whom I shared many academic and personal interests and learned so much. They helped me with a part of this research and for that I am greatly indebted to all of you. I'd also like to express my gratitude to Angy for her help improving my speaking and writing English skills, moral support, entertainment, leisure distractions, and travel adventures. Many thanks to my colleagues and friends, particularly Ana, Carlos, Luís, Mauro, Marcelo, and Pedro for the good moments. Big thanks to the oldest and most recent GEOMODLAB friends and colleagues.

I am profoundly grateful to my friends Catarina, Joana(s), and Soraia for always providing me with all kinds of much-needed distractions such as amazing dinners and travel adventures. Thanks for your friendship and guidance.

I reserved my last acknowledgements for the ones that will always come first, my family. I want to thank my parents and brother for their unconditional love, immeasurable support, and

patience. There are certainly many whom I forgot to list here that have helped me somehow along this trajectory. A big thank you to you all!

Abstract

This thesis arises from the understanding of how the integration of concepts, tools, techniques, and methods from geographic information science (GIS) can provide a formalised knowledge base for agricultural land systems in response to future agricultural and food system challenges. To that end, this thesis focuses on understanding the potential application of GIS-based approaches and available spatial data sources for modelling regional agricultural land-use and production dynamics in Portugal.

The specific objectives of this thesis are addressed in seven chapters in Parts II through V, each corresponding to one scientific article that was either published or is being considered for publication in peer-reviewed international scientific journals. In Part II, Chapter 2 summarises the body of knowledge and provides the context for the contribution of this thesis within the scientific domain of agricultural land systems. In Part III, Chapters 3 and 4 explore remotely sensed and Volunteered Geographic Information (VGI) data, multitemporal and multisensory approaches, and a variety of statistical methods for mapping, quantifying, and assessing regional agricultural land dynamics in the Beja district. In Part IV, Chapters 5–7 explore the CA-Markov model, Markov chain model, machine learning, and model-agnostic approach, as well as a set of spatial metrics and statistical methods for modelling the factors and spatiotemporal changes of agricultural land use in the Beja district. In Part V, Chapter 8 explores an area-weighting GIS-based technique, a spatiotemporal data cube, and statistical methods to model the spatial distribution across time for regional agricultural production in Portugal.

The case studies in the thesis contribute practical and theoretical knowledge by demonstrating the strengths and limitations of several GIS-based approaches. Together, the case studies demonstrate the underlying principles that underpin each approach in a way that allows us to infer their potentiality and appropriateness for modelling regional agricultural land-use and production dynamics, stimulating further research along this line. Generally, this thesis partly reflects the state-of-art of land-use modelling and contribute significantly to the introduction of advances in agricultural system modelling research and land-system science.

Keywords: cropland; regional agriculture; agricultural production; food security; land changes

Resumo

Os sistemas de solos agrícolas são cruciais para a segurança alimentar, para assegurar benefícios nutricionais, assim como uma variedade de serviços ambientais, culturais, sociais e econômicos. Contudo, nas próximas décadas, os sistemas de solos agrícolas enfrentarão desafios complexos, e até que ponto eles serão capazes de apoiar a segurança do sistema alimentar global será determinado pela sua eficiência, sustentabilidade e equitabilidade. Como tal, é necessário compreender melhor os fatores, padrões espaciais, e dinâmicas dos sistemas de solos agrícolas, de modo a auxiliar os processos de tomada de decisão, antecipar mudanças futuras e projetar estratégias robustas a longo prazo para lidar com os desafios iminentes.

Assim, num contexto mais amplo, em resposta aos futuros desafios do sistema alimentar e produção global, esta tese surge da relevância de compreender como a integração de conceitos, ferramentas, técnicas e métodos das Ciências da Informação Geográfica (CIG) permitem fornecer uma base de conhecimento formalizada sobre os sistemas de solos agrícolas em múltiplas escalas espaciais e temporais e diferentes contextos geográficos. Em particular, esta tese centra-se na compreensão do potencial de aplicação de abordagens baseadas em CIG e das fontes de dados espaciais disponíveis para modelar o uso do solo agrícola no distrito de Beja e as dinâmicas de produção agrícola em Portugal.

Quatro perguntas de partida associadas ao objetivo geral foram desenvolvidas. A pesquisa original conduzida foi preparada com base numa tese estruturada por artigos científicos, portanto, os objetivos específicos desta tese são abordados ao longo de sete capítulos nas Partes II a V, cada um correspondendo a um artigo científico que foi publicado ou está a ser considerado para publicação em revistas científicas internacionais revisadas por pares. Ao todo, a tese divide-se em sete capítulos, além da introdução e conclusões gerais.

A tese inicia-se com a Parte I, que inclui o Capítulo 1. Este capítulo apresenta o contexto geral de pesquisa sobre o tema da produção e segurança alimentar e a ligação com os sistemas de solos agrícolas. Além disso, o progresso no desenvolvimento e integração de modelos de sistema no campo dos sistemas agrícolas é discutido, assim como as técnicas de modelação e tecnologias geoespaciais podem ajudar as necessidades de pesquisa atuais e futuras. No fim, o capítulo detalha os objetivos, questões de pesquisa, estrutura da tese e estudos de caso.

Na Parte II, o Capítulo 2 é dedicado à revisão da literatura, procurando resumir e fornecer o contexto da contribuição desta tese no domínio científico dos sistemas de solos agrícolas. Deste modo, investigações prévias, publicadas em revistas científicas, relacionadas com solos agrícolas para produção de alimentos para consumo humano foram revisadas e sistematizadas com o objetivo de identificar e descrever os principais campos de pesquisa, direções metodológicas e fontes de dados. Além disso, este capítulo procurou vincular os principais campos de pesquisa ao cumprimento dos Objetivos de Desenvolvimento Sustentável (ODS) e, ainda, discutir a

importância dos dados espaciais, da análise espacial e da pesquisa interdisciplinar, bem como os *trade-offs* associados aos sistemas agrícolas.

Na Parte III, os Capítulos 3 e 4 exploram o uso de imagens de satélite, dados da plataforma VGI (*Volunteered Geographic Information*) OpenStreetMap (OSM), assim como abordagens multitemporais e multissensoriais e uma variedade de métodos estatísticos com o objetivo de mapear, quantificar e avaliar as dinâmicas dos solos agrícolas no distrito de Beja. Em particular, o Capítulo 3 explora uma abordagem metodológica usando a técnica de agrupamento *K-means* para refinar amostra para treino do algoritmo de classificação *Time-Weighted Dynamic Time Warping* (TWDTW) com o objetivo de dar suporte à classificação multi-temporal de longo prazo e à deteção das mudanças no uso do solo a partir de séries temporais de imagens Landsat. O Capítulo 4 avalia a potencialidade dos dados OSM no apoio do mapeamento multi-temporal do uso do solo à escala regional, bem como para fins de treino em classificações multi-temporal supervisionada.

Na Parte IV, os Capítulos 5, 6 e 7 exploram várias abordagens em CIG, tais como o modelo CA-Markov, o modelo de Cadeias de Markov, *machine learning* e modelos agnósticos, bem como um conjunto de métricas espaciais e métodos estatísticos, para modelar os fatores e as mudanças espácio-temporais do uso do solo agrícola no distrito de Beja. Em particular, o Capítulo 5 apresenta uma estrutura para avaliar o comportamento e a utilidade de um modelo baseado em autómatos celulares (CA-Markov) para simular futuras mudanças no uso do solo agrícola. O Capítulo 6 fornece uma análise estatística aprofundada com o objetivo de compreender as mudanças espácio-temporais dos solos agrícolas e projetar o desenvolvimento futuro do uso da com base no modelo de cadeias de Markov. O Capítulo 7 introduz uma abordagem quantitativa que combina *machine learning* e modelos agnósticos para avaliar os vários fatores humanos e ambientais que explicam o uso dos solos agrícolas para culturas de trigo, milho e olival.

Na Parte V, o Capítulo 8 explora várias abordagens baseadas em CIG, tais como a técnica de ponderação de área, o cubo espácio-temporal, bem como métodos estatísticos, para produzir uma longa série temporal e espacial de dados de produção agrícola ao nível regional com o objetivo de mapear, quantificar e avaliar a evolução temporal e a distribuição espacial da produção de cereais ao longo de 169 anos em Portugal.

Na Parte VI, o Capítulo 9 apresenta uma reflexão geral das principais conclusões, destaca várias implicações práticas, pontos fracos e oportunidades das várias abordagens baseadas em CIG, com recomendações para pesquisas futuras.

Do ponto de vista científico e metodológico, os estudos de caso da tese contribuem com o conhecimento prático e teórico, demonstrando os pontos fortes e as limitações de vários métodos, ferramentas e técnicas de CIG. As abordagens baseadas em CIG exploradas não pretendem ser exaustivas em termos da grande variedade de abordagens disponíveis atualmente. Em vez disso, procurou-se demonstrar os princípios subjacentes que sustentam cada abordagem de forma a

permitir inferir a sua potencialidade e adequação para modelar as dinâmicas do uso do solo agrícola e da produção, fornecendo um conhecimento formalizado sobre abordagens promissoras para análise dos sistemas de solos agrícolas e estimulando novas pesquisas nesta linha.

Os estudos de caso da tese procuraram combinar as componentes espaciais e temporais, bem como uma perspetiva de longo prazo, o que permitiu fornecer uma explicação histórica com evidências geográficas sob uma leitura cruzada do espaço-tempo. Contudo, em termos de implicações práticas, foi demonstrado que as aplicações de abordagens baseadas em CIG tendencialmente variam de acordo com os objetivos e uso pretendido, limitações no desenho do próprio modelo e disponibilidade de dados.

De modo geral, esta tese apresentou inovação teórica e prática, com diferentes soluções interdisciplinares que refletem em parte o estado da arte no campo da modelação do uso dos solos e contribui significativamente para a introdução de avanços na pesquisa de modelação dos sistemas agrícolas e da ciência dos sistemas de uso do solo. Embora a investigação tenha sido conduzida no contexto português, de um modo geral, a maior parte das constatações relativas às abordagens científicas e metodológicas podem constituir um acréscimo significativo ao estado atual do conhecimento nestes domínios científicos. Os conhecimentos obtidos contribuem para as ciências da terra em vários níveis, fornecendo informações eficazes para a monitorização e avaliação económica e ambiental, resultando em decisões políticas informadas, caminhos para um futuro mais sustentável, e no desenho de medidas antecipatórias em resposta aos desafios da segurança alimentar e desenvolvimento sustentável.

Palavras-chave: solos agrícolas; agricultura regional; produção agrícola; segurança alimentar; alterações do solo

Table of Contents

List of Figures

List of Tables

Part I. Introduction

Chapter 1. Introduction

1.1. General statement

In the 1970s, it was believed that insufficient food production was an explanatory factor for millions of people suffering from hunger in the world. However, even with a global food production increment of 138% from 1.84×10^9 t in 1961 to 4.38×10^9 t in 2007 in response to rapid population growth (from 3×10^9 in 1960 to 6.7×10^9 in 2009), millions of people continued to experience food insecurity (FAOSTAT, 2009; Tilman et al., 2011). Therefore, in the mid-1970s, it became clear that agricultural production and food security challenges were not only a matter of food quantity, but also related to other aspects such as food safety/quality, food access, and the stability of food chain supply (Barrett, 2010; Ericksen, 2008; Ericksen et al., 2009; Fan & Brzeska, 2016; Schmidhuber & Tubiello, 2007).

Since the 1996 World Food Summit, the concept of food security has evolved into a more complex definition based on four fundamental principles: the availability, accessibility, use/utilisation, and stability of foodstuffs (Ericksen, 2008). Currently, the concept of food security aggregates elements related to, for instance, food equity with distribution to match production with need, safe and healthy food, respect for the eating habits of different cultural groups, the idea of food nutritional benefits, and the environmental sustainability of technological change (Ericksen, 2008; Sobal et al., 1998). Today's food systems are not exclusively focused on how much will be produced, but who will have access, how it will be used and with what stability that same production will be achieved (FAO et al., 2013). Overall, the concept of 'fullness' in food security evolved to meet other basic needs, such as health, education, culture, and housing (Alonso et al., 2018; Burchi & De Muro, 2016).

In 2000, the Millennium Development Goals (MDGs) aimed to halve the number of hungry people by 2015. More recently, upon the expiration of the MDGs, the United Nations (UN) set 17 Sustainable Development Goals (SDGs) for the period 2015–2030, in which global food security remains a central goal. Indeed, ending hunger is the second SDG goal. As hunger has an adverse impact on human health, society, culture, and economic welfare (Marsden & Morley, 2015), reducing hunger contributes to accelerating the meeting of the other SDGs. Thus, at the very core of global initiatives, several significant adaptive management and policy approaches are being discussed and redirected towards the achievement of global food production objectives and food security for the present and future generations towards the achievement of SDG 2. Nonetheless, food security remains a major global concern, mainly fuelled by the double pressure of demographic growth and accelerated urbanisation (FAO, 2017; Tomlinson, 2013), a significant increase in consumption of and dependence on finite natural and energy resources (Popp et al., 2014; Tilman et al., 2011), climate variability (Abd-Elmabod et al., 2020; IPCC, 2014), and the broadening of food preferences and consumption (Godfray et al., 2010; Hatab et al., 2019; Kastner et al., 2012; Willett et al., 2019).

Despite the great endeavour made over the last decades to develop strategies for global food security, approximately 10% of people worldwide are suffering from severe levels of food insecurity (FAO, 2020). Moreover, the number of people worldwide suffering from chronic hungry increased between 2006 and 2016 to approximately 11% (FAO, IFAD, UNICEF, 2017). Unfortunately, the steady pace of population growth in some regions of the world means that the tendency towards food insecurity will intensify (FAO, IFAD, UNICEF, 2017). However, there are different scales of food insecurity worldwide, from poor quality food consumption to large-scale hunger levels (FAO, 2021; World Bank Group, 2015). The situation may be chronic or transient, and access may be limited within a specific period due to lack of financial and other resources. For instance, Southeast Asia and Sub-Saharan Africa are areas of the world where hunger is an alarming problem (Alexandratos & Bruinsma, 2012), although cropland area (100 million ha in Asia and 59 in Africa) (OECD/FAO, 2009) and the foreign investment (Mason-D'Croz et al., 2019) increased substantially. The challenge of food security dominating the EU and other developed countries has been related not to severe hunger but to the irregular access to nutritious and sufficient food, as well as the widespread rapid growth of obesity (FAO et al., 2020). In these regions, the great challenges in achieving food security are related to improving nutrition and reducing obesity, decreasing food waste and overconsumption, and overall environmental footprint (Allen & Prosperi, 2016; FAO et al., 2020; FAO, 2021; Willett et al., 2019; World Bank Group, 2015). Therefore, ensuring global food security is a defining challenge of the 21st century; the food and nutritional security agenda requires urgent international efforts based on effective strategies to tackle food security according to specific challenges of each region (Foley et al., 2011; Wu et al., 2014).

In particular, agricultural land systems are the principal biogeophysical source for delivering food security and ensuring nutritional benefits, as well as a variety of environmental, cultural, social, and economic services (Scown et al., 2019; Stephens et al., 2018). For instance, worldwide agricultural land derives (directly or indirectly) ~90% of food calories (Cassidy et al., 2013) and ~80% of livestock production (protein and fat) (Steinfeld et al., 2006). Considering that agricultural land systems provide the largest proportion of global food supplies, it is a triumph to guarantee future food security and achieve some of the targets of the SDGs (Avtar et al., 2020; FAO, 2017; Godfray et al., 2010; Wu et al., 2014). A main problem is that agricultural land systems constantly change over time. For instance, encroachment of agricultural land (Foley et al., 2011; Radwan et al., 2019), agricultural land abandonment (Castillo et al., 2021; MacDonald et al., 2000), agricultural land-use intensification (Lambin et al., 2000; Olesen et al., 2011), and agricultural land fragmentation (Gomes, Banos, et al., 2019; Postek et al., 2019), have an influence on global food production and food security objectives (Godfray et al., 2010; Wu et al., 2014). The current agricultural land dynamics have significant implications on the provision of food, feed, and fibre, and also, among other things, on climate change mitigation (Freibauer et al., 2004). Thus, in the coming decades, agricultural land systems will face complex challenges; the extent to which they will be able to support the security of the global food system will be determined by their efficiency, sustainability, and equitability (Kastner et al., 2012; Nelson et al., 2010). Therefore, there is a need to better understand the factors, spatial patterns, and dynamics of agricultural land systems across multiple spatial and temporal scales and geographical contexts to assist decision-making processes, anticipate future changes, and design robust strategies in the long term to manage impending challenges (FAO et al., 2020; Ruben et al., 2018; Schneider et al., 2011).

1.2. Research context

Agricultural land is a man-made system created for the purpose of producing livestock and croplands from the Earth's natural resources for human consumption (food, feed, fibre, energy) (Dillon & McConnell, 1997). An agricultural land system can be conceptualised as a complex adaptive system with a collection of interconnected components, including physical, cultural, social, behavioural, political, and economic components, all of which interact nonlinearly and dynamically in both space and time with the natural terrestrial environment (Dillon & McConnell, 1997; Verburg et al., 2015). Accordingly, elements such as spatiotemporal nonlinearity, emergence, self-organization, and self-similarity may be used to describe agricultural land systems (Antle et al., 2017; Müller et al., 2020).

Given the complex characteristics of agricultural land systems and their critical importance to human life, a growing number of scholars have studied for decades to better understand and represent this system from numerous disciplinary perspectives. A system analysis approach has registered increasing applications in agricultural systems research (Dillon & McConnell, 1997; Jones et al., 2016). This type of approach searches for generalisations based on the whole rather than on individual parts, providing better understanding of a complex system characterised by non-linear interactions among its components that produce feedback (positive or negative) that serves as the foundation for self-regulatory and emergent properties (Ackoff, 1973; Ackoff & Emery, 1972; Antle et al., 2017). However, with the size of an actual system, the inability to directly access it, or other barriers that limit this type of approach, development of a conceptual model reflecting the system components and interactions has been proposed as an appropriate framework. A model is a simplified representation or approximation of certain aspects of the structure, behaviour, functioning, or other characteristics of a real-world process, concept, or system (Jeffers, 1978; Odum & Odum, 2000). Thus, a model that describes the behaviour of a system can be used to link data and theories about a system (Rana, 2015). The logic underlying the use of models is that there is a reality external to our existence that can be captured (albeit roughly) using the principles of logic and mathematics (modelling).

Although von Thünen developed one of the first models related to agricultural land use and the importance of spatial location in 1826 (also known as location theory) (von Thünen, 1826), significant progress in the development and application of system models in the field of agricultural systems occurred in the 1950s (Jones et al., 2016). During this decade, Earl Heady published one of the first works on agricultural system modelling that attempted to maximise farm-scale decisions and evaluate the effects of public policies (Heady, 1957). Dent and Blackie (1979) published a book that included agricultural economic and biological models. The International Biological Program (IBP) encouraged scientists from different disciplines to contribute to the development and application of models in the field of agricultural systems (Worthington, 1975). In the late 1960s, Duncan et al. (1967) worked on canopy photosynthesis modelling by developing a crop-specific simulation model for different agricultural products (corn, cotton, and peanut). In 1974, the FAO developed a land evaluation framework model integrating soil, climate vegetation, and socio-economic variables (FAO, 1976). Between 1982 and 1986, crop models (CERES–Wheat, CERES–Maize) were developed in the United States to predict the production of major croplands (Boote et al., 1985; Wilkerson et al., 1983). In the 1990s, crop and economic models were developed to study the potential impact of climate change on agricultural systems (Curry et al., 1990; Rosenzweig & Parry, 1994).

Between the 1990s and the 2000s, mathematical models of crop and livestock diseases and insect pests were developed (Delgado et al., 2016; Freer et al., 1997; Herrero et al., 1996, 1999; Jones et al., 2003). The Great Plains Framework for Agricultural Resource Management (GPFARM) was developed to track croplands and livestock production for decision support (Rauff et al., 2015). Tanure et al. (2015) developed a mathematical model for farm management. (Jones et al., 2016) provided a synthesis of the history of agricultural system modelling and current and future research needs. Jones et al. (2017) reviewed the current state of agricultural system modelling, focusing on model applicability. In both studies, the authors concluded that agricultural system modelling has been promoted for different purposes across agricultural science disciplines, and that reinforcement of agricultural system modelling approaches is largely related to the development of mathematical and statistical techniques (Jones et al., 2016, 2017).

Over the last decade, there has been an increase in the number of studies using different scientific approaches and techniques from an agricultural land system perspective (Viana et al., 2022). Scholars from different disciplines have learned that mapping, monitoring, quantifying, and modelling agricultural land dynamics across multiple spatial and temporal scales and geographical contexts are practices that assist conscious and thrifty management of this natural resource, resulting in effective environmental planning and informed policy decisions for the future (Calleja et al., 2012; Cao et al., 2019; Foley et al., 2011; Kühling et al., 2016; Lambin et al., 2000; Nakalembe et al., 2017; Radwan et al., 2019; van Vliet et al., 2015; Verburg, Schot, et al., 2004; Weiss et al., 2020). The ability to map and monitor agricultural land dynamics provides spatially and temporally accurate information that contributes to effective land management and promotes proper, efficient, and rational agricultural land use (Akpoti et al., 2019; Diogo, 2018; Esgalhado et al., 2020; Piiroinen et al., 2015; Torbick et al., 2017; Viana, Santos, et al., 2021; Viana & Rocha, 2020). Likewise, evaluating the influence of different factors (human and environmental) on agricultural land production, or providing information on available and suitable land for agricultural production can contribute to identification of the best areas for crop production, establish sustainable intensification, and maximise food production (EEA, 2017; Jin et al., 2017; Shen et al., 2013; Struik & Kuyper, 2017; Wu et al., 2014). Forecasting the geographic distribution of crop yield and agricultural production in different climate scenarios can provide essential knowledge concerning consequences on agricultural systems, improving food safety and health, and strengthening resilience to climate variability (Abd-Elmabod et al., 2020; Arora, 2019; Fanzo et al., 2018; Leng & Hall, 2019; Manners & van Etten, 2018; Q. Yu et al., 2012). Modelling agricultural land changes using land-change models provides a more spatially detailed and plausible assessment of agricultural land systems, advancing our understanding of their long-term sustainability and global implications (Foley et al., 2011; Jones et al., 2016; Kelly et al., 2013; Radwan et al., 2019).

The agricultural land changes we observe today are the direct result of a combination of social, economic, and environmental factors, technological innovations, ecological and political priorities, varying at a range of spatial and temporal scales that have far-reaching implications for the environment and human well-being (Marcos-Martinez et al., 2017; Müller et al., 2020; van Vliet et al., 2015; Verburg, Ritsema van Eck, et al., 2004; Verburg, Schot, et al., 2004). The study of agricultural land changes can be supported by integrated modelling of multiple factors based on complex linkages and feedback among them, considering that the concept of change is linked to time, and changes in the temporal component are gradual (Agarwal et al., 2002; van Vliet et al., 2013; Verburg, Schot, et al., 2004; Verburg & Overmars, 2009). As a result, a plethora of land-change models based on geographic theory and reconstructions of past developments have been applied as tools to extrapolate past agricultural land changes and to predict future changes (Foley et al., 2011; Jones et al., 2016; Radwan et al., 2019). Covering the past using long-term perspectives (crossing different years, decades, or centuries) offers historical viewpoints that can provide insights to form more effective solutions for current and future food production challenges (Verburg et al., 2015; Viana, Freire, et al., 2021). Historical events can be explored as historical data, which in turn can be considered to improve both climate change and land-change models or to establish more informed baselines for different scientific disciplines (Boivin & Crowther, 2021; Winkler et al., 2021).

In a broad sense, the emergence of land-change modelling approaches has been driven by increased computing power, exponential development of geographic information science (GIS), particularly geocomputation discipline and system theories (such as complexity, self-organizing, and non-linear theories), and data collection techniques and available database software (Albrecht, 2005; Goodchild, 2009; Verburg et al., 2013; Zscheischler & Rogga, 2015). The need for a broader understanding has motivated scholars to review and categorise the growing availability of GIS-based approaches for modelling land change. One of the first reviews published in the literature was that of Baker (1989) in the context of urban ecology. Lambin (1997) reviewed land-change models that used mathematical, empirical/statistical, and spatial simulation approaches. Lambin et al. (2000) focused on agricultural intensification models and grouped them into four categories: stochastic, empirical–statistical, optimisation, and dynamic simulation. Briassoulis (2000) classified land-change models according to their methodological and functional attributes, such as statistical or econometric models, spatial interaction models, optimisation models (including linear programming, multicriteria decision-making models, and utility maximisation models), integrated models combining natural–human subsystems (gravitic, simulation, and input–output models), models based on natural sciences, models based on GIS, and models based on Markov chains. Irwin and Geoghegan (2001) compared non-economic models, many of which were included in the approaches described by Baker (1989) and Lambin (1997). They also investigated models based on the cellular automata approach and economic land-change models, which were classified as non-spatial models or spatially explicit models. Veldkamp and Lambin (2001) revised spatially explicit models; Agarwal et al. (2002) reviewed a range of models in which the dimensions of space, time, and human decision-making were present. Verburg et al. (2004) and Heistermann et al. (2006) focused their reviews on the theoretical and practical characteristics of land change models. With the diversity and nature of each model, Heistermann et al. (2006) proposed categorising land-change models into three types: geographic land-use models (CLUE, SLEUTH), economic land-use models (CAPRI, MIRAGE), and integrated models (LUMOCAP). Parker et al. (2008) focused on agent-based models as a tool to represent human decision-making and simulate possible future land use transitions. Brown et al. (2013) identified five key types of models based on machine learning, cellular models, sector-based economic models, spatially disaggregated economic models, and agent-based models. More recently, Soesbergen (2016) and Noszczyk (2019) reviewed the current state of land change modelling. Currently, models integrated or interfacing with artificial intelligence (AI), in particular machine learning, and linking to disciplines such as computing and mathematics are popular (artificial neural networks (ANN), cellular automata (CA), fuzzy logic, machine learning, and agent-based models (ABM)) (Noszczyk, 2019; Soesbergen, 2016).

Although many land-change models have been developed over the years, some technical issues related to their predictive abilities and effective applicability remain a challenge (Verburg et al., 2019). Objectively, land-change models are developed as simplifications of a complex real world; they depend on the level of complexity of the system being studied, the model parameterisation process, and the empirical data used. Thus, it is crucial to evaluate the fidelity of the model (van Vliet, 2013; van Vliet et al., 2016). Model calibration and validation are two required processes that must be evaluated separately to assess the level of confidence in the model performance against its intended application and provide information about their usability and fidelity to real-world conditions (Brown et al., 2013; Pontius Jr et al., 2008; van Vliet, 2013). Calibration is the iterative process of adjusting model parameter values to reduce divergences between the actual system and the modelling results to produce accurate outcomes; validation is the process of evaluating the quality of the produced results, assessing the acceptable range of accuracy (Chen & Pontius Jr, 2010; van Vliet, 2013).

Land-change models are mostly calibrated based on historical observations of land-use changes, whereas validation is based on land cover outcomes (van Vliet, 2013). Ideally, model validation should be applied independently; the input data used for validation must be different from the data used for calibration (Batty & Torrens, 2005; Kok et al., 2001). Although validation is an important consideration for many users, other users apply flawed criteria for validation or do not validate at all (Agarwal et al., 2002; Bennett et al., 2013; Bradley et al., 2016; Olmedo et al., 2015; van Vliet et al., 2016; Verburg et al., 2019). A proper validation assessment requires distinguishing the correct simulation due to change or due to persistence, because if during the simulation time interval there are small amounts of land-use changes, then the validation assessment will return high accuracy values even if all changes are simulated incorrectly. This is why standard Kappa statistic is not recommended for validation of simulation models, because the amount of simulated change and persistence is not included in the assessment (Pontius Jr & Millones, 2011). A proper alternative for validation assessment is to distinguish between errors caused by quantity and those caused by allocation (Olmedo et al., 2015; Pontius Jr et al., 2011). Nevertheless, several calibration and validation approaches can be adopted for land-change modelling accuracy assessments (Pontius Jr et al., 2004; Pontius Jr and Millones, 2011; van Vliet et al., 2016). For instance, another assessment method is the relative operating characteristic (ROC) (Pontius & Schneider, 2001). In addition, landscape metrics can be used to assess overall spatial pattern accuracy (Peng et al., 2010). As a result, in this framework, it is necessary to produce and update data, particularly with spatial and temporal properties, for calibration and validation of land-change models (Goodchild, 2009; Weiss et al., 2020).

Undoubtedly, the expanding capabilities of GIS-based approaches and land-change models are highly dependent on data input, resolution, and spatial extent (van Vliet, 2013; van Vliet et al., 2016). Consequently, data availability, and timely and accurate spatial data are crucial. Advances in data collection techniques, greater data processing efficiency, and expansion of public domain databases have provided important and detailed information for environmental monitoring, assessment, and sustainable use of natural resources, and political, industrial, and social development applications (Hendler, 2013; Tiropanis et al., 2014). Advances in remote sensing technology and the increasing availability of open georeferenced data collected and produced by government agencies, social media, and mobile devices have contributed to a plethora of spatial data (Yu et al., 2020). Remotely sensed data providing large-scale high-resolution land-use data are the most common data used in land-change modelling (Lu et al., 2004; Weiss et al., 2020; Wulder et al., 2012). Landsat images at 30 m spatial resolution are widely used for multi-temporal mapping, quantification, and assessing land changes for their high temporal frequency, regular coverage (almost 40 years of data records), and free availability (Alam et al., 2019; Godinho, Gil, et al., 2016; Hansen & Loveland, 2012; Viana, Girão, et al., 2019; Zhu & Woodcock, 2014). In addition, MODIS Land Cover, a global dataset available for every year from 2000 to 2012 at 500 m spatial resolution has been used for detecting land changes (Friedl et al., 2010; Lunetta et al., 2006; Xiao et al., 2005). Sentinel-2 and Sentinel-1 images at a spatial resolution of 10 m have been widely used because they represent the best available high-resolution remotely sensed data at a global scale and are freely available (Campos-Taberner et al., 2020; Immitzer et al., 2016; Phiri et al., 2020; Schulz et al., 2021).

Regional time-series datasets derived from satellite image processing and in situ data, such as the CORINE Land Cover (CLC), have been used in a variety of studies as datasets for mapping and assessing land changes (Feranec et al., 2017; Godinho, Guiomar, et al., 2016). The CLC dataset is coordinated and integrated by the European Environment Agency (EEA) and available for the 28 European Union member states and other European countries for 1990, 2000, 2006, 2012, and 2018 (Büttner & Feranec, 2002; European Union, 2021). In addition, thematic cartography produced with regularity and reliability by governmental institutions at the country level is a valuable resource providing insights for territorial planning and management, ecosystem service mapping, and the identification of areas vulnerable to rural fires (e.g., Abrantes et al. 2016, Boavida-Portugal et al. 2016, Meneses et al. 2018, Gomes et al. 2019, Viana & Rocha 2020). An example of a national dataset is the official Portuguese Land Cover Map (COS) produced by the Portuguese General Directorate for Territorial Development (DGT), which is freely available for 1995, 2007, 2010, 2015, and 2018 (DGT, 2018). Alternatively, historical datasets relying on local and regional statistics/records (census data, historical data on climate), old analogical documents (maps, cadasters, registers), and volunteered geographic information (VGI) platforms (OpenStreetMap - OSM) have been gathered and widely used as data sources (Arsanjani & Vaz, 2015; Estima & Painho, 2015; Fonte et al., 2015; Fonte et al., 2017; Viana, Encalada, et al., 2019; Viana, Freire, et al., 2021; Yang et al., 2017).

In a broad sense, advancement in GIS technology, development of more useful and efficient tools and techniques, and data collection and availability provide a formalised knowledge base concerning agricultural land systems at multiple spatial and temporal scales and geographical contexts. With the integration of concepts, methods, and tools from GIS, there is a practical opportunity for modelling agricultural land-use and production dynamics in response to climatic change or policy reform scenarios, especially when operated in a spatially explicit, integrated, and multi-scale manner. This, in turn, can provide meaningful information for long-term policy formulation, planning, management, and investment, all of which are critical for tackling global food production and security challenges.

1.3. Case studies: Beja district and mainland Portugal

Global agricultural land covers approximately 38% of the total ice-free land surface, of which approximately one-third is used for growing crops; the remaining two-thirds is used for grazing livestock (FAO, 2020). Geographically, Asia has the highest percentage of cropland area worldwide (38%), followed by the Americas (24%), Africa (18%), Europe (18%), and Oceania (2%) (FAO, 2020). The latest data from the FAO indicate significant changes in cropland area dynamics in different regions over the last two decades (FAO, 2020; Winkler et al., 2021). For instance, although agriculture is the most prominent land use in Europe (48%), between 1900 and 2010 approximately 38% of cropland was lost to other land uses, whereas only approximately 29% was gained (Fuchs et al., 2015).

The same trend has been observed in Portugal. In the westernmost country on the European continent, where the main activity was agriculture until the 1960s, substantial changes in agricultural land dynamics have occurred over the past decades (Freire & Lains, 2015, 2017; Medeiros, 2009). In 1875, agricultural land was estimated to cover approximately 21% of mainland Portugal; from 1902–1907 it increased to 35% (Medeiros, 2009). Agricultural land continued to increase with the wheat campaign, which began with 1889 legislation and the wheat policy in 1929. In 1956, agricultural land was estimated to cover approximately 54% of mainland Portugal (Medeiros, 2009). In Portugal, in addition to the natural and socioeconomic characteristics that have impacted rural areas and agricultural activities over the last century, agricultural land has almost always been conditioned to the main trends in public policies and national authorities (Ferreira, 2001). One of the policies and measures was Portugal's accession to the European Union in 1986, which marked the beginning of the application of the “Common Agricultural Policy” (CAP). Since then, the evolution of Portuguese agriculture has been inextricably linked with the evolution of the CAP, in its basic assumptions and in the instruments of practical application.

According to the official Portuguese Land Cover Map (COS) data (DGT, 2018), between 1995 and 2007, approximately 4.3% of agricultural land was lost to other land uses; until 2015, the trend has been a decrease in agricultural land. The latest data (2018) indicate that agricultural land covers 41% of mainland Portugal (Figure 1.1).

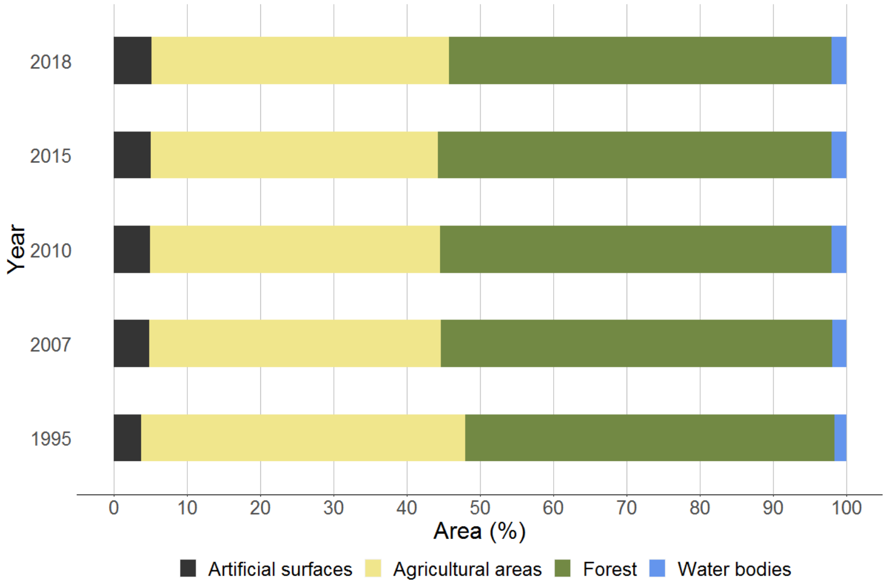


Figure 1.1. Area of each land use/cover class in each year (%) (Data source, DGT)

Although agricultural land has undergone major changes, a considerable portion of mainland Portugal is still dedicated to agricultural production. Nevertheless, the agricultural land area by type of use shows large fluctuations over the last few decades (Figure 1.2 and Table 1.1). Heterogeneous areas decreased almost 2.1% between 1995 and 2018, while arable land and pastures decreased by approximately 1.6% and 0.6%, respectively. Permanent crops increased by approximately 0.6%. According to the latest data (2018), arable land (13%) and heterogeneous areas (12%) represent the main agricultural land use, followed by permanent crops (9%) and pastures (6%).

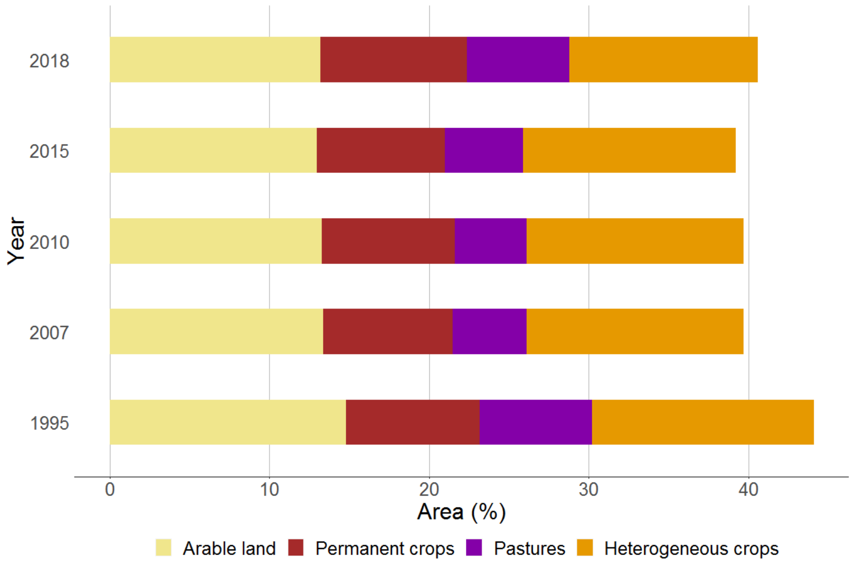


Figure 1.2. Area of each land use/cover class in each year (%) (Data source, DGT)

Table 1.1. Net changes of agricultural land classes for each time interval (%) (Data source, DGT)

LULC Class	1995-2007	2007-2010	2010-2015	2015-2018	1995-2018
Arable land	-1.32	-0.18	-0.27	0.15	-1.6
Permanent crops	-0.32	0.19	-0.34	1.24	0.8
Pastures	-2.34	-0.13	0.38	1.51	-0.6
Heterogeneous areas	-0.34	0.00	-0.29	-1.48	-2.1

The importance of agriculture in different Portuguese regions is pronounced due to edaphoclimate, soil fertility, water availability, and land structure (Figure 1.3). In the northern, central, and southern regions, different types of agricultural land have different agricultural products cultivated, with each having an important role in the primary sector. In terms of the main products, the northern region is known for vineyards in the Douro Valley and the cultivation of various cereals (wheat, barley, and corn); the southern region, specifically in Alentejo, developed an extensive monoculture of cereals, vines, and olive trees where the Mediterranean influences are more accentuated (Medeiros, 2009). Other important products include fruits, namely oranges from Algarve, cherries from the Central region, and pêra rocha from the Oeste region, and a wide variety of crops, such as green vegetables, oilseeds, nuts, and cork.

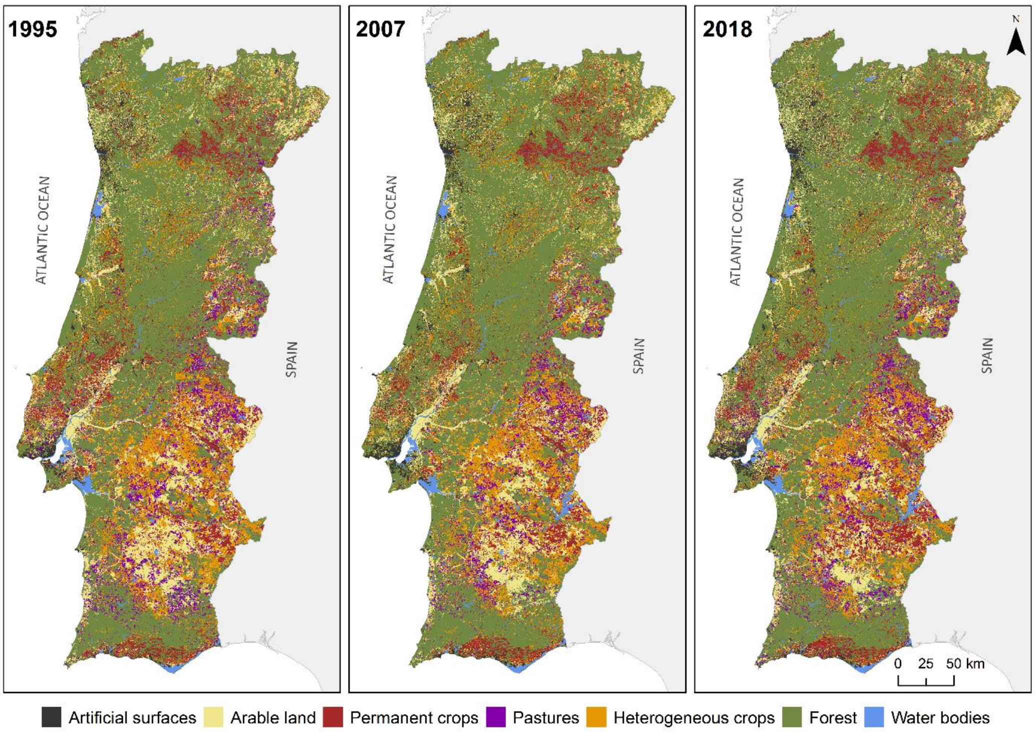


Figure 1.3. Main agricultural land uses in Portugal for 1995, 2007, and 2018

In the case of cereals, significant production changes have been observed over the last three and a half decades; 71% of cereal-growing land has been lost (ENPPC, 2018). Although cereals accounted for approximately 10% of the national territory (900 million hectares) in the late 1980s, the latest data show that this area has decreased to approximately 3% (260 million hectares), a loss of 640 million hectares (ENPPC, 2018). Regionally, a substantial portion of cereal production, including oats, wheat, and barley, is concentrated in Alentejo. One feature that distinguishes the Alentejo region from other Portuguese regions is the great support under the CAP policy for its specialisation in agricultural products such as cereals (Ferreira, 2001). In the Alentejo region, cereals represent approximately 19.1% of the regional agricultural production and more than 40% of the national cereal production (ENPPC, 2018).

Particularly, Beja, the largest Portuguese district in terms of area (constituting approximately 11% of the total area of Portugal), and the second in terms of the percentage of agricultural land, is located in Alentejo region. This district is characterised by a low urban density and the dominance of a mixed agrosilvopastoral environment with a vast landscape of intermingling cultures including wheat, olive groves, vineyards, cork oak forests, and pastures that have a high economic importance in the regional and national Portuguese agricultural industry (Muñoz-Rojas et al., 2019). In this rural region, the Montado (Montado in Portugal and Dehesa in Spain), an agrosilvopastoral agricultural heritage system (Correia, 1993), has been indicated as a globally important agricultural system according to the Globally Important Agricultural Heritage System (GIAHS) program promoted by the Food and Agriculture Organization of the United Nations (FAO) (Koohafkan & Altieri, 2016). Although agriculture remains the predominant land use in the Beja district (approximately 64% in 2018), rapid agricultural land change has been observed over the past decades. Between 1995 and 2018, arable land substantially decreased (−5.9%), along with heterogeneous areas (−2%) and pastures (-2.6%); permanent crops experienced growth in the same period (+5.13%) (Viana & Rocha, 2020). Accordingly, the dynamics of agricultural land-use and production, especially in the Beja district, have affected the rural areas, challenging food production objectives and justifying the case studies of this thesis.

1.4. Objectives, research questions, and structure

The overarching objective of this thesis is to demonstrate the potential application of GIS-based approaches and available spatial data sources for modelling regional agricultural land-use and production dynamics. To this end, this thesis focuses on the integration of concepts, tools, techniques, and methods from GIS to model the factors and spatiotemporal changes of agricultural land use, as well as the spatial distribution over time of agricultural production. The discrete-time Markov chain model, cellular automata, statistical methods, machine learning, and model-agnostic methods are examples of the GIS-based approaches considered. Remotely sensed data, agricultural statistical data, census information, environmental spatial data, VGI, and land use/cover datasets are examples of the spatial data explored. For the purpose of this thesis, the Beja district of Portugal and mainland Portugal were used as case studies.

The objective of this thesis is addressed through the following research questions (RQs):

RQ.1. What research fields, methodological directions, and data sources emerge from an agricultural land system perspective?

RQ.2. What are the available spatial data sources that can be used to map, quantify, and assess agricultural land dynamics?

RQ.3. To what extent are GIS-based approaches able to model agricultural land-use dynamics?

RQ.4. To what extent are GIS-based approaches able to model agricultural production dynamics?

The original research developed within the scope of this thesis was prepared within a research scientific article-based framework. Accordingly, this thesis has six parts and nine chapters. The specific objectives of this thesis and the RQs are addressed in Chapters 2–8, each corresponding to one scientific article. The scientific articles were either published or are being considered for publication in peer-reviewed scientific international journals. Chapters 2–8 maintain the original content of the information published in (or submitted to) scientific journals; however, the structure and formatting is slightly adapted in order to improve consistency with other chapters. For better understanding, a brief description of the research conducted, and the objectives of each chapter are provided. Table 1.2 summarises the thesis structure, objectives, methodology, and data sources.

The thesis starts with Part I, which includes Chapter 1. This chapter introduces the general thesis statement and the research context under the topic of food production and food security and the link to agricultural land systems. In addition, the progress in the development and integration of system models in the field of agricultural systems and how modelling techniques and geospatial technologies can assist current and future research needs are discussed. Finally, the chapter details the objectives, research questions, thesis structure, and case studies.

Part II introduces Chapter 2, which summarises the body of knowledge and provides the context for the contribution of this thesis within the scientific domain of agricultural land systems. This chapter presents a systematic literature review that identifies and describes the main research fields of the published research studies related to agricultural land for food grain crops (food human consumption), methodological research directions, and data sources. Additionally, this chapter links the main fields of research to the achievement of SDGs and discusses the importance of spatial data, spatial analysis, and interdisciplinary research, as well as the trade-offs associated with agricultural systems.

Parts III–V introduce six research case studies exploring the potential application of GIS-based approaches and available spatial data sources for modelling agricultural land-use and production dynamics in Portugal. Part III, which includes Chapters 3 and 4, introduces two studies that focus on understanding how remotely sensed and VGI data can be used to map, quantify, and assess agricultural land use in the Beja district. Chapter 3 integrates a set of methods and tools and uses Landsat products to refine a sample source for training the TWDTW classifier algorithm to support long-term satellite image time-series classification and multi-temporal land-use change detection. Chapter 4 explores a set of spatial metrics and evaluates the potential of OSM data to support regional and rural multi-temporal land-use mapping and provides sampling data sources to support multi-temporal land-use mapping.

Part IV, which includes Chapters 5–7, introduces three studies that focus on modelling the spatiotemporal changes and factors of agricultural land-use in the Beja district. Chapter 5 introduces a framework to evaluate a simulation of future agricultural land-use changes with an experimental application of the CA-Markov model and insightful methods of validation. Chapter 6 integrates a set of spatial metrics and statistical methods as well as the discrete-time Markov chain model to map, quantify, assess, and predict spatiotemporal changes in agricultural land use. Chapter 7 integrates an analytical framework combining machine learning and model-agnostic methods to understand the factors potentially explaining the use of agricultural land for three crop plantations.

Part V introduces Chapter 8, which focuses on modelling the spatial distribution over time of regional agricultural production in Portugal. This chapter explores a set of GIS-based approaches to produce a long spatial time-series of agricultural production data at the regional level to map, quantify, and assess the temporal evolution and spatial distribution of cereal production.

Part VI, which includes Chapter 9, provides a general conclusion; the main findings, practical implications, weaknesses, and opportunities of the research conducted in Chapters 2–8 is discussed, and recommendations for further research are outlined.

Table 1.2. Summary of thesis structure, methodological approach, sources, and status of scientific articles

Part	**Title**	**Chapter**	**Article title**	**Summary of objectives**	**RQs**	**Study area**	**Approach/ Methods**	**Main Data Source**	**Article Status**
I	Introduction	1	Not applicable	Problem statement; Research context; Case studies; Objectives, research questions, and structure	N/A	N/A	N/A	N/A	N/A
II	Agricultural land systems: state-of-art	2	Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review	Literature review	RQ 1	N/A	PRISMA protocol	N/A	Published
III	Dynamics of agricultural land: exploring remotely-sensed and VGI data	3	Long-Term Satellite Image Time-Series for Land Use/Land Cover Change Detection Using Refined Open Source Data in a Rural Region	Supports long-term satellite image time-series classification and multi-temporal land-use changes detection	RQ 2	Beja munici-pallity	K-means; TWDTW	Landsat; COS	Published
		4	The Value of OpenStreetMap Historical Contributions as a Source of Sampling Data for Multi-Temporal Land Use/Cover Maps	Supports regional and rural multi-temporal land-use mapping, and provides sampling data sources	RQ 2	Beja district	Spatial metrics	OSM; COS; CLC	Published
IV	Dynamics of agricultural land use: modelling spatiotemporal changes and factors	5	A framework to evaluate land change simulation software with an illustration of a Cellular Automata – Markov model	Investigates the implications of using a cellular automata-based model to simulate future agricultural land-use changes	RQ 3	Beja district	CA-Markov; Spatial metrics	COS	Submitted
		6	Evaluating Dominant Land Use/Land Cover Changes and Predicting Future Scenario in a Rural Region Using a Memoryless Stochastic Method	Measures the spatiotemporal agricultural land-use changes	RQ 3	Beja district	Spatial metrics; statistical methods; DTMC	COS	Published
		7	Evaluation of the factors explaining the use of agricultural land: A machine learning and model-agnostic approach	Evaluates the factors explaining the use of agricultural land	RQ 3	Beja district	Machine learning; model-agnostic methods	Statistics Portugal; IFAP	Published

Table 1.2. Cont.

Part	Title	Chapter	Article title	Summary of objectives	RQs	Study area	Approach/ Methods	Main Data Source	Article Status
V	Dynamics of agricultural production: modelling spatial distributions across time	8	Evolution of Agricultural Production in Portugal during 1850–2018: A Geographical and Historical Perspective	Describes the temporal evolution and spatial distribution of cereal production	RQ 4	Portugal mainland	Area-weighting GIS-based technique, spatiotemporal data cube	Statistics Portugal; National archives	Published
VI	General conclusion	9	N/A	Answers to the research questions; Limitations of the methods and data used; Outline of future research perspectives	N/A	N/A	N/A	N/A	N/A

References

Abd-Elmabod, S. K., Muñoz-Rojas, M., Jordán, A., Anaya-Romero, M., Phillips, J. D., Laurence, J., Zhang, Z., Pereira, P., Fleskens, L., van der Ploeg, M., & de la Rosa, D. (2020). Climate change impacts on agricultural suitability and yield reduction in a Mediterranean region. Geoderma, 374, 114453. https://doi.org/10.1016/j.geoderma.2020.114453

Abrantes, P., Fontes, I., Gomes, E., & Rocha, J. (2016). Compliance of land cover changes with municipal land use planning: Evidence from the Lisbon metropolitan region (1990–2007). Land Use Policy, 51, 120–134. https://doi.org/10.1016/J.LANDUSEPOL.2015.10.023

Ackoff, R. L. (1973). Science in the Systems Age: Beyond IE, OR, and MS. Operations Research, 21(3), 661–671.

Ackoff, R. L., & Emery, F. E. (1972). On Purposeful Systems. Tavistock.

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P., & Schweik, C. M. (2002). A Review and Assessment of Land-Use Change Models: Dynamics of Space, Time, and Human Choice. Apollo The International Magazine Of Art And Antiques, 62. https://doi.org/10.1289/ehp.6514

Akpoti, K., Kabo-bah, A. T., & Zwart, S. J. (2019). Agricultural land suitability analysis: State-of-the-art and outlooks for integration of climate change analysis. In Agricultural Systems (Vol. 173, pp. 172–208). Elsevier Ltd. https://doi.org/10.1016/j.agsy.2019.02.013

Alam, A., Bhat, M. S., & Maheen, M. (2019). Using Landsat satellite data for assessing the land use and land cover change in Kashmir valley. GeoJournal 2019 85:6, 85(6), 1529–1543. https://doi.org/10.1007/S10708-019-10037-X

Albrecht, J. (2005). A New Age for Geosimulation. Transactions in GIS, 9(4), 451–454. https://doi.org/10.1111/J.1467-9671.2005.00228.X

Alexandratos, N., & Bruinsma, J. (2012). World agriculture towards 2030/2050: the 2012 revision. (ESA Working Paper No. 12-03).

Allen, T., & Prosperi, P. (2016). Modeling Sustainable Food Systems. Environmental Management, 57(5), 956–975. https://doi.org/10.1007/s00267-016-0664-8

Alonso, E. B., Cockx, L., & Swinnen, J. (2018). Culture and food security. Global Food Security, 17, 113–127. https://doi.org/10.1016/J.GFS.2018.02.002

Antle, J. M., Basso, B., Conant, R. T., Godfray, H. C. J., Jones, J. W., Herrero, M., Howitt, R. E., Keating, B. A., Munoz-Carpena, R., Rosenzweig, C., Tittonell, P., & Wheeler, T. R. (2017). Towards a new generation of agricultural system data, models and knowledge products: Design and improvement. Agricultural Systems, 155, 255–268. https://doi.org/10.1016/j.agsy.2016.10.002

Arora, N. K. (2019). Impact of climate change on agriculture production and its sustainable solutions. Environmental Sustainability, 2(2), 95–96. https://doi.org/10.1007/s42398-019-00078-w

Arsanjani, Jamal Jokar, & Vaz, E. (2015). An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises. International Journal of Applied Earth Observation and Geoinformation, 35, 329–337. https://doi.org/10.1016/J.JAG.2014.09.009

Avtar, R., Aggarwal, R., Kharrazi, A., Kumar, P., & Kurniawan, T. A. (2020). Utilizing geospatial information to implement SDGs and monitor their Progress. In Environmental Monitoring and Assessment (Vol. 192, Issue 1, pp. 1–21). Springer. https://doi.org/10.1007/s10661-019-7996-9

Baker, W. L. (1989). A review of models of landscape change. Landscape Ecology, SPB Academic Publishing, The Hague, Landscape Ecology Vol. 2 n.o 2, 111-133.

Barrett, C. B. (2010). Measuring Food Insecurity. Science, 327(5967), 825–828. https://doi.org/10.1126/SCIENCE.1182768

Batty, M., & Torrens, P. M. (2005). Modelling and prediction in a complex world. Futures, 37, 745–766. https://doi.org/10.1016/j.futures.2004.11.003

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., & Andreassian, V. (2013). Characterising performance of environmental models. Environmental Modelling & Software, 40, 1–20. https://doi.org/10.1016/J.ENVSOFT.2012.09.011

Boavida-Portugal, I., Rocha, J., & Ferreira, C. C. (2016). Exploring the impacts of future tourism development on land use/cover changes. Applied Geography, 77, 82–91. https://doi.org/10.1016/j.apgeog.2016.10.009

Boivin, N., & Crowther, A. (2021). Mobilizing the past to shape a better Anthropocene. Nature Ecology & Evolution 2021, 1–12. https://doi.org/10.1038/s41559-020-01361-4

Boote, K. J., Jones, J. W., Mishoe, J. W., & Wilkerson, G. G. (1985). Modeling growth and yield of groundnut. Agrometeorology of Groundnut: Proceedings of an International Symposium, 243–254.

Bradley, A. V., Rosa, I. M. D., Pontius, R. G., Ahmed, S. E., Araújo, M. B., Brown, D. G., Brandão, A., Câmara, G., Carnerio, T. G. S., Hartley, A. J., Smith, M. J., & Ewers, R. M. (2016). SimiVal, a multi-criteria map comparison tool for land-change model projections. Environmental Modelling & Software, 82, 229–240. https://doi.org/10.1016/J.ENVSOFT.2016.04.016

Briassoulis, H. (2000). Analysis of Land Use Change: Theoretical and Modeling Approaches. In R. (2019) Loveridge, S. (1999), Jackson (Ed.), The web book of regional science (3rd ed.). Regional Research Institute, West Virginia University.

Brown, D. G., Verburg, P. H., Pontius, R. G., & Lange, M. D. (2013). Opportunities to improve impact, integration, and evaluation of land change models. Current Opinion in Environmental Sustainability, 5(5), 452–457. https://doi.org/10.1016/J.COSUST.2013.07.012

Burchi, F., & De Muro, P. (2016). From food availability to nutritional capabilities: Advancing food security analysis. Food Policy, 60, 10–19. https://doi.org/10.1016/J.FOODPOL.2015.03.008

Büttner, G., & Feranec, J. (2002). The CORINE Land Cover upddate 2000. Techinical guidelines. EEA Technical Report, 89.

Calleja, E. J., Ilbery, B., & Mills, P. R. (2012). Agricultural change and the rise of the British strawberry industry, 1920-2009. Journal of Rural Studies, 28(4), 603–611. https://doi.org/10.1016/j.jrurstud.2012.07.005

Campos-Taberner, M., García-Haro, F. J., Martínez, B., Izquierdo-Verdiguier, E., Atzberger, C., Camps-Valls, G., & Gilabert, M. A. (2020). Understanding deep learning in land use classification based on Sentinel-2 time series. Scientific Reports 2020 10:1, 10(1), 1–12. https://doi.org/10.1038/s41598-020-74215-5

Cao, M., Zhu, Y., Lü, G., Chen, M., & Qiao, W. (2019). Spatial Distribution of Global Cultivated Land and Its Variation between 2000 and 2010, from Both Agro-Ecological and Geopolitical Perspectives. Sustainability, 11(5), 1242. https://doi.org/10.3390/su11051242

Cassidy, E. S., West, P. C., Gerber, J. S., & Foley, J. A. (2013). Redefining agricultural yields: from tonnes to people nourished per hectare. Environmental Research Letters, 8(3), 034015. https://doi.org/10.1088/1748-9326/8/3/034015

Castillo, C. P., Jacobs-Crisioni, C., Diogo, V., & Lavalle, C. (2021). Modelling agricultural land abandonment in a fine spatial resolution multi-level land-use model: An application for the EU. Environmental Modelling and Software, 136, 104946. https://doi.org/10.1016/j.envsoft.2020.104946

Chen, H., & Pontius Jr, R. G. (2010). Diagnostic tools to evaluate a spatial land change projection along a gradient of an explanatory variable. Landscape Ecology, 25(9), 1319–1331. https://doi.org/10.1007/s10980-010-9519-5

Correia, T. P. (1993). Threatened landscape in Alentejo, Portugal: the 'montado' and other 'agro-silvo-pastoral' systems. Landscape and Urban Planning, 24(1–4), 43–48. https://doi.org/10.1016/0169-2046(93)90081-N

Curry, R. B., Peart, R. M., Jones, J. W., Boote, K. J., Allen, L. H., & Jr. (1990). Response of crop yield to predicted changes in climate and atmospheric CO2 using simulation. Transactions of the ASAE, 33(4), 1383–1390. https://doi.org/10.13031/2013.31484

Delgado, C., Rosegrant, M., Steinfeld, H., Ehui, S., & Courbois, C. (2016). Livestock to 2020: The Next Food Revolution: Http://Dx.Doi.Org/10.5367/000000001101293427, 30(1), 27–29. https://doi.org/10.5367/000000001101293427

Dent, J. B., & Blackie, M. J. (1979). Systems Simulation in Agriculture. In Applied Science Publishers.

DGT. (2018). Especificações Técnicas da Carta de Uso e Ocupação do Solo (COS) de Portugal Continental para 1995, 2007, 2010 e 2015.

Dillon, J. L., & McConnell, D. J. (1997). Farm management for Asia: a systems approach. FAO Farm Systems Management Series - 13.

Diogo, V. (2018). Agricultural land systems: Explaining and simulating agricultural land-use patterns. Vrije Universiteit Amsterdam.

Duncan, W. G., Loomis, R. S., Williams, W. A., & Hanau, R. (1967). A model for simulating photosynthesis in plant communities. Hilgardia, 38(4), 181–205. https://doi.org/10.3733/HILG.V38N04P181

EEA. (2017). Food in a green light - A systems approach to sustainable food.

ENPPC. (2018). Estratégia Nacional para a Promoção da Produção de Cereais.

Ericksen, P. J. (2008). Conceptualizing food systems for global environmental change research. Global Environmental Change, 18(1), 234–245. https://doi.org/10.1016/j.gloenvcha.2007.09.002

Ericksen, P. J., Ingram, J. S. I., & Liverman, D. M. (2009). Food security and global environmental change: emerging challenges. In Environmental Science and Policy (Vol. 12, Issue 4, pp. 373–377). Elsevier. https://doi.org/10.1016/j.envsci.2009.04.007

Esgalhado, C., Guimarães, H., Debolini, M., Guiomar, N., Lardon, S., & Ferraz de Oliveira, I. (2020). A holistic approach to land system dynamics – The Monfurado case in Alentejo, Portugal. Land Use Policy, 95, 104607. https://doi.org/10.1016/J.LANDUSEPOL.2020.104607

Estima, J., & Painho, M. (2015). Investigating the Potential of OpenStreetMap for Land Use/Land Cover Production: A Case Study for Continental Portugal. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience (pp. 273–293). Springer, Cham. https://doi.org/10.1007/978-3-319-14280-7

European Union. (2021). Copernicus Land Monitoring Service.

Fan, S., & Brzeska, J. (2016). Sustainable food security and nutrition: Demystifying conventional beliefs. Global Food Security, 11, 11–16. https://doi.org/10.1016/J.GFS.2016.03.005

Fanzo, J., Davis, C., McLaren, R., & Choufani, J. (2018). The effect of climate change across food systems: Implications for nutrition outcomes. In Global Food Security (Vol. 18, pp. 12–19). Elsevier B.V. https://doi.org/10.1016/j.gfs.2018.06.001

FAO, IFAD, UNICEF, W. and W. (2017). The State of Food Security and Nutrition in the World 2017. Building resilience for peace and food security.

FAO. (1976). A framework for land evaluation. Soils bulletin 32.

FAO. (2017). The Future of Food and Agriculture - Trends and Challenges.

FAO. (2020). World Food and Agriculture - Statistical Yearbook 2020. https://doi.org/https://doi.org/10.4060/cb1329en

FAO. (2021). Crop Prospects and Food Situation - Quarterly Global Report No. 3.

FAO, IFAD, UNICEF, WFP, & WHO. (2020). The State of Food Security and Nutrition in the World 2020. Transforming food systems for affordable healthy diets.

FAO, IFAD, & WPF. (2013). The state of food insecurity in the world 2013. The Multiple Dimensions of Food Security.

FAOSTAT. (2009). Food and agricultural commodities production.

Feranec, J., Soukup, T., Taff, G., Stych, P., & Bicik, I. (2017). Overview of changes in land cover and land use in Eastern Europe. In G. Gutman & V. Radeloff (Eds.), Land-cover and land-use changes in Eastern Europe after the collapse of the Soviet Union in 1991 (Springer).

Ferreira, D. de B. (2001). Evolução Da Paisagem De Montado No Alentejo Interior Ao Longo Do Século XX: Dinâmica e incidências ambientais. Finisterra, 36(72), 179–193.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., … Zaks, D. P. M. (2011). Solutions for a cultivated planet. Nature, 478(7369), 337–342. https://doi.org/10.1038/nature10452

Fonte, Cidália C., Bastin, L., See, L., Foody, G., & Lupia, F. F. (2015). Usability of VGI for validation of land cover maps. International Journal of Geographical Information Science, 29(7), 1269–1291. https://doi.org/10.1080/13658816.2015.1018266

Fonte, Cidália Costa, Patriarca, J. A., Minghini, M., Antoniou, V., See, L., & Brovelli, M. A. (2017). Using OpenStreetMap to Create Land Use and Land Cover Maps: Development of an Application. In C. E. C. Campelo, M. Bertolotto, & P. Corcoran (Eds.), Volunteered Geographic Information and the Future of Geospatial Data: Vol. i (pp. 113–137). IGI Global. https://doi.org/10.4018/978-1-5225-2446-5

Freer, M., Moore, A. D., & Donnelly, J. R. (1997). GRAZPLAN: decision support systems for Australian grazing enterprises. I. Overview of the GRAZPLAN project and a description of the MetAccess and LambAlive DSS. Agricultural Systems, 54, 57–76.

Freibauer, A., Rounsevell, M. D. A., Smith, P., & Verhagen, J. (2004). Carbon sequestration in the agricultural soils of Europe. Geoderma, 122(1), 1–23. https://doi.org/10.1016/J.GEODERMA.2004.01.021

Freire, D., & Lains, P. (2015). An Agrarian History of Portugal (year 1000 to 2000). Leiden&Boston: Brill (no prelo).

Freire, D., & Lains, P. (2017). An agrarian history of Portugal, 1000-2000: economic development on the European frontier (D. Freire & P. Lains (eds.)). Brill.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sensing of Environment, 114(1), 168–182. https://doi.org/10.1016/J.RSE.2009.08.016

Fuchs, R., Herold, M., Verburg, P. H., Clevers, J. G. P. W., & Eberle, J. (2015). Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010. Global Change Biology, 21(1), 299–313. https://doi.org/10.1111/gcb.12714

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. In Science (Vol. 327, Issue 5967, pp. 812–818). American Association for the Advancement of Science. https://doi.org/10.1126/science.1185383

Godinho, S., Gil, A., Guiomar, N., Neves, N., & Pinto-Correia, T. (2016). A remote sensing-based approach to estimating montado canopy density using the FCD model: a contribution to identifying HNV farmlands in southern Portugal. Agroforestry Systems, 90(1), 23–34. https://doi.org/10.1007/s10457-014-9769-3

Godinho, S., Guiomar, N., Machado, R., Santos, P., Sá-Sousa, P., Fernandes, J. P., Neves, N., & Pinto-Correia, T. (2016). Assessment of environment, land management, and spatial variables on recent changes in montado land cover in southern Portugal. Agroforestry Systems, 90(1), 177–192. https://doi.org/10.1007/s10457-014-9757-7

Gomes, E., Abrantes, P., Banos, A., Rocha, J., & Buxton, M. (2019). Farming under urban pressure: Farmers' land use and land cover change intentions. Applied Geography, 102, 58–70. https://doi.org/10.1016/j.apgeog.2018.12.009

Gomes, E., Banos, A., Abrantes, P., Rocha, J., Kristensen, S. B. P., & Busck, A. (2019). Agricultural land fragmentation analysis in a peri-urban context: From the past into the future. Ecological Indicators, 97, 380–388. https://doi.org/10.1016/j.ecolind.2018.10.025

Goodchild, M. F. (2009). Geographic information systems and science: today and tomorrow. Procedia Earth and Planetary Science, 1(1), 1037–1043. https://doi.org/10.1016/j.proeps.2009.09.160

Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. Remote Sensing of Environment, 122, 66–74. https://doi.org/10.1016/j.rse.2011.08.024

Hatab, A. A., Cavinato, M. E. R., & Lagerkvist, C. J. (2019). Urbanization, livestock systems and food security in developing countries: A systematic review of the literature. Food Security, 11(2), 279–299. https://doi.org/10.1007/s12571-019-00906-1

Heady, E. O. (1957). An Econometric Investigation of the Technology of Agricultural Production Functions. Econometrica, 25(2), 249. https://doi.org/10.2307/1910253

Heistermann, M., Müller, C., & Ronneberger, K. (2006). Land in sight?: Achievements, deficits and potentials of continental to global scale land-use modeling. Agriculture, Ecosystems & Environment, 114(2–4), 141–158. https://doi.org/10.1016/J.AGEE.2005.11.015

Hendler, J. (2013). Broad Data: Exploring the Emerging Web of Data. Big Data, 1(1), 18–20. https://doi.org/10.1089/BIG.2013.1506

Herrero, M., Fawcett, R. H., & Dent, J. B. (1996). Integrating simulation models to optimise nutrition and management for dairy farms: a methodology. In J. B. Dent & et al. (Eds.), Livestock Farming Systems: Research, Development, Socio-economics and the Land Manager, European Association for Animal Production Publication No. 79 (79th ed., pp. 322–326). Wageiningen Pers.

Herrero, M., Fawcett, R. H., & Dent, J. B. (1999). Bio-economic evaluation of dairy farm management scenarios using integrated simulation and multiple-criteria models. Agricultural Systems, 62, 149–168.

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First experience with Sentinel-2 data for crop and tree species classifications in central Europe. Remote Sensing, 8(3). https://doi.org/10.3390/rs8030166

IPCC. (2014). Food security and food production systems. In J. R. Porter, L. Xie, A. J. Challinor, K. Cochrane, S. M. Howden, M. M. Iqbal, D. B. Lobell, & M. I. Travasso (Eds.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., Barros, V.R., Dokken, D.J., Ma.

Irwin, E. G., & Geoghegan, J. (2001). Theory, data, methods: developing spatially explicit economic models of land use change. Agriculture, Ecosystems & Environment, 85(1–3), 7–24. https://doi.org/10.1016/S0167-8809(01)00200-6

Jeffers, J. N. R. (1978). An introduction to system analysis - with ecological applications. Arnold.

Jin, X., Shao, Y., Zhang, Z., Resler, L. M., Campbell, J. B., Chen, G., & Zhou, Y. (2017). The evaluation of land consolidation policy in improving agricultural productivity in China. Scientific Reports, 7(2792).

Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2016). Brief history of agricultural systems modeling. Agricultural Systems, 155, 240–254. https://doi.org/10.1016/j.agsy.2016.05.014

Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017). Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. Agricultural Systems, 155, 269–288. https://doi.org/10.1016/j.agsy.2016.09.021

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT cropping system model. European Journal of Agronomy, 18(3–4), 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7

Kastner, T., Rivas, M. J. I., Koch, W., & Nonhebel, S. (2012). Global changes in diets and the consequences for land requirements for food. Proceedings of the National Academy of Sciences of the United States of America, 109(18), 6868–6872. https://doi.org/10.1073/pnas.1117054109

Kelly, R. A., Jakeman, A. J., Barreteau, O., Borsuk, M. E., ElSawah, S., Hamilton, S. H., Henriksen, H. J., Kuikka, S., Maier, H. R., Rizzoli, A. E., van Delden, H., & Voinov, A. A. (2013). Selecting among five common modelling approaches for integrated environmental assessment and management. Environmental Modelling & Software, 47, 159–181. https://doi.org/10.1016/J.ENVSOFT.2013.05.005

Kok, K., Farrow, A., Veldkamp, A., & Verburg, P. H. (2001). A method and application of multi-scale validation in spatial land use models. Agriculture, Ecosystems & Environment, 85(1–3), 223–238. https://doi.org/10.1016/S0167-8809(01)00186-4

Koohafkan, P. (Parviz), & Altieri, M. A. (2016). Forgotten agricultural heritage : reconnecting food systems and sustainable development.

Kühling, I., Broll, G., & Trautz, D. (2016). Spatio-temporal analysis of agricultural land-use intensity across the Western Siberian grain belt. Science of the Total Environment, 544, 271–280. https://doi.org/10.1016/j.scitotenv.2015.11.129

Lambin, E. F. (1997). Modelling and monitoring land-cover change processes in tropical regions. Progress in Physical Geography, 21(3), 375–393. https://doi.org/10.1177/030913339702100303

Lambin, E. F., Rounsevell, M. D. A., & Geist, H. J. (2000). Are agricultural land-use models able to predict changes in land-use intensity? Agriculture, Ecosystems and Environment, 82(1–3), 321–331. https://doi.org/10.1016/S0167-8809(00)00235-8

Leng, G., & Hall, J. (2019). Crop yield sensitivity of global major agricultural countries to droughts and the projected changes in the future. Science of the Total Environment, 654, 811–821. https://doi.org/10.1016/j.scitotenv.2018.10.434

Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004). Change detection techniques. International Journal of Remote Sensing, 25(12), 2365–2401. https://doi.org/10.1080/0143116031000139863

Lunetta, R. S., Knight, J. F., Ediriwickrema, J., Lyon, J. G., & Worthy, L. D. (2006). Land-cover change detection using multi-temporal MODIS NDVI data. Remote Sensing of Environment, 105(2), 142–154. https://doi.org/10.1016/j.rse.2006.06.018

MacDonald, D., Crabtree, J. R., Wiesinger, G., Dax, T., Stamou, N., Fleury, P., Gutierrez Lazpita, J., & Gibon, A. (2000). Agricultural abandonment in mountain areas of Europe: Environmental consequences and policy response. Journal of Environmental Management, 59(1), 47–69. https://doi.org/10.1006/JEMA.1999.0335

Manners, R., & van Etten, J. (2018). Are agricultural researchers working on the right crops to enable food and nutrition security under future climates? Global Environmental Change, 53, 182–194. https://doi.org/10.1016/j.gloenvcha.2018.09.010

Marcos-Martinez, R., Bryan, B. A., Connor, J. D., & King, D. (2017). Agricultural land-use dynamics: Assessing the relative importance of socioeconomic and biophysical drivers for more targeted policy. Land Use Policy, 63, 53–66. https://doi.org/10.1016/j.landusepol.2017.01.011

Marsden, T., & Morley, A. (2015). Sustainable Food Systems: Building a New Paradigm (T. Marsden & A. Morley (eds.)). Routledge.

Mason-D'Croz, D., Sulser, T. B., Wiebe, K., Rosegrant, M. W., Lowder, S. K., Nin-Pratt, A., Willenbockel, D., Robinson, S., Zhu, T., Cenacchi, N., Dunston, S., & Robertson, R. D. (2019). Agricultural investments and hunger in Africa modeling potential contributions to SDG2 – Zero Hunger. World Development, 116, 38–53. https://doi.org/10.1016/j.worlddev.2018.12.006

Medeiros, C. A. (2009). Geografia de Portugal : ambiente natural e ocupação humana : uma introdução. Editorial Estampa.

Meneses, B. M., Reis, E., Vale, M. J., & Reis, R. (2018). Modelling the Land Use and Land cover changes in Portugal: A multi-scale and multi-temporal approach. Finisterra, 53(107). https://doi.org/10.18055/Finis12258

Müller, B., Hoffmann, F., Heckelei, T., Müller, C., Hertel, T. W., Polhill, J. G., van Wijk, M., Achterbosch, T., Alexander, P., Brown, C., Kreuer, D., Ewert, F., Ge, J., Millington, J. D. A., Seppelt, R., Verburg, P. H., & Webber, H. (2020). Modelling food security: Bridging the gap between the micro and the macro scale. Global Environmental Change, 63, 102085. https://doi.org/10.1016/j.gloenvcha.2020.102085

Muñoz-Rojas, J., Pinto-Correia, T., Hvarregaard Thorsoe, M., & Noe, E. (2019). The Portuguese Montado : A Complex System under Tension between Different Land Use Management Paradigms . In Silvicultures - Management and Conservation. IntechOpen. https://doi.org/10.5772/intechopen.86102

Nakalembe, C., Dempewolf, J., & Justice, C. (2017). Agricultural land use change in Karamoja Region, Uganda. Land Use Policy, 62, 2–12. https://doi.org/10.1016/j.landusepol.2016.11.029

Nelson, G. C., Rosegrant, M. W., Palazzo, A., Gray, I., Ingersoll, C., Robertson, R. D., Tokgoz, S., Zhu, T., Sulser, T. B., Ringler, C., Msangi, S., & You, L. (2010). Food Security, Farming, and Climate Change to 2050: Scenarios, Results, Policy Options. In Food Security, Farming, and Climate Change to 2050: Scenarios, Results, Policy Options. International Food Policy Research Institute. https://doi.org/10.2499/9780896291867

Noszczyk, T. (2019). A review of approaches to land use changes modeling. Human and Ecological Risk Assessment: An International Journal, 25(6), 1377–1405. https://doi.org/10.1080/10807039.2018.1468994

Odum, H. T., & Odum, E. C. (2000). Modeling for all Scales: An introduction to System Simulation (A. Press (ed.)).

OECD/FAO. (2009). OECD-FAO Agricultural Outlook 2009-2018. OECD Publishing and FAO.

Olesen, J. E., Trnka, M., Kersebaum, K. C., Skjelvåg, A. O., Seguin, B., Peltonen-Sainio, P., Rossi, F., Kozyra, J., & Micale, F. (2011). Impacts and adaptation of European crop production systems to climate change. European Journal of Agronomy, 34(2), 96–112. https://doi.org/10.1016/J.EJA.2010.11.003

Olmedo, M. T. C., Pontius Jr, R. G., Paegelow, M., & Mas, J. F. (2015). Comparison of simulation models in terms of quantity and allocation of land change. Environmental Modelling and Software, 69, 214–221. https://doi.org/10.1016/j.envsoft.2015.03.003

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2008). Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review. Http://Dx.Doi.Org/10.1111/1467-8306.9302004, 93(2), 314–337. https://doi.org/10.1111/1467-8306.9302004

Peng, J., Wang, Y., Zhang, Y., Wu, J., Li, W., & Li, Y. (2010). Evaluating the effectiveness of landscape metrics in quantifying spatial patterns. Ecological Indicators, 10(2), 217–223. https://doi.org/10.1016/J.ECOLIND.2009.04.017

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. Remote Sensing 2020, Vol. 12, Page 2291, 12(14), 2291. https://doi.org/10.3390/RS12142291

Piiroinen, R., Heiskanen, J., Mõttus, M., & Pellikka, P. (2015). Classification of crops across heterogeneous agricultural landscape in Kenya using AisaEAGLE imaging spectroscopy data. International Journal of Applied Earth Observation and Geoinformation, 39, 1–8. https://doi.org/10.1016/j.jag.2015.02.005

Pontius Jr, R. G., Boersma, W., Castella, J. C., Clarke, K., Nijs, T., Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C. D., McConnell, W., Mohd Sood, A., Pijanowski, B., Pithadia, S., Sweeney, S., Trung, T. N., Veldkamp, A. T., & Verburg, P. H. (2008). Comparing the input, output, and validation maps for several models of land change. Annals of Regional Science, 42(1), 11–37. https://doi.org/10.1007/s00168-007-0138-2

Pontius Jr, R. G., Huffaker, D., & Denman, K. (2004). Useful techniques of validation for spatially explicit land-change models. Ecological Modelling, 179(4), 445–461. https://doi.org/10.1016/J.ECOLMODEL.2004.05.010

Pontius Jr, R. G., & Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing, 32(15), 4407–4429. https://doi.org/10.1080/01431161.2011.552923

Pontius Jr, R. G., Peethambaram, S., & Castella, J.-C. (2011). Comparison of Three Maps at Multiple Resolutions: A Case Study of Land Change Simulation in Cho Don District, Vietnam. Annals of the Association of American Geographers, 101(1), 45–62. https://doi.org/10.1080/00045608.2010.517742

Pontius, R. G., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems & Environment, 85(1–3), 239–248. https://doi.org/10.1016/S0167-8809(01)00187-6

Popp, J., Lakner, Z., Harangi-Rákos, M., & Fári, M. (2014). The effect of bioenergy expansion: Food, energy, and environment. Renewable and Sustainable Energy Reviews, 32, 559–578. https://doi.org/10.1016/j.rser.2014.01.056

Postek, P., Leń, P., & Stręk, Ż. (2019). The proposed indicator of fragmentation of agricultural land. Ecological Indicators, 103, 581–588. https://doi.org/10.1016/j.ecolind.2019.04.023

Radwan, T. M., Blackburn, G. A., Whyatt, J. D., & Atkinson, P. M. (2019). Dramatic Loss of Agricultural Land Due to Urban Expansion Threatens Food Security in the Nile Delta, Egypt. Remote Sensing, 11(3), 332. https://doi.org/10.3390/rs11030332

Rana, L. (2015). Models, theory & systems analysis in geography. In The Association for Geographical Studies.

Rauff, K. O., Bello, R., Rauff, K. O., & Bello, R. (2015). A Review of Crop Growth Simulation Models as Tools for Agricultural Meteorology. Agricultural Sciences, 6(9), 1098–1105. https://doi.org/10.4236/AS.2015.69105

Rosenzweig, C., & Parry, M. L. (1994). Potential impact of climate change on world food supply. Nature 1994 367:6459, 367(6459), 133–138. https://doi.org/10.1038/367133a0

Ruben, R., Verhagen, J., & Plaisier, C. (2018). The Challenge of Food Systems Research: What Difference Does It Make? Sustainability, 11(1), 171. https://doi.org/10.3390/su11010171

Schmidhuber, J., & Tubiello, F. N. (2007). Global food security under climate change. In Proceedings of the National Academy of Sciences of the United States of America (Vol. 104, Issue 50, pp. 19703–19708). National Academy of Sciences. https://doi.org/10.1073/pnas.0701976104

Schneider, U. A., Havlík, P., Schmid, E., Valin, H., Mosnier, A., Obersteiner, M., Böttcher, H., Skalský, R., Balkovič, J., Sauer, T., & Fritz, S. (2011). Impacts of population growth, economic development, and technical change on global food production and consumption. Agricultural Systems, 104(2), 204–215. https://doi.org/10.1016/j.agsy.2010.11.003

Schulz, D., Yin, H., Tischbein, B., Verleysdonk, S., Adamou, R., & Kumar, N. (2021). Land use mapping using Sentinel-1 and Sentinel-2 time series in a heterogeneous landscape in Niger, Sahel.

ISPRS Journal of Photogrammetry and Remote Sensing, 178, 97–111. https://doi.org/10.1016/J.ISPRSJPRS.2021.06.005

Scown, M. W., Winkler, K. J., & Nicholas, K. A. (2019). Aligning research with policy and practice for sustainable agricultural land systems in Europe. Proceedings of the National Academy of Sciences of the United States of America, 116(11), 4911–4916. https://doi.org/10.1073/pnas.1812100116

Shen, J., Cui, Z., Miao, Y., Mi, G., Zhang, H., Fan, M., Zhang, C., Jiang, R., Zhang, W., Li, H., Chen, X., Li, X., & Zhang, F. (2013). Transforming agriculture in China: From solely high yield to both high yield and high resource use efficiency. In Global Food Security (Vol. 2, Issue 1, pp. 1–8). Elsevier. https://doi.org/10.1016/j.gfs.2012.12.004

Sobal, J., Khan, L. K., & Bisogni, C. (1998). A conceptual model of the food and nutrition system. Social Science and Medicine, 47(7), 853–863. https://doi.org/10.1016/S0277-9536(98)00104-X

Soesbergen, A. van. (2016). A review of land-use change models. United Nations Environment Programme.

Steinfeld, H., Gerber, P., Wassenaar, T., Castel, V., Rosales, M., & de Haan, C. (2006). Livestock's Long Shadow: Environmental Issues and Options. Food and Agriculture Organization of the United Nations (FAO).

Stephens, E. C., Jones, A. D., & Parsons, D. (2018). Agricultural systems research and global food security in the 21st century: An overview and roadmap for future opportunities. Agricultural Systems, 163, 1–6. https://doi.org/10.1016/j.agsy.2017.01.011

Struik, P. C., & Kuyper, T. W. (2017). Sustainable intensification in agriculture: the richer shade of green. A review. In Agronomy for Sustainable Development (Vol. 37, Issue 5, pp. 1–15). Springer-Verlag France. https://doi.org/10.1007/s13593-017-0445-7

Tanure, S., Nabinger, C., & Becker, J. L. (2015). Bioeconomic Model of Decision Support System for Farm Management: Proposal of a Mathematical Model. Systems Research and Behavioral Science, 32(6), 658–671. https://doi.org/10.1002/SRES.2252

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences of the United States of America, 108(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

Tiropanis, T., Hall, W., Hendler, J., & De Larrinaga, C. (2014). The Web Observatory: A Middle Layer for Broad Data. Big Data, 2(3), 129–130. https://doi.org/10.1089/BIG.2014.0035

Tomlinson, I. (2013). Doubling food production to feed the 9 billion: A critical perspective on a key discourse of food security in the UK. Journal of Rural Studies, 29, 81–90. https://doi.org/10.1016/j.jrurstud.2011.09.001

Torbick, N., Chowdhury, D., Salas, W., & Qi, J. (2017). Monitoring Rice Agriculture across Myanmar Using Time Series Sentinel-1 Assisted by Landsat-8 and PALSAR-2. Remote Sensing, 9(2), 119. https://doi.org/10.3390/rs9020119

van Vliet, J. (2013). Calibration and validation of land--use models. Wageningen University, Wageningen, The Netherlands.

van Vliet, J., Bregt, A. K., Brown, D. G., van Delden, H., Heckbert, S., & Verburg, P. H. (2016). A review of current calibration and validation practices in land-change modeling. In Environmental Modelling and Software (Vol. 82, pp. 174–182). Elsevier Ltd. https://doi.org/10.1016/j.envsoft.2016.04.017

van Vliet, J., de Groot, H. L. F., Rietveld, P., & Verburg, P. H. (2015). Manifestations and underlying drivers of agricultural land use change in Europe. In Landscape and Urban Planning (Vol. 133, pp. 24–36). Elsevier B.V. https://doi.org/10.1016/j.landurbplan.2014.09.001

van Vliet, J., Naus, N., van Lammeren, R. J. A., Bregt, A. K., Hurkens, J., & van Delden, H. (2013). Measuring the neighbourhood effect to calibrate land use models. Computers, Environment and Urban Systems, 41, 55–64. https://doi.org/10.1016/j.compenvurbsys.2013.03.006

Veldkamp, A., & Lambin, E. F. (2001). Predicting land-use change. Agriculture, Ecosystems & Environment, 85(1–3), 1-6Veldkamp, a, Lambin, E. . (2001). Predicting. https://doi.org/10.1016/S0167-8809(01)00199-2

Verburg, P. H., Alexander, P., Evans, T., Magliocca, N. R., Malek, Z., Rounsevell, M. DA, & van Vliet, J. (2019). Beyond land cover change: towards a new generation of land use models. In Current Opinion in Environmental Sustainability (Vol. 38, pp. 77–85). Elsevier B.V. https://doi.org/10.1016/j.cosust.2019.05.002

Verburg, P. H., Crossman, N., Ellis, E. C., Heinimann, A., Hostert, P., Mertz, O., Nagendra, H., Sikor, T., Erb, K. H., Golubiewski, N., Grau, R., Grove, M., Konaté, S., Meyfroidt, P., Parker, D. C., Chowdhury, R. R., Shibata, H., Thomson, A., & Zhen, L. (2015). Land system science and sustainable development of the earth system: A global land project perspective. Anthropocene, 12, 29–41. https://doi.org/10.1016/J.ANCENE.2015.09.004

Verburg, P. H., Erb, K. H., Mertz, O., & Espindola, G. (2013). Land System Science: between global challenges and local realities. Current Opinion in Environmental Sustainability, 5(5), 433–437. https://doi.org/10.1016/J.COSUST.2013.08.001

Verburg, P. H., & Overmars, K. P. (2009). Combining top-down and bottom-up dynamics in land use modeling: Exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. Landscape Ecology, 24(9), 1167–1181. https://doi.org/10.1007/s10980-009-9355-7

Verburg, P. H., Ritsema van Eck, J. R., de Nijs, T. C. M., Dijst, M. J., & Schot, P. (2004). Determinants of land-use change patterns in the Netherlands. Environment and Planning B: Planning and Design, 31(1), 125–150. https://doi.org/10.1068/b307

Verburg, P. H., Schot, P. P., Dijst, M. J., & Veldkamp, A. (2004). Land use change modelling: Current practice and research priorities. In GeoJournal (Vol. 61, Issue 4, pp. 309–324). Springer Netherlands. https://doi.org/10.1007/s10708-004-4946-y

Viana, C. M., Encalada, L., & Rocha, J. (2019). The value of OpenStreetMap Historical Contributions as a Source of Sampling Data for Multi-temporal Land Use/Cover Maps. ISPRS International Journal of Geo-Information, 8(3), 116. https://doi.org/10.3390/ijgi8030116

Viana, C. M., Freire, D., Abrantes, P., & Rocha, J. (2021). Evolution of Agricultural Production in Portugal during 1850–2018: A Geographical and Historical Perspective. Land 2021, Vol. 10, Page 776, 10(8), 776. https://doi.org/10.3390/LAND10080776

Viana, C. M., Freire, D., Abrantes, P., Rocha, J., & Pereira, P. (2022). Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review. Science of The Total Environment, 806, 150718. https://doi.org/10.1016/J.SCITOTENV.2021.150718

Viana, C. M., Girão, I., & Rocha, J. (2019). Long-Term Satellite Image Time-Series for Land Use/Land Cover Change Detection Using Refined Open Source Data in a Rural Region. Remote Sensing, 11(9), 1104. https://doi.org/10.3390/rs11091104

Viana, C. M., & Rocha, J. (2020). Evaluating Dominant Land Use/Land Cover Changes and Predicting Future Scenario in a Rural Region Using a Memoryless Stochastic Method. Sustainability, 12(10), 4332. https://doi.org/10.3390/su12104332

Viana, C. M., Santos, M., Freire, D., Abrantes, P., & Rocha, J. (2021). Evaluation of the factors explaining the use of agricultural land: A machine learning and model-agnostic approach. Ecological Indicators, 131, 108200. https://doi.org/10.1016/J.ECOLIND.2021.108200

von Thünen, J. H. (1826). Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalokonomie.

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. Remote Sensing of Environment, 236, 111402. https://doi.org/10.1016/j.rse.2019.111402

Wilkerson, G. G., Jones, J. W., Boote, K. J., Ingram, K. T., & Mishoe, J. W. (1983). Modeling soybean growth for crop management. Transactions of the American Society of Agricultural Engineers, 26(1), 63–73.

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L. J., Fanzo, J., Hawkes, C., Zurayk, R., Rivera, J. A., De Vries, W., Majele Sibanda, L., … Murray, C. J. L. (2019). Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. The Lancet, 393(10170), 447–492. https://doi.org/10.1016/S0140-6736(18)31788-4

Winkler, K., Fuchs, R., Rounsevell, M., & Herold, M. (2021). Global land use changes are four times greater than previously estimated. Nature Communications 2021 12:1, 12(1), 1–10. https://doi.org/10.1038/s41467-021-22702-2

World Bank Group. (2015). Ending Poverty and Hunger by 2030 An Agenda for the global food system.

Worthington, E. B. (1975). The Evolution of IBP. Cambridge University Press.

Wu, W. Bin, Yu, Q. Y., Peter, V. H., You, L. Z., Yang, P., & Tang, H. J. (2014). How could agricultural land systems contribute to raise food production under global change? Journal of Integrative Agriculture, 13(7), 1432–1442. https://doi.org/10.1016/S2095-3119(14)60819-4

Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. Remote Sensing of Environment, 122, 2–10. https://doi.org/10.1016/j.rse.2012.01.010

Xiao, X., Boles, S., Liu, J., Zhuang, D., Frolking, S., Li, C., Salas, W., & Moore, B. (2005). Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. Remote Sensing of Environment, 95(4), 480–492. https://doi.org/10.1016/j.rse.2004.12.009

Yang, Y., Zhang, S., Liu, Y., Xing, X., & De Sherbinin, A. (2017). Analyzing historical land use changes using a Historical Land Use Reconstruction Model: A case study in Zhenlai County, northeastern China. Scientific Reports, 7(January). https://doi.org/10.1038/srep41275

Yu, M., Bambacus, M., Cervone, G., Clarke, K., Duffy, D., Huang, Q., Li, J., Li, W., Li, Z., Liu, Q., Resch, B., Yang, J., & Yang, C. (2020). Spatiotemporal event detection: a review. International Journal of Digital Earth, 13(12), 1339–1365. https://doi.org/10.1080/17538947.2020.1738569

Yu, Q., Wu, W., Yang, P., Li, Z., Xiong, W., & Tang, H. (2012). Proposing an interdisciplinary and cross-scale framework for global change and food security researches. Agriculture, Ecosystems and Environment, 156, 57–71. https://doi.org/10.1016/j.agee.2012.04.026

Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. Remote Sensing of Environment, 144, 152–171. https://doi.org/10.1016/J.RSE.2014.01.011

Zscheischler, J., & Rogga, S. (2015). Transdisciplinarity in land use science – A review of concepts, empirical findings and current practices. Futures, 65, 28–44. https://doi.org/10.1016/J.FUTURES.2014.11.005

Part II. Agricultural land systems: state-of-art

Chapter 2. Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review

Cláudia M. Viana, Dulce Freire, Patrícia Abrantes, Jorge Rocha, Paulo Pereira

Article published at the Science of the Total Environment (2022), Volume 806, Part 3, 150718. https://doi.org/10.1016/j.scitotenv.2021.150718

The structure and formatting are slightly adapted.

Abstract

Agriculture provides the largest share of food supplies and ensures a critical number of ecosystem services (e.g., food provisioning). Therefore, agriculture is vital for food security and supports the Sustainable Development Goal (SDGs) 2 (SDG 2 - zero hunger) as others SDG's. Several studies have been published in different world areas with different research directions focused on increasing food and nutritional security from an agricultural land system perspective. The heterogeneity of the agricultural research studies calls for an interdisciplinary and comprehensive systematization of the different research directions and the plethora of approaches, scales of analysis, and reference data used. Thus, this work aims to systematically review the contributions of the different agricultural research studies by systematizing the main research fields and present a synthesis of the diversity and scope of research and knowledge. From an initial search of 1151 articles, 260 meet the criteria to be used in the review. Our analysis revealed that most articles were published between 2015-2019 (59%), and most of the case studies were carried out in Asia (36%) and Africa (20%). The number of studies carried out in the other continents was lower. In the last 30 years, most of the research was centred in six main research fields: land-use changes (28%), agricultural efficiency (27%), climate change (16%), farmer's motivation (12%), urban and peri-urban agriculture (11%), and land suitability (7%). Overall, the research fields identified are directly or indirectly linked to 11 of the 17 SDGs. There are essential differences in the number of articles among research fields, and future efforts are needed in the ones that are less represented to support food security and the SDGs.

Keywords: Agricultural research; Zero hunger; Land-use changes; Agricultural efficiency; Climate change

2.1. Introduction

Since the 1996 World Food Summit (WFS), massive efforts have been made in increasing agriculture food production and security (Ericksen, 2008; FAO, 2017). More recently, in 2015, the United Nations (UN) set the 17 Sustainable Development Goals (SDGs), where an essential goal is Zero Hunger (SDG 2). Despite the great efforts carried out in the last decades in developing strategies and policies towards the achievement of global food security, nowadays, approximately one in ten persons worldwide are suffering from severe levels of food insecurity (FAO et al., 2020). The demographic growth, accelerated urbanization (FAO, 2017; Tomlinson, 2013), the non-sustainable consumption of non-renewable resources (Popp et al., 2014; Tilman et al., 2011), climate change (Abd-Elmabod et al., 2020; IPCC, 2014; Schmidhuber & Tubiello, 2007), the changing of food consumption pattern (e.g., increase in overall calorie intakes; diet structure changes towards increase of meat, eggs, among others products) (Godfray et al., 2010; Guyomard et al., 2012; Hatab et al., 2019; Kastner et al., 2012), will put important challenges in food security. Population growth is expected to increase undernourishment (Hall et al., 2017), while the intensive exploitation of resources may lead to land degradation and reduce soil productivity (Tóth et al., 2018). The increase of extreme events (e.g., droughts and floods) and the increasing frequency of pests and diseases associated with climate change can be responsible for crop failure or destruction (e.g., Richardson et al. 2018, Spence et al. 2020). Finally, the changing of food patterns and the demand for more products is increasing the demand for land and water resources, exhausting the resources and increasing the uncertainty regarding food security. Therefore, the food and nutrition security agenda calls for urgent international efforts with effective global food security insurance (FAO et al., 2020; Ruben et al., 2018; Schneider et al., 2011).

Agricultural land provides the largest share of food supplies and ensures an essential number of ecosystem services (e.g., providing food, fuel, fibre) (Pereira, Bogunovic, et al., 2018; Scown et al., 2019; Stephens et al., 2018). Mainly, agricultural land contributes (directly or indirectly) to approximately 90% of food calories (Cassidy et al., 2013) and 80% of protein and fats (livestock production) (Steinfeld et al., 2006). Therefore, agricultural areas support food security and SDGs achievement (Avtar et al., 2020; DeClerck et al., 2016; FAO, 2017; Godfray et al., 2010; Wu et al., 2014). Also, agriculture, especially when practiced sustainably, is dependent, connected or essential to improve other SDG's. For instance, it is vital to reduce poverty (Goal 1-No poverty-e.g., targets 1.4 and 1.5). Increase population wellbeing (Goal 3-Good Health and Wellbeing-e.g., target 3.9) and support knowledge and R&D (Goal 4-Quality education; e.g., Target 4.7). Improve water quality and use efficiency (Goal 5-Clean water and Sanitation; e.g., Targets 6.3, 6.4), energy efficiency and investment in clean energy (Goal 6- Affordable and Clean Energy; e.g., Targets 7.2. and 7.3). Also, it is essential to improve the farmers working conditions and resource efficiency use (Goal 8- Decent work and economic growth; e.g., Targets 8.2, 8.3 and 8.4), support small scale farmers and promote innovation (Goal 9- Industry Innovation and Infrastructure; e.g., Targets 9.3 and 9.4) and improve a fair trade between producers and consumers (Goal 10- Reduced Inequalities; e.g., Target 10.a). Agriculture contributes to the increase of urban areas livability and access to green spaces (e.g., urban gardens, green roofs), reduce the impact of natural hazards and pollution and ensure food security (Goal 11- Sustainable Cities and Communities'; e.g., Targets 11.5, 11.6, 11.7 and 11.a). Agriculture friendly practices contribute to the efficient management of natural resources (e.g., soil and water) and reduce food waste and waste production (Goal 12- Sustainable Cities and Communities; e.g., Targets 12.1, 12.2, 12.3, 12.4 and 12.5), to reduce the greenhouse gases emissions and mitigate the impacts of climate change-related events (Goal 13-Climate Action; e.g., Target 13.1), to decrease the agrochemicals application and the pollution of surface water bodies (Goal 14- Life Bellow Water; e.g., Target 14.1) and to reduce the intensive agriculture practices (e.g., deep tillage, agrochemicals application), deforestation and land degradation (Goal 15-Life on land; e.g., Targets 15.1; 15.2; 15.3, 15.4 and 15.5). Unsustainable agriculture practices that may lead to resource exhaustion or land degradation may trigger conflicts. Therefore, sustainable land management is key to reducing the conflicts resulting from the lack of food (Goal 16- Peace Justice and Strong Institutions) (United Nations, 2015b). Although agriculture has an essential role in improving an important number of SDG's, several works highlighted the existence of tradeoffs between SDG's. For instance, the increase of food production to support No Poverty (Goal 1) or Zero Hunger (Goal 2), may have negative implications in the achievement of other goals such as Climate Action (Goal 13), Life Bellow Water (Goal 14) and Life on Land (Goal 15) (e.g., Bowen et al. 2017, Moyer and Bohl 2019, Yang et al. 2020). To minimize the tradeoffs associated with agriculture impacts, it is vital to invest and develop new technologies for data acquisition (e.g., remote and proximal sensing) and create robustly and validated models that consider data from multiple sources (Brevik et al., 2016; Gomes et al., 2021). This will be essential to identify accurately where the agriculture areas are more productive and where they can have more detrimental impacts on the ecosystems. This will be vital to have better agricultural land management. However, for this to be fully operational interdisciplinary research is needed (Pereira, Brevik, et al., 2018).

Several studies have been focused on the global challenge of increasing food security from an agricultural land systems perspective (i.e., a man-made system created with the purpose of livestock and cropland production) (e.g., Antle et al., 2017; Stephens et al., 2018; Wu et al., 2014; Yu et al., 2012). The works carried out are very heterogenous (Koutsos et al., 2019) from different scientific disciplines (e.g., geography, ecology, soil science, agronomy, economy) (Jones et al., 2016). Few works (e.g., van Noordwijk *et al.* 2018, Nicholls *et al.* 2020) have been carried out on revision or systematization of the published research studies about agriculture and how they contribute to SDG's achievement. Therefore, based on a systematic literature review, this study aimed to assess knowledge on agricultural land system research. The specific objectives are 1) Identify and describe the principal research fields of the published research studies from an agricultural land system perspective; 2) Link the identified research fields to SDG's; 3) Assess the tradeoffs associated with agricultural systems and 4) Demonstrate the current methodological research directions. This study provides an essential synthesis to understand the main focus of agricultural land system research, where they were conducted, the methodological approach, and how they support food security and the SDGs.

2.2. Systematic review framework

2.2.1. Search strategy, keywords and criteria of selection

A systematic review of the literature was conducted following the framework developed by Koutsos et al. (2019). The method follows six steps: 1) scoping, 2) planning, 3) identification and search, 4) screening articles, 5) eligibility assessment, and 6) presentation and interpretation. 1) scoping: the protocol for the review was defined. In this study, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement protocol (Moher, 2009) was chosen to guarantee the scientific quality and ensure a transparent, systematic review. 2) planning: the suitable databases were identified, and the search strategies were developed. This study opted to include only the WOS database because we wanted to select the articles going through a rigorous peer-review process and (theoretically) are considered in indexed journals. Furthermore, because the agricultural land system is a vast topic, we combined different relevant keywords and boolean operators related to the study aim using the search string: "TI =(agricultur* AND land* OR cultivat* AND land* OR crop* AND agricultur*) AND TS = (food)" (we assumed "Agricultural land" and "Cultivated land" interchangeably). 3) identification and search: the articles to include from the database were retrieved and identified. The search was conducted in January 2020. In total, 1151 articles were selected. 4) screening: the identified articles were filtered to meet the determined criteria: documents typeset as articles; peer-review scientific journals; published until 2019; English language; and related to agricultural land for food-grain crops (food human consumption). Therefore, after the title and abstract reading, 851 articles were excluded because they were focused on other types of agricultural production, such as, e.g. the production of feed, energy fuel/biofuel, the study of mammals, impacts of agricultural land on the ecosystems. 5) eligibility assessment: the full-text articles that do not meet the criteria were excluded. In total, 260 articles were identified as eligible for the review. After a detailed analysis of the articles, they were grouped into six different topics: land suitability (i.e. best location for specific land use), land management (i.e., control the use and development of land resources), land conversion/change (i.e., the transformation of the natural landscape), land use (i.e., the human use of land), land efficiency (i.e., managing land use under land policies and principles of sustainable development), and climate adaptation (i.e., adaptation of natural or human systems to the current or predicted climate change and their effects). These research topics are broadly related to food security and agricultural land systems and were adopted from Wu et al. (2014) and Yu et al. (2012), both review articles propose strategies to raise future food production and describe global change/food security studies. Group the included research articles is a crucial step because of the heterogeneity of the studies and allows to summarize and describe their methodological directions. Thus, if the article primarily assessed one of these six research topics, it was assigned to one broad research field. 6) interpretation and presentation: the article's content, i.e., the publication year, geographic coverage, approach, methodological characteristics, reference data used, were identified and synthesized. Since this is a review article, more minor detail will be provided regarding the theoretical and methodological approaches of each individual article and statistical analysis due to differences in each article objectives, scientific field background, data availability, or other technological used sources. Finally, after grouping the included articles by each research field, we provide a logical relation between the different research fields and the SDG's based on current literature and the author's understanding of the topic. The description and supporting references are not directly related to the selected articles from the systematic literature review. The framework applied in this work is sensitized in Figure 2.1.

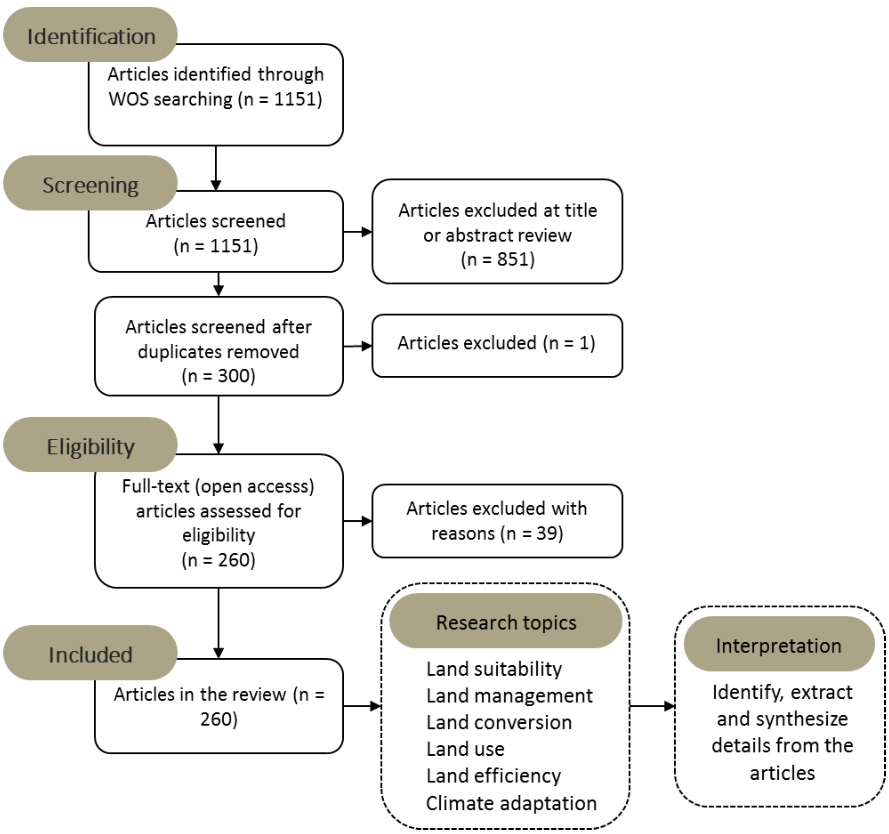


Figure 2.1. Systematic review framework based on PRISMA protocol and inductive approach

2.3. Results and discussion

2.3.1. The historical and geographical context of the articles

The number of articles published per year is shown in Figure 2.2. The evolution of the number of works can be divided into three stages: 1) 1991-1999: reduced number of papers (6 articles); 2) 2000-2009: an increase in the number of published articles (32 articles), and 3) 2010-2019: a large number of published works (222 articles). Overall, approximately 59% of the articles were published during 2015-2019, simultaneously with the establishment of different global strategies such as the Millennium Development Goals (MDGs) settled in 2000 for the subsequent 15 years (United Nations, 2015a), and the UN SDGs established in 2015 (United Nations, 2015b). These international strategies highlighted the challenges that need to be addressed by humanity. In particular, the SDGs consider a comprehensive approach, involving poverty, hunger, prosperity, environment, climate, peace, and justice (Griggs et al., 2013), paving the road towards a more sustainable world to meet the needs of the current generation without compromising the needs of future generations (United Nations, 2015b). The majority of the articles were published in Asia (96 articles) and Africa (52 articles). Developed regions in America, Europe, and Oceania received less attention than Africa and Asia with 45, 38 and 7 published articles, respectively. Besides, 31 articles were focused on the global scale. Some articles focused on more than one region (Figure 2.3).

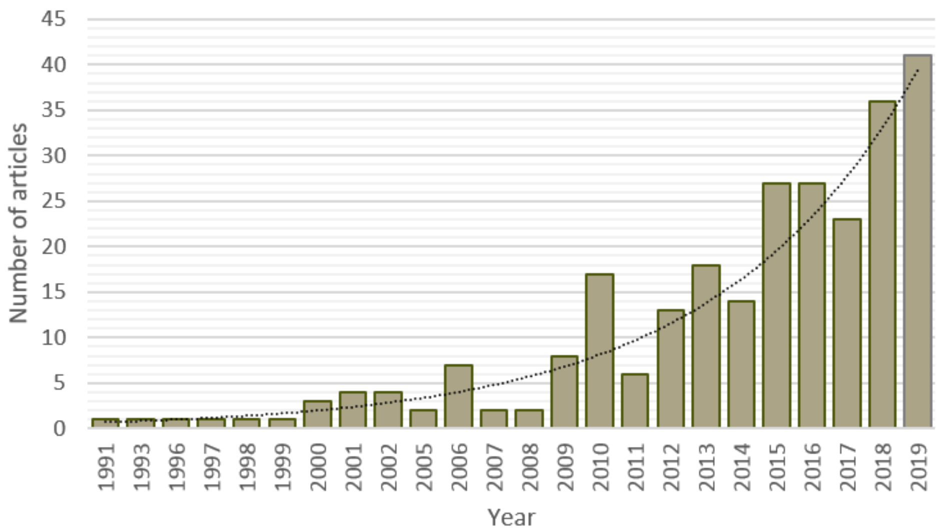


Figure 2.2. Number of articles published by year

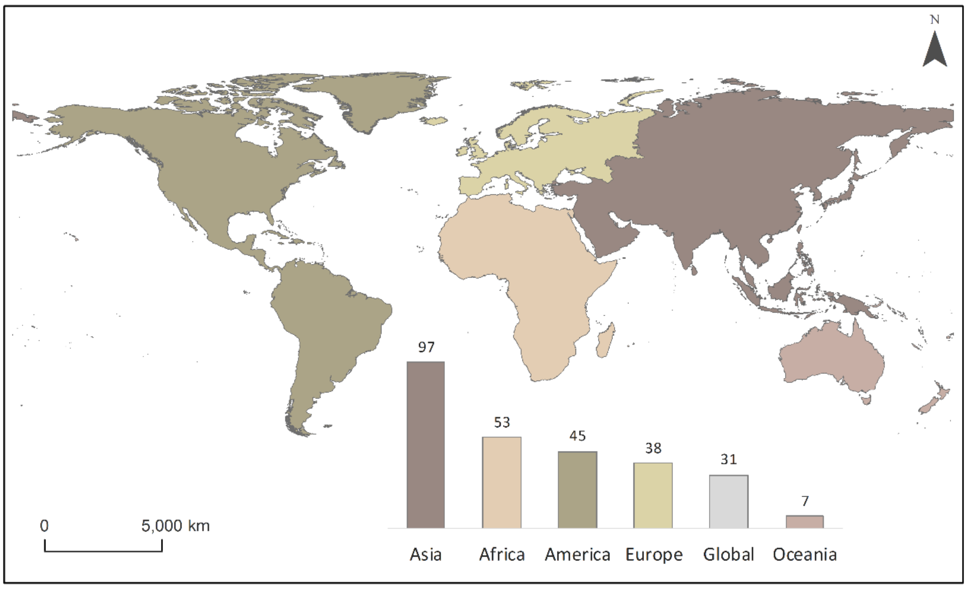


Figure 2.3. Geographic coverage of the article case study

The 260 selected articles cover six principal research fields (Figure 2.4). Three research fields become more preeminent during the 1990s: 1) *efficiency of agricultural systems*, 2) *urban and peri-urban agriculture movement*, and 3) *effect of climate change in agriculture*. During the first decade of the 21st century, the other research fields started to be more attractive. Overall, more than 77% of the articles were published between 2010-2019.

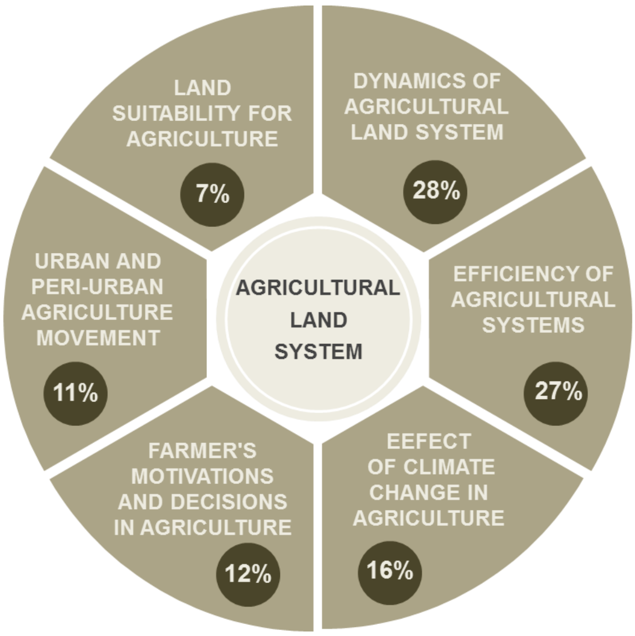


Figure 2.4. Synthesis of the research fields and percentage of articles per field

Most of the articles were included in the field of *efficiency of agricultural systems (50%), dynamics of agricultural land systems, land suitability for agriculture*, and *farmer's motivations* *and decisions in agriculture* were conducted in Asia (50%, 34%, 35%, 33%, respectively) and Africa (15%, 24%, 18%, 33%, respectively) (Figure 2.5). Southeast Asia and Subsarian Africa are the areas of the globe where hunger is an alarming problem (Alexandratos & Bruinsma, 2012), although cropland area (100 million ha in Asia and 59 in Africa) (OECD/FAO, 2009) and the foreign investment (Mason-D'Croz et al., 2019) increased substantially. In these regions, poverty, population and urbanization growth rates, climate change effects, vulnerability to extreme events, and food insecurity are much higher than elsewhere in the world (FAO et al., 2020; United Nations, 2019). Moreover, these are also the regions where land degradation is a serious problem (Millennium Ecosystem Assessment, 2005). Therefore, most of the efforts should be conducted in Southeast Asia and Subsarian Africa to increase the capacity of these regions to achieve the SDGs (FAO et al., 2020; FAO, 2021).

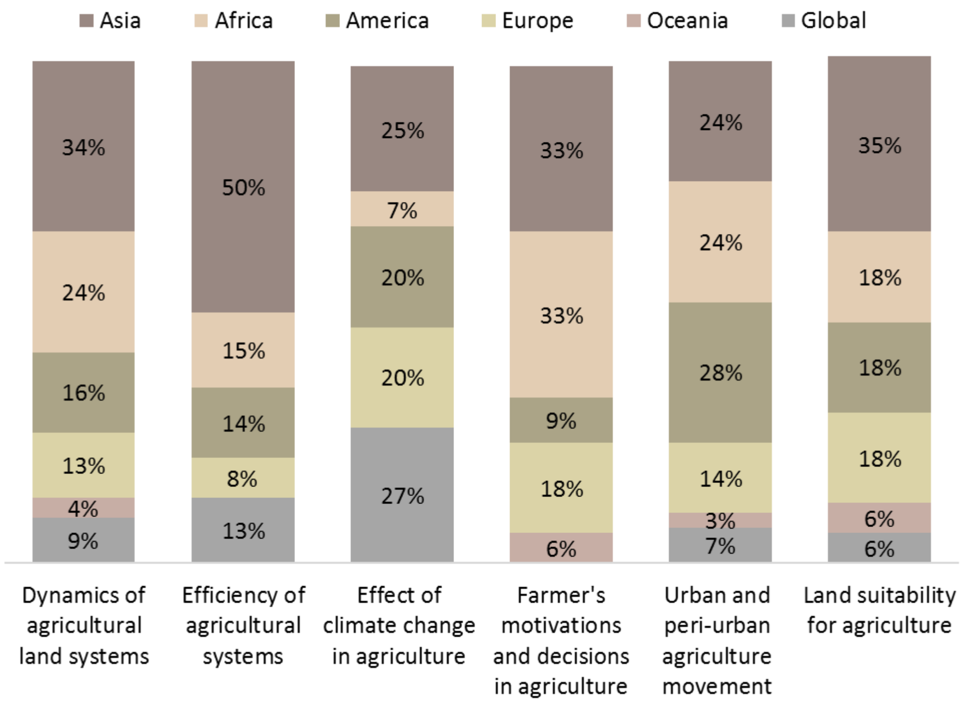


Figure 2.5. Location of the article case study per research field

The *Urban and peri-urban agriculture movement* research field was mainly carried out in (North) America (28%), Asia (24%) and Africa (24%). Urbanization in Northern America is not a new phenomenon, and 82% of people live in urban areas (2018) (United Nations, 2019). For instance, in the United States, there has been an increasing interest from different institutions (e.g., academic, civic) in urban agriculture issues in the last decade (Siegner et al., 2018). Therefore, this increased the research carried out. In Africa and Asia, urbanization is recent due to the massive rural exodus (United Nations, 2019). In this context, it is important to study the impacts of cities growth on land consumption and how this can affect food security in the following decades (United Nations, 2019; World Bank Group, 2015). Therefore, it is not surprising that many of the selected works are in Africa and Asia. The articles focused on *climate change's impact on agriculture* were mainly developed globally (27%) and in Asia (25%). Climate change is a global issue, irrespective of the geographic area. Therefore, the importance of international studies is high (e.g., Van Meijl et al., 2018). On the other hand, some areas are more affected than others, and the impacts are unequal (FAO et al., 2020). One of these areas is Asia, where the poverty rate and population density is high (e.g., Southeast Asia), exacerbating the impacts of climate change on agriculture (e.g., Im et al., 2017). Therefore, it was rather expected that a substantial number of the selected works were carried out in this continent. The challenges to achieving food security and the SDGs are aggravated by the vulnerability to climate change, mainly because it affects the most vulnerable and have a highly heterogeneous pattern (FAO, 2020; IPCC, 2014; Schmidhuber & Tubiello, 2007). Ensure global food security is a tremendous global challenge (FAO et al., 2020; FAO, 2021; World Bank Group, 2015). While most of the selected articles were conducted on developing regions in Asia and Africa, further studies are needed in areas with different realities where food security has improved substantially but still can be affected by climate change. For each research field, a differentiated cross-timescale analysis should be prioritized to help to face the challenges outlined by the SDGs (World Bank Group, 2015).

2.3.2. Main research fields from an agricultural land system perspective

2.3.2.1. Dynamics of agricultural land system

The *dynamics of the agricultural land system* was the research field with the highest number of published articles (28%). Several works were developed to detect, characterize, monitor, map, and model agricultural land. There were two main methodological directions: 1) Spatiotemporal expansion or contraction of agricultural land areas in different periods and geographic contexts (e.g., Cao et al., 2019; Kühling et al., 2016; Nakalembe et al., 2017). Several driving forces and processes, acting individually or coupled affect agricultural land-use changes, were identified and analyzed in the context of different scenarios. They are mainly related to the functioning of local and national financial markets, demographic trends, environmental factors, and internal and external policies (Eklund et al., 2017; Garrett et al., 2018; Piquer-Rodríguez et al., 2018). These drivers are incorporated in different land-use models (e.g., SLEUTH, cellular automata, Markov chains) to simulate spatiotemporal land changes and forecast future agricultural land changes (e.g., Grundy et al., 2016; Martellozzo et al., 2018) and 2) Mapping and/or monitoring agricultural land at different temporal and spatial scales (e.g., Piiroinen et al., 2015; Torbick et al., 2017). For instance, by compile and analyze real-time data and using multitemporal and multisensor methodologies, several articles studied the crops phenophase, crop nitrogen stress, the cropland rotation and diversity, and the crop yields (e.g., Monteleone et al., 2018; Samasse et al., 2018; Veloso et al., 2017). Overall, by combining geographic information systems (GIS) and the use of different data types (e.g., remote sensing data products, historical and statistical data), the dynamics of agricultural land is simulated in both the past, present, and the future at regional, national, and global scales (e.g., Grundy et al., 2016; Shi et al., 2016; Torbick et al., 2017).

2.3.2.2. The efficiency of agricultural systems

The *efficiency of agricultural systems* was the second field of research, with more articles selected (27%). Different factors (e.g., human and environmental) on agricultural land productivity and food production received important research attention. Three measures are generally mentioned in the literature: 1) agricultural production (the net produce or output of cropland), 2) agricultural crop yield (the amount of crop harvested per unit area of land), and 3) agricultural productivity (income produced per unit area of land or person employed, i.e. the market value of the final output) (Lin & Hülsbergen, 2017; Shen et al., 2013). Agricultural efficiency and productivity have been synonymously and interchangeably used. This is explained by the fact that agricultural productivity ((measured in terms of the amount of output (referred to as yield) per unit of area of input)) refers to the productive efficiency sector of the total agricultural efficacy. Thus, agricultural productivity is a part of agricultural efficiency, a broader concept expressed in crop productivity levels per unit area or other inputs or nutrition provided per unit area yield. These measures are used to assess the positive or negative influence of operational and structural factors in agricultural land production at different scales (e.g., Jin et al., 2017). Through various methods such as econometric analysis, crop model simulations, and non-parametric techniques (e.g., Van Ittersum et al., 2013), using aggregated (at the national or global level) or disaggregated data (at regional level), a wide range of environmental, institutional, organizational, managemental, and socioeconomic factors, are put in perspective to clarify their influence on productive efficiency (e.g., Hong et al., 2019; Mbata, 2001). Environmental and human factors such as water availability (e.g., Yan et al., 2015), land degradation and land fragmentation (e.g., Looga et al., 2018), terrain slope (e.g., Li et al., 2014), pest pressure (e.g., Drechsler et al., 2016) are often considered. Likewise, the influence of agricultural practices innovations adoption, from land consolidation (e.g., Hong et al., 2019), (bio) fertilizer applications (e.g., Nayak et al., 2019), herbicide/pesticide applications (e.g., Schreinemachers and Tipraqsa, 2012), conventional tillage (e.g., Das et al., 2014), are also being measured.

2.3.2.3. Effect of climate change in agriculture

Sixteen % of the articles selected were focused on the *effect of climate change in agriculture*. Briefly, several works evaluated short and long-term changes in the climatic conditions and their consequences on agricultural systems, considering different model's techniques (e.g., Leng and Hall, 2019; Manners and van Etten, 2018; Yu et al., 2012). The influence of individual factors within these scenario sets was used to forecast the geographic distribution of crop yield and agricultural productivity gains and losses in several world regions (e.g., Europe, Africa and Asia). In addition, the definition of scenarios identify the existing relationship between the factors associated with the climate change impacts on agricultural land systems and simulate the effect of future scenarios considering different future narratives (e.g., Ahmed et al., 2016). Overall, the methods applied biophysical and agro-ecosystem process-based models and statistical simulation models to estimate agricultural yields at different scales and climate change scenarios (e.g., Basso et al., 2015). Process-based models simulate crop growth processes, according to different climate factors, such as temperature or rainfall variations, and extreme upward events (particularly floods and droughts) (e.g., Kukal and Irmak, 2018; Leng and Hall, 2019), or soil properties and management (e.g., Basso et al., 2015). Statistical models estimate future trends in agricultural yields (e.g., by changes in temperature, CO_2, or fertilization) based on statistical correlations from historical trends (e.g., Mori et al., 2010).

2.3.2.4. Farmer's motivations and decisions in agriculture

The *farmer's motivations and decisions in agriculture* were studied by 12% of the works selected. Farmers' land-use decisions and their management strategies and motivations are often conditioned by agro-ecological, climatic, and political conditions, influencing local practices (e.g., Brady et al., 2012; Kvakkestad et al., 2015). Farmers' motivations depend on what they consider advantageous. They can minimize the risk of losses (e.g., which food crops to grow, the fallow period duration) or management decisions (e.g., fertilizers/pesticides input, mechanization, tillage, or irrigation) (e.g., TerAvest et al., 2019; Yang et al., 2017). Other articles were focused on the farmers' decisions regarding their choices towards uncertain factors that affect the agricultural productivity and farm food self-sufficiency (e.g., agricultural policy reform, government incentive/subsidy programs) (e.g., Brady et al., 2012; Chibwana et al., 2012), or others externalities (e.g., pest, plant disease, technology, soil contaminants, weather variability, or land tenure agreements) (e.g., Boz, 2016; Nkomoki et al., 2018). Furthermore, the different management options and the socioeconomic factors were analyzed at the farm level, i.e., as experimentation, to understand how the farm revenue or the crop yield is impacted depending by the practices adopted (e.g., Leonardo et al., 2015; Lin and Hülsbergen, 2017; Vasile et al., 2015). The constraints or opportunities of the farmland production situations were assessed through surveys or group discussions at the household/field level (e.g., Hao et al., 2015; Leonardo et al., 2015). In some cases, spatial and/or statistical data were included (e.g., Gunda et al., 2017; Lyle et al., 2015).

2.3.2.5. Urban and peri-urban agriculture movement

The urban and peri-urban agriculture movement was studied by 11% of the articles selected. Urban and peri-urban agriculture is a popular topic worldwide (Cerrada-Serra, Colombo, Ortiz-Miranda, & Grando, 2018; Li, Wang, Liu, & Zhu, 2019; Mackay, 2018). Depending on the research objectives, urban agriculture (UA) or peri-urban agriculture (PUA) are studied individually, while others are integrated (Urban and Peri-urban agriculture - UPA). UPA is a multidimensional concept since it covers different production techniques (e.g., aquaponics, hydroponics, permaculture productions, food crops and livestock) and purposes (e.g., pedagogy, consumption, farmers markets) in different locations (rooftops, communal or private gardens) and scales (Mougeot, 2011; UNDP, 1996). UPA consider subjects such as access to food, health, income, the environment or natural resources (Tornaghi, 2014). Overall, there are three essential approaches. 1) Governance and policies, i.e., articulation between urbanism, land-use planning and agriculture preservation in the urban and peri-urban areas (e.g., She et al., 2015). There are several issues related to land use conflicts and governance challenges. For example, 1) in the access to land that can be public or private, or in the form of organization that may be more spontaneous or more institutional (e.g. community gardens) (e.g., Ayambire et al., 2019; Cerrada-Serra et al., 2018); 2) Locate vacant, abandoned, or marginal land within a city or on its periphery to use it for agriculture practice (e.g., Pothukuchi, 2018; Saha and Eckelman, 2017) and 3) Identification of patterns and trends in agriculture land uses affected by rapid urban expansion (e.g., Li et al., 2019; Yu et al., 2018).

2.3.2.6. Land suitability for agriculture

Seven % of the papers selected focused on *assessing land suitability for agriculture*. These articles mainly focused on identifying suitable areas for agriculture use (e.g., Mesgaran et al., 2017; Musakwa, 2018) and producing specific agricultural food-grain crops (e.g., Boix and Zinck, 2008; Kazemi et al., 2016). The approaches used in this research field were focused on spatial analytical methods to identify the multiple factors that affect the suitability of the land (e.g., topography, soil properties, climatic characteristics, socioeconomic drivers) (e.g., Kazemi et al., 2016; Schiefer et al., 2016; Zabel et al., 2014). Usually, the weight of the variables is assessed using an expert's evaluation. Frequently, multi-criteria evaluation (MCE) methods (e.g., Musakwa, 2018) were applied, such as the analytic hierarchy process (AHP) (e.g., Geng et al., 2019), fuzzy logic (e.g., Zabel et al., 2014), and weighted linear combination (WLC) (e.g., Li et al., 2017; Montgomery et al., 2016). For instance, two or more methods can be combined. Overall, the output of the analysis is an agricultural land suitability map that commonly presents four categories (not suitable, marginally suitable, moderately suitable, and highly suitable) according to the land suitability index of the FAO (e.g., Geng et al., 2019; Mendas and Delali, 2012).

2.3.3. Insights towards "a better and more sustainable future for all."

Boosting agricultural production and productivity of agricultural land currently under production is a recognized strategy to enhance and maintain food supply and reduce hunger (Millennium Ecosystem Assessment, 2005; The Royal Society, 2009; Wu et al., 2014). The *efficiency of the agricultural systems* is a relevant field of research, where the influence of different factors on agricultural land productivity is evaluated to meet food needs (e.g., Hong et al., 2019; Mbata, 2001). Moreover, the progress in this field of research will contribute directly to SDG 2 (Zero hunger) and SDG 1 (No Poverty). As shown in World Bank (2015) report, an increase of 1% in food production reduce 0.48% and 0.72% of the poverty in South Asia and sub-Saharan Africa. In addition, the increase in agricultural efficiency of the main crops could substantially increase farmers' income and stimulates domestic trade in the countries (SDG 8 - Decent Work and Economic Growth) and promote good health and wellbeing (SDG 3 - Good Health and Well-being) (World Bank Group, 2015). It is necessary to incorporate different factors (e.g., environmental, institutional, organizational, and socioeconomic) to have a more e efficient agricultural management and reduce the impact on ecosystem services (FAO, 2011; Millennium Ecosystem Assessment, 2005). For instance, factors such as pests and pathogens are estimated to be responsible for reducing about 35% of crop yields (Oerke, 2006; József Popp et al., 2013). This may influence the progress toward the achievement of SDG 2. Expanding genetically modified crops (van Esse et al., 2020) or applying organic pesticides (Kalkura et al., 2021) could be viable solutions for decreasing crop yield losses associated with pests diseases. Nevertheless, the use of genetically modified plants can raise concerns regarding human health and biodiversity loss (e.g., Tsatsakis *et al.* 2017, Raman 2018). In addition, there are several shreds of evidence that herbicide-resistant crops do not provide better yields or decrease the application of herbicides. The investment in herbicide-resistant crops and the use of herbicides had several detrimental effects such as 1) decreased crop rotation and increased weed management based on herbicides; 2) the application of glyphosate-based herbicides affect soil microbiology and plant diseases resistance and 3) the use and abuse of glyphosate in the last 20 years increased the appearance of 34 glyphosate-resistant weed species (Schütte et al., 2017). The application of pesticides and herbicides in agriculture have been linked to the emergence of several chronical (e.g., diabetes, asthma, cancer) and other short-term diseases (e.g., headaches, dizziness, nausea, skin and eye irritation) (Brevik et al., 2020; Kim et al., 2017). Also, climate change and biodiversity loss increase pest and disease frequency, as highlighted in previous works (e.g., Anderson et al., 2004; Potter and Urquhart, 2017; Rosenzweig et al., 2001). Overall, the efforts carried out to increase food security and improve SDG 2 may be detrimental to the achievement of another (e.g., SDG 3 – good health and wellbeing; SDG 15 - life on land) (OECD, 2020; United Nations, 2015b).

Likewise, providing information on available and suitable land for agricultural production can contribute to the identification of the best areas for crop production, establish a sustainable intensification and maximize food production (EEA, 2017; Shen et al., 2013; Struik & Kuyper, 2017; Wu et al., 2014), which is in line with the SDG 2. Therefore, *land suitability for the agriculture* field of research is very relevant from a land management perspective (Akpoti et al., 2019), promoting proper, efficient, rational land use. This is highly relevant to the achievement of SDG 1 (no poverty), SDG 2 (zero hunger), SDG 6 (clean water and Sanitation), SDG 11 (sustainable cities and communities), SDG 12 (responsible consumption and production), SDG 13 (climate action), SDG 14 (life below water) and SDG 15 (life on land) (Akpoti et al., 2019). From this perspective, assessing the land potentially can also be an effective strategy to implement sustainable agriculture (OECD, 2020), which would strengthen population health (SDG 3-good wealth and wellbeing) (Li et al., 2017).

Agriculture covers approximately 38% of the land surface. However, in some areas (e.g., urban and peri-urban areas), food security is decreasing (Foley et al., 2011; Godfray et al., 2010; Grundy et al., 2016; Radwan et al., 2019). Therefore, evaluate the changes in agricultural lands, and the drivers responsible for such changes are the motivation of the *dynamics of the agricultural land* field of research. The agriculture spatiotemporal changes will support food production and safety policy decisions, in line with SDG 2 (Sun et al., 2018; van Vliet et al., 2015; Viana & Rocha, 2020). Moreover, spatially and temporally accurate information contributes to effective land management, which is a key towards sustainable land use (OECD, 2020) and important for meeting SDG 6, SDG 7 (transitioning to clean energy), SDG 13 (Kasperson & Kasperson, 2001; Ogle et al., 2017; Tyson et al., 2001), and improve the ecosystems (SDG 15) (Weiss et al., 2020).

Meeting global food supply demand for a growing population (FAO, 2017) is one of the 21st-century challenges that will be exacerbated by climate change (FAO, 2020; IPCC, 2014). Moreover, the areas threatened by climate change and high population growth are located in the same geographical area (Sub-Saharan Africa and Southeast Asia) (Alexandratos & Bruinsma, 2012; United Nations, 2019). Despite technological progress, food production will be negatively affected by the changing climate patterns and increases in the frequency and intensity of extreme weather events (Abd-Elmabod et al., 2020; IPCC, 2014). Therefore, the *effect of climate change in the agriculture* field is key to develop essential knowledge to forecast agricultural production under different climate scenarios (Abd-Elmabod et al., 2020; Fanzo et al., 2018; Leng & Hall, 2019) and improve food safety, better health, and strengthen resilience to climate variability (SDGs 2, 3 and 13), as well as improve the ecosystems (SDG 15) (Arora, 2019).

The production of food is also developed outside the rural areas. The research focused on this topic is timely. Since 2018, more than half of the human population has lived in urban environments, and by 2050 this proportion is expected to increase to 68% (United Nations, 2019). As *the urban and peri-urban agriculture movement* field of research suggests, agriculture in urban and peri-urban contexts is a global trend that has been enforced as a strategy to combat climate change (SDG 13), increase food security (SDG 2) and make the urban areas more liveable (SDG 11) (Brevik et al., 2020; Ferreira et al., 2018; Opitz et al., 2016). Agricultural activities in urban and peri-urban areas have many benefits (Lin et al., 2015; Saha & Eckelman, 2017). Contributes to meet the nutritional needs by providing access to fresh and healthy food products (SDG 2), improve human health and wellbeing (SDG 3) (Aubry et al., 2012; Santo et al., 2016), generates local food economies contributing to poverty alleviation (SDG 1) (Lwasa et al., 2015); and promotes local educational (SDG 4 – quality education) (Eigenbrod & Gruda, 2015). Moreover, urban and peri-urban agriculture have relevant ecological and social functions in air quality regulation, soil erosion regulation, floods regulation, increase the population accessibility to green spaces, and promote efficient water management (SDGs 6, 11, and 15) (Ayambire et al., 2019), increasing the sustainability of the urban areas (Sioen et al., 2018; Tsuchiya et al., 2015; Zezza & Tasciotti, 2010).

The response to our right to food and security depends on the interests and motivations of local land use decision-makers. While the government put policies and design incentives to induce changes in individuals' behaviour, the farmer manages the farm according to their interests (Foguesatto et al., 2020). The *farmer's motivations and decisions in agriculture* is a very relevant field of research since it puts in perspective the farmers' management decisions (Malek et al., 2019) and supports SDGs. Specifically, it will contribute to improving the income and livelihood levels of farmers reducing rural poverty (SDGs 1 and 8) or enhance the local and regional food production needs (SDG 2) (FAO, 2020; World Bank Group, 2015). In addition, more informed decisions can be taken to evaluate food system vulnerability, reduce the impacts on climate factors (SDG 13), and foster sustainable agricultural practices and natural resources exploitation (SDGs 6 and 12) (FAO, 2020; United Nations, 2015b).

2.3.4. The trade-offs dilemma

For the coming decades, agricultural areas need to double the food production to ensure a stable and accessible food supply (Foley et al., 2011; Tomlinson, 2013). However, the current agri-food systems (e.g., agricultural practices, food preferences and consumption shifts) are increasing greenhouse gases emissions and ecosystems degradation (e.g., soil degradation, loss of biodiversity, water scarcity) (Goucher et al., 2017; Pereira, Bogunovic, et al., 2018; Sánchez-Bayo & Wyckhuys, 2019; Zuo et al., 2018). This agriculture expansion and the associated impacts (e.g., greenhouse gases emissions and ecosystems degradation) are evident in areas near the sub-tropical and tropical forests (e.g., Amazon) (Aizen et al., 2019; Monteles et al., 2021). The tradeoffs associated with agriculture may cause a global crisis in food security or environmental degradation at an unprecedented scale (Michel-Villarreal et al., 2019; Yang et al., 2020). As a result, any progress achieved in food security and SDG 2 may not represent an increase in the sustainable environment because it can be very detrimental to the environment (Foley et al., 2011; Marsden & Morley, 2015; Scherer et al., 2018; World Bank Group, 2015). Therefore, important decisions must be made to minimize the tradeoff between increasing food production and reducing greenhouse gas emissions and biodiversity loss from agriculture (FAO, 2017; FAO et al., 2020; Foley et al., 2011; United Nations, 2019). This is a challenge faced by the agri-food sector since a reduction in food system greenhouse gas emissions, water use, biodiversity loss, and soil degradation is key to decrease the agriculture footprint (EEA, 2017; Foley et al., 2011; United Nations, 2019; van de Kamp et al., 2018). The wide implementation of well-known solutions is needed such as reduce food waste (e.g., Rosenzweig *et al.* 2020), water reuse (e.g., Ricart and Rico 2019), agriculture intensification (e.g., tillage, pesticides and herbicides use) (e.g., Islam *et al.* 2020, Sattler *et al.* 2020), meat consumption (Stoll-Kleemann & Schmidt, 2017) and invest in sustainable practices based on no-tillage (e.g., Dachraoui and Sombrero 2020), crop diversification, use of organic fertilizers, increase rotation periods and cover cropping (e.g., Feng *et al.* 2018) that are beneficial to increase the crop resilience to pests (e.g., Murrell 2017) and can increase yield, as observed in several works (e.g., Huang *et al.* 2018, Beillouin *et al.* 2019, Jat *et al.* 2019). In addition, more efforts are needed to increase food quality, a target that can be achieved using sustainable agriculture practices (e.g., Lampridi et al., 2019; Michel-Villarreal et al., 2019; Zulfiqar et al., 2019).

Nevertheless, the concept of sustainable development is multidimensional in time and space and is achieved if there is socioeconomic development and environmental protection (Allen & Prosperi, 2016). Remarkably, the solutions to the long-term sustainability and food supply require adopting sustainable agricultural practices as an effective strategy with reduced environmental impact (Allen & Prosperi, 2016; EEA, 2017; Wu et al., 2014). It is essential to promote local, diverse, and sustainable agriculture that respects the environment and understanding international trade as a complement to local production. The local and national systems need to be strengthened to adapt to the climate crisis and diversify the farmed products. Crop diversity can reduce crop vulnerability to pests and diseases risks, open markets for different food crops, break their dependence on commercial crops, increase biodiversity, and reduce the impacts on climate change (The Royal Society, 2009). In addition, the success of agricultural transformation depends mainly on smallholders' capacity to adopt sustainable practices and adapt to climate change (FAO, 2017; Li et al., 2020). All in all, the effectiveness of research, policies, planning, and investment to build a resilient agricultural system and increase food production depends on local and global challenges and how they mitigate the tradeoffs caused by food production (The Royal Society, 2009).

2.3.5. Geospatial data, spatial analysis, integrated models, and interdisciplinary research

Agricultural land systems depend on different environmental and socioeconomic factors that interact in space and time (Scown et al., 2019; Stephens et al., 2018). Therefore, it is a dynamic and complex social-ecological system and a top research priority for global development and sustainability (Allen & Prosperi, 2016; Müller et al., 2020; Yu et al., 2012). Therefore, agricultural land systems research needs an interdisciplinary approach and incorporates different science branches (Ingram et al., 2020; Ruben et al., 2018; Scown et al., 2019; Yu et al., 2012) at different temporal and spatial scales (Yu et al., 2020).

From this literature review, it is clear that research and innovation activities, coupled with systemic theories, multitemporal and multisensory technologies, GIS techniques, scenario development and analysis, land-use models, agricultural economic/trade modelling, and geospatial data, are essential to address several SDG's challenges (Avtar et al., 2020; Müller et al., 2020; Weiss et al., 2020). For instance, many SDG targets are related to geoinformation (Avtar et al., 2020; Yuan, 2021). Specifically, data availability, timely, spatially, and accurately, is crucial to monitoring SDGs achievement in different countries (United Nations, 2012, 2015b). In addition, long-term databases are helpful to improve both climate-change and land-change models or establish more informed baselines for the different scientific disciplines, particularly those involving food security (Boivin & Crowther, 2021; World Bank Group, 2015). Also, the improvements in data collection, methodological advances, and robust models are essential to improve the current knowledge and likely open new questions that need to be addressed in different geographical and temporal contexts.

2.3.6. Limitations of this systematic review

The literature review is limited to WOS articles. Grey literature (e.g., reports) was discarded because we only want to review peer-reviewed articles in indexed journals to ensure quality control and scientific credibility. Likewise, the screening procedure involved synthesizing the information contained in the articles, which has undoubtedly led to generalization, and some information might be underestimated. It should be noted that it was outside of the article scope to analyze individual articles' theoretical and methodological approaches. Therefore, a vital implication of this study is that there has been no critical analysis of the methodologies applied in the different publications. Moreover, the empirical identification of research fields was based on the adaptation of six broad research topics from Yu et al. (2012) and Wu et al. (2014) works and the author's academic background, meaning there is always room for further improvement. The number of papers per continent may be related to the country's investment in R&D. Despite the limitations; this work provides important insights regarding the current state of knowledge of food security and their relation with the different SDG's.

2.4. Conclusions

This paper presented a systematic literature review, describing the main research fields in agricultural land systems and their linkage with the SDGs. Our analysis revealed that most articles were published during 2015-2019 (59%), with the case studies focused mainly on developing regions in Asia (36%) and Africa (20%). In the last 30 years, the body of research has been centred in six main research fields in the subjects of land-use changes (28%), agricultural efficiency (27%), climate change (16%), farmer's motivation (12%), urban and peri-urban agriculture (11%), and land suitability (7%). Each research field is diversified and highly important for long-term global development, providing approaches with different cross-scale frameworks and geographical contexts. The six areas are directly or indirectly linked to 11 of the 17 SDGs. However, the discrepancy in the percentage of publications by research field emphasizes the need for future studies to fulfil this gap because each domain has a vital role in providing knowledge to food security and the SDGs. In this context, more studies are needed in the different geographic areas and research fields, and this can be improved by using new datasets, methodological approaches and robust models.

References

Abd-Elmabod, S. K., Muñoz-Rojas, M., Jordán, A., Anaya-Romero, M., Phillips, J. D., Laurence, J., Zhang, Z., Pereira, P., Fleskens, L., van der Ploeg, M., & de la Rosa, D. (2020). Climate change impacts on agricultural suitability and yield reduction in a Mediterranean region. Geoderma, 374, 114453. https://doi.org/10.1016/j.geoderma.2020.114453

Ahmed, K. F., Wang, G., You, L., & Yu, M. (2016). Potential impact of climate and socioeconomic changes on future agricultural land use in West Africa. Earth System Dynamics, 7(1), 151–165. https://doi.org/10.5194/esd-7-151-2016

Akpoti, K., Kabo-bah, A. T., & Zwart, S. J. (2019). Agricultural land suitability analysis: State-of-the-art and outlooks for integration of climate change analysis. In Agricultural Systems (Vol. 173, pp. 172–208). Elsevier Ltd. https://doi.org/10.1016/j.agsy.2019.02.013

Alexandratos, N., & Bruinsma, J. (2012). World agriculture towards 2030/2050: the 2012 revision. (ESA Working Paper No. 12-03).

Allen, T., & Prosperi, P. (2016). Modeling Sustainable Food Systems. Environmental Management, 57(5), 956–975. https://doi.org/10.1007/s00267-016-0664-8

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends in Ecology & Evolution, 19(10), 535–544. https://doi.org/10.1016/J.TREE.2004.07.021

Antle, J. M., Basso, B., Conant, R. T., Godfray, H. C. J., Jones, J. W., Herrero, M., Howitt, R. E., Keating, B. A., Munoz-Carpena, R., Rosenzweig, C., Tittonell, P., & Wheeler, T. R. (2017). Towards a new generation of agricultural system data, models and knowledge products: Design and improvement. Agricultural Systems, 155, 255–268. https://doi.org/10.1016/j.agsy.2016.10.002

Arora, N. K. (2019). Impact of climate change on agriculture production and its sustainable solutions. Environmental Sustainability, 2(2), 95–96. https://doi.org/10.1007/s42398-019-00078-w

Aubry, C., Ramamonjisoa, J., Dabat, M. H., Rakotoarisoa, J., Rakotondraibe, J., & Rabeharisoa, L. (2012). Urban agriculture and land use in cities: An approach with the multi-functionality and sustainability concepts in the case of Antananarivo (Madagascar). Land Use Policy, 29(2), 429–439. https://doi.org/10.1016/j.landusepol.2011.08.009

Avtar, R., Aggarwal, R., Kharrazi, A., Kumar, P., & Kurniawan, T. A. (2020). Utilizing geospatial information to implement SDGs and monitor their Progress. In Environmental Monitoring and Assessment (Vol. 192, Issue 1, pp. 1–21). Springer. https://doi.org/10.1007/s10661-019-7996-9

Ayambire, R. A., Amponsah, O., Peprah, C., & Takyi, S. A. (2019). A review of practices for sustaining urban and peri-urban agriculture: Implications for land use planning in rapidly urbanising Ghanaian cities. Land Use Policy, 84, 260–277. https://doi.org/10.1016/j.landusepol.2019.03.004

Basso, B., Hyndman, D. W., Kendall, A. D., Grace, P. R., & Robertson, G. P. (2015). Can Impacts of Climate Change and Agricultural Adaptation Strategies Be Accurately Quantified if Crop Models Are Annually Re-Initialized? PLOS ONE, 10(6), e0127333. https://doi.org/10.1371/journal.pone.0127333

Beillouin, D., Ben-Ari, T., & Makowski, D. (2019). Evidence map of crop diversification strategies at the global scale. Environmental Research Letters, 14(12), 123001. https://doi.org/10.1088/1748-9326/AB4449

Boivin, N., & Crowther, A. (2021). Mobilizing the past to shape a better Anthropocene. Nature Ecology & Evolution 2021, 1–12. https://doi.org/10.1038/s41559-020-01361-4

Boix, L. R., & Zinck, J. A. (2008). Land-use planning in the chaco plain (Burruyacú, Argentina). Part 1: Evaluating land-use options to support crop diversification in an agricultural frontier area using physical land evaluation. Environmental Management, 42(6), 1043–1063. https://doi.org/10.1007/s00267-008-9208-1

Bowen, K. J., Cradock-Henry, N. A., Koch, F., Patterson, J., Häyhä, T., Vogt, J., & Barbi, F. (2017). Implementing the "Sustainable Development Goals": towards addressing three key governance challenges—collective action, trade-offs, and accountability. Current Opinion in Environmental Sustainability, 26–27, 90–96. https://doi.org/10.1016/J.COSUST.2017.05.002

Boz, I. (2016). Effects of environmentally friendly agricultural land protection programs: Evidence from the Lake Seyfe area of Turkey. Journal of Integrative Agriculture, 15(8), 1903–1914. https://doi.org/10.1016/S2095-3119(15)61271-0

Brady, M., Sahrbacher, C., Kellermann, K., & Happe, K. (2012). An agent-based approach to modeling impacts of agricultural policy on land use, biodiversity and ecosystem services. Landscape Ecology, 27(9), 1363–1381. https://doi.org/10.1007/s10980-012-9787-3

Brevik, E. C., Calzolari, C., Miller, B. A., Pereira, P., Kabala, C., Baumgarten, A., & Jordán, A. (2016). Soil mapping, classification, and pedologic modeling: History and future directions. Geoderma, 264, 256–274. https://doi.org/10.1016/J.GEODERMA.2015.05.017

Brevik, E. C., Slaughter, L., Singh, B. R., Steffan, J. J., Collier, D., Barnhart, P., & Pereira, P. (2020). Soil and Human Health: Current Status and Future Needs: Https://Doi.Org/10.1177/1178622120934441, 13. https://doi.org/10.1177/1178622120934441

Cao, M., Zhu, Y., Lü, G., Chen, M., & Qiao, W. (2019). Spatial Distribution of Global Cultivated Land and Its Variation between 2000 and 2010, from Both Agro-Ecological and Geopolitical Perspectives. Sustainability, 11(5), 1242. https://doi.org/10.3390/su11051242

Cassidy, E. S., West, P. C., Gerber, J. S., & Foley, J. A. (2013). Redefining agricultural yields: from tonnes to people nourished per hectare. Environmental Research Letters, 8(3), 034015. https://doi.org/10.1088/1748-9326/8/3/034015

Cerrada-Serra, P., Colombo, L., Ortiz-Miranda, D., & Grando, S. (2018). Access to agricultural land in peri-urban spaces: social mobilisation and institutional frameworks in Rome and Valencia. Food Security, 10(6), 1325–1336. https://doi.org/10.1007/s12571-018-0854-8

Chibwana, C., Fisher, M., & Shively, G. (2012). Cropland Allocation Effects of Agricultural Input Subsidies in Malawi. World Development, 40(1), 124–133. https://doi.org/10.1016/j.worlddev.2011.04.022

Dachraoui, M., & Sombrero, A. (2020). Effect of tillage systems and different rates of nitrogen fertilisation on the carbon footprint of irrigated maize in a semiarid area of Castile and Leon, Spain. Soil and Tillage Research, 196, 104472. https://doi.org/10.1016/J.STILL.2019.104472

Das, T. K., Bhattacharyya, R., Sudhishri, S., Sharma, A. R., Saharawat, Y. S., Bandyopadhyay, K. K., Sepat, S., Bana, R. S., Aggarwal, P., Sharma, R. K., Bhatia, A., Singh, G., Datta, S. P., Kar, A., Singh, B., Singh, P., Pathak, H., Vyas, A. K., & Jat, M. L. (2014). Conservation agriculture in an irrigated cotton-wheat system of the western Indo-Gangetic Plains: Crop and water productivity and economic profitability. Field Crops Research, 158, 24–33. https://doi.org/10.1016/j.fcr.2013.12.017

DeClerck, F. A. J., Jones, S. K., Attwood, S., Bossio, D., Girvetz, E., Chaplin-Kramer, B., Enfors, E., Fremier, A. K., Gordon, L. J., Kizito, F., Lopez Noriega, I., Matthews, N., McCartney, M., Meacham, M., Noble, A., Quintero, M., Remans, R., Soppe, R., Willemen, L., … Zhang, W. (2016). Agricultural ecosystems and their services: the vanguard of sustainability? In Current Opinion in Environmental Sustainability (Vol. 23, pp. 92–99). Elsevier B.V. https://doi.org/10.1016/j.cosust.2016.11.016

Drechsler, M., Touza, J., White, P. C. L., & Jones, G. (2016). Agricultural landscape structure and invasive species: the cost-effective level of crop field clustering. Food Security 2016 8:1, 8(1), 111–121. https://doi.org/10.1007/S12571-015-0539-5

EEA. (2017). Food in a green light - A systems approach to sustainable food.

Eigenbrod, C., & Gruda, N. (2015). Urban vegetable for food security in cities. A review. In Agronomy for Sustainable Development (Vol. 35, Issue 2, pp. 483–498). Springer-Verlag France. https://doi.org/10.1007/s13593-014-0273-y

Eklund, L., Degerald, M., Brandt, M., Prishchepov, A. V., & Pilesjö, P. (2017). How conflict affects land use: Agricultural activity in areas seized by the Islamic State. Environmental Research Letters, 12(5), 054004. https://doi.org/10.1088/1748-9326/aa673a

Ericksen, P. J. (2008). Conceptualizing food systems for global environmental change research. Global Environmental Change, 18(1), 234–245. https://doi.org/10.1016/j.gloenvcha.2007.09.002

Fanzo, J., Davis, C., McLaren, R., & Choufani, J. (2018). The effect of climate change across food systems: Implications for nutrition outcomes. In Global Food Security (Vol. 18, pp. 12–19). Elsevier B.V. https://doi.org/10.1016/j.gfs.2018.06.001

FAO. (2011). The State of the World's Land and Water Resources for Food and Agriculture (SOLAW)—Managing Systems at Risk.

FAO. (2017). The Future of Food and Agriculture - Trends and Challenges.

FAO. (2020). The State of Agricultural Commodity Markets 2020. In The State of Agricultural Commodity Markets 2020. FAO. https://doi.org/10.4060/cb0665en

FAO. (2021). Crop Prospects and Food Situation No.1, 03.

FAO, IFAD, UNICEF, WFP, & WHO. (2020). The State of Food Security and Nutrition in the World 2020. Transforming food systems for affordable healthy diets.

Feng, J., Li, F., Zhou, X., Xu, C., Ji, L., Chen, Z., & Fang, F. (2018). Impact of agronomy practices on the effects of reduced tillage systems on CH4 and N2O emissions from agricultural fields: A global meta-analysis. PLOS ONE, 13(5), e0196703. https://doi.org/10.1371/JOURNAL.PONE.0196703

Ferreira, A. J. D., Guilherme, R. I. M. M., Ferreira, C. S. S., & Oliveira, M. de F. M. L. de. (2018). Urban agriculture, a tool towards more resilient urban communities? In Current Opinion in

Environmental Science and Health (Vol. 5, pp. 93–97). Elsevier B.V. https://doi.org/10.1016/j.coesh.2018.06.004

Foguesatto, C. R., Borges, J. A. R., & Machado, J. A. D. (2020). A review and some reflections on farmers' adoption of sustainable agricultural practices worldwide. In Science of the Total Environment (Vol. 729, p. 138831). Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2020.138831

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., … Zaks, D. P. M. (2011). Solutions for a cultivated planet. Nature, 478(7369), 337–342. https://doi.org/10.1038/nature10452

Garrett, R. D., Koh, I., Lambin, E. F., le Polain de Waroux, Y., Kastens, J. H., & Brown, J. C. (2018). Intensification in agriculture-forest frontiers: Land use responses to development and conservation policies in Brazil. Global Environmental Change, 53, 233–243. https://doi.org/10.1016/j.gloenvcha.2018.09.011

Geng, S., Shi, P., Zong, N., & Zhu, W. (2019). Agricultural Land Suitability of Production Space in the Taihang Mountains, China. Chinese Geographical Science, 29(6), 1024–1038. https://doi.org/10.1007/s11769-019-1075-6

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. In Science (Vol. 327, Issue 5967, pp. 812–818). American Association for the Advancement of Science. https://doi.org/10.1126/science.1185383

Gomes, E., Inácio, M., Bogdzevič, K., Kalinauskas, M., Karnauskaitė, D., & Pereira, P. (2021). Future land-use changes and its impacts on terrestrial ecosystem services: A review. Science of The Total Environment, 781, 146716. https://doi.org/10.1016/J.SCITOTENV.2021.146716

Goucher, L., Bruce, R., Cameron, D. D., Lenny Koh, S. C., & Horton, P. (2017). The environmental impact of fertilizer embodied in a wheat-to-bread supply chain. Nature Plants, 3(3), 1–5. https://doi.org/10.1038/nplants.2017.12

Griggs, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M. C., Shyamsundar, P., Steffen, W., Glaser, G., Kanie, N., & Noble, I. (2013). Policy: Sustainable development goals for people and planet. In Nature (Vol. 495, Issue 7441, pp. 305–307). Nature Publishing Group. https://doi.org/10.1038/495305a

Grundy, M. J., Bryan, B. A., Nolan, M., Battaglia, M., Hatfield-Dodds, S., Connor, J. D., & Keating, B. A. (2016). Scenarios for Australian agricultural production and land use to 2050. Agricultural Systems, 142, 70–83. https://doi.org/10.1016/j.agsy.2015.11.008

Gunda, T., Bazuin, J. T., Nay, J., & Yeung, K. L. (2017). Impact of seasonal forecast use on agricultural income in a system with varying crop costs and returns: An empirically-grounded simulation. Environmental Research Letters, 12(3), 034001. https://doi.org/10.1088/1748-9326/aa5ef7

Guyomard, H., Darcy-Vrillon, B., Esnouf, C., Marin, M., Russel, M., & Guillou, M. (2012). Eating patterns and food systems: critical knowledge requirements for policy design and implementation. Agriculture & Food Security 2012 1:1, 1(1), 1–21. https://doi.org/10.1186/2048-7010-1-13

Hall, C., Dawson, T. P., Macdiarmid, J. I., Matthews, R. B., & Smith, P. (2017). The impact of population growth and climate change on food security in Africa: looking ahead to 2050. Https://Doi.Org/10.1080/14735903.2017.1293929, 15(2), 124–135. https://doi.org/10.1080/14735903.2017.1293929

Hao, H., Li, X., Tan, M., Zhang, J., & Zhang, H. (2015). Agricultural land use intensity and its determinants: A case study in Taibus Banner, Inner Mongolia, China. Frontiers of Earth Science, 9(2), 308–318. https://doi.org/10.1007/s11707-014-0471-6

Hatab, A. A., Cavinato, M. E. R., & Lagerkvist, C. J. (2019). Urbanization, livestock systems and food security in developing countries: A systematic review of the literature. Food Security, 11(2), 279–299. https://doi.org/10.1007/s12571-019-00906-1

Hong, C., Jin, X., Ren, J., Gu, Z., & Zhou, Y. (2019). Satellite data indicates multidimensional variation of agricultural production in land consolidation area. Science of the Total Environment, 653, 735–747. https://doi.org/10.1016/j.scitotenv.2018.10.415

Huang, Y., Ren, W., Wang, L., Hui, D., Grove, J. H., Yang, X., Tao, B., & Goff, B. (2018). Greenhouse gas emissions and crop yield in no-tillage systems: A meta-analysis. Agriculture, Ecosystems & Environment, 268, 144–153. https://doi.org/10.1016/J.AGEE.2018.09.002

Im, E. S., Pal, J. S., & Eltahir, E. A. B. (2017). Deadly heat waves projected in the densely populated agricultural regions of South Asia. Science Advances, 3(8), e1603322. https://doi.org/10.1126/sciadv.1603322

Ingram, J., Ajates, R., Arnall, A., Blake, L., Borrelli, R., Collier, R., de Frece, A., Häsler, B., Lang, T., Pope, H., Reed, K., Sykes, R., Wells, R., & White, R. (2020). A future workforce of food-system analysts. Nature Food, 1(1), 9–10. https://doi.org/10.1038/s43016-019-0003-3

IPCC. (2014). Food security and food production systems. In J. R. Porter, L. Xie, A. J. Challinor, K. Cochrane, S. M. Howden, M. M. Iqbal, D. B. Lobell, & M. I. Travasso (Eds.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., Barros, V.R., Dokken, D.J., Ma.

Islam, S. F. ul, Sander, B. O., Quilty, J. R., de Neergaard, A., van Groenigen, J. W., & Jensen, L. S. (2020). Mitigation of greenhouse gas emissions and reduced irrigation water use in rice production through water-saving irrigation scheduling, reduced tillage and fertiliser application strategies. Science of The Total Environment, 739, 140215. https://doi.org/10.1016/J.SCITOTENV.2020.140215

Jat, H. S., Datta, A., Choudhary, M., Yadav, A. K., Choudhary, V., Sharma, P. C., Gathala, M. K., Jat, M. L., & McDonald, A. (2019). Effects of tillage, crop establishment and diversification on soil organic carbon, aggregation, aggregate associated carbon and productivity in cereal systems of semi-arid Northwest India. Soil and Tillage Research, 190, 128–138. https://doi.org/10.1016/J.STILL.2019.03.005

Jin, X., Shao, Y., Zhang, Z., Resler, L. M., Campbell, J. B., Chen, G., & Zhou, Y. (2017). The evaluation of land consolidation policy in improving agricultural productivity in China. Scientific Reports, 7(2792).

Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2016). Brief history of agricultural systems modeling. Agricultural Systems, 155, 240–254. https://doi.org/10.1016/J.AGSY.2016.05.014

Kalkura, P., Puneeth, R. B., Suhas, K. N., Surya, & Ramyashree, M. (2021). Pest Control Management System using Organic Pesticides. Global Transitions Proceedings. https://doi.org/10.1016/J.GLTP.2021.08.058

Kasperson, J. X., & Kasperson, R. E. (2001). International workshop on vulnerability and global environmental change. 17–19 May 2001, Stockholm Environment Institute (SEI), Stockholm, Sweden. A Workshop Summary Prepared on Behalf of Workshop Participants. Report 2001-01. Stockholm, Stockholm Environment Institute. SEI Risk and Vulnerability Programme.

Kastner, T., Rivas, M. J. I., Koch, W., & Nonhebel, S. (2012). Global changes in diets and the consequences for land requirements for food. Proceedings of the National Academy of Sciences of the United States of America, 109(18), 6868–6872. https://doi.org/10.1073/pnas.1117054109

Kazemi, H., Sadeghi, S., & Akinci, H. (2016). Developing a land evaluation model for faba bean cultivation using geographic information system and multi-criteria analysis (A case study: Gonbad-Kavous region, Iran). Ecological Indicators, 63, 37–47. https://doi.org/10.1016/j.ecolind.2015.11.021

Kim, K. H., Kabir, E., & Jahan, S. A. (2017). Exposure to pesticides and the associated human health effects. Science of The Total Environment, 575, 525–535. https://doi.org/10.1016/J.SCITOTENV.2016.09.009

Koutsos, T. M., Menexes, G. C., & Dordas, C. A. (2019). An efficient framework for conducting systematic literature reviews in agricultural sciences. In Science of the Total Environment (Vol. 682, pp. 106–117). Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2019.04.354

Kühling, I., Broll, G., & Trautz, D. (2016). Spatio-temporal analysis of agricultural land-use intensity across the Western Siberian grain belt. Science of the Total Environment, 544, 271–280. https://doi.org/10.1016/j.scitotenv.2015.11.129

Kukal, M. S., & Irmak, S. (2018). Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production. Scientific Reports, 8(1), 1–18. https://doi.org/10.1038/s41598-018-21848-2

Kvakkestad, V., Rørstad, P. K., & Vatn, A. (2015). Norwegian farmers' perspectives on agriculture and agricultural payments: Between productivism and cultural landscapes. Land Use Policy, 42, 83–92. https://doi.org/10.1016/j.landusepol.2014.07.009

Lampridi, M. G., Sørensen, C. G., & Bochtis, D. (2019). Agricultural sustainability: A review of concepts and methods. Sustainability (Switzerland), 11(18). https://doi.org/10.3390/su11185120

Leng, G., & Hall, J. (2019). Crop yield sensitivity of global major agricultural countries to droughts and the projected changes in the future. Science of the Total Environment, 654, 811–821. https://doi.org/10.1016/j.scitotenv.2018.10.434

Leonardo, W. J., van de Ven, G. W. J., Udo, H., Kanellopoulos, A., Sitoe, A., & Giller, K. E. (2015). Labour not land constrains agricultural production and food self-sufficiency in maize-based smallholder farming systems in Mozambique. Food Security, 7(4), 857–874. https://doi.org/10.1007/s12571-015-0480-7

Li, G., Messina, J. P., Peter, B. G., & Snapp, S. S. (2017). Mapping Land Suitability for Agriculture in Malawi. Land Degradation & Development, 28(7), 2001–2016. https://doi.org/10.1002/ldr.2723

Li, M., Xu, J., Gao, Z., Tian, H., Gao, Y., & Kariman, K. (2020). Genetically modified crops are superior in their nitrogen use efficiency-A meta-analysis of three major cereals. Scientific Reports, 10(1), 1–9. https://doi.org/10.1038/s41598-020-65684-9

Li, W., Wang, D., Liu, S., & Zhu, Y. (2019). Measuring urbanization-occupation and internal conversion of peri-urban cultivated land to determine changes in the peri-urban agriculture of the black soil region. Ecological Indicators, 102, 328–337. https://doi.org/10.1016/j.ecolind.2019.02.055

Li, Y., Yang, X., Cai, H., Xiao, L., Xu, X., & Liu, L. (2014). Topographical Characteristics of Agricultural Potential Productivity during Cropland Transformation in China. Sustainability, 7(1), 96–110. https://doi.org/10.3390/su7010096

Lin, B. B., Philpott, S. M., & Jha, S. (2015). The future of urban agriculture and biodiversity-ecosystem services: Challenges and next steps. Basic and Applied Ecology, 16(3), 189–201. https://doi.org/10.1016/J.BAAE.2015.01.005

Lin, H. C., & Hülsbergen, K. J. (2017). A new method for analyzing agricultural land-use efficiency, and its application in organic and conventional farming systems in southern Germany. European Journal of Agronomy, 83, 15–27. https://doi.org/10.1016/j.eja.2016.11.003

Looga, J., Jürgenson, E., Sikk, K., Matveev, E., & Maasikamäe, S. (2018). Land fragmentation and other determinants of agricultural farm productivity: The case of Estonia. Land Use Policy, 79, 285–292. https://doi.org/10.1016/j.landusepol.2018.08.021

Lwasa, S., Mugagga, F., Wahab, B., Simon, D., Connors, J. P., & Griffith, C. (2015). A meta-analysis of urban and peri-urban agriculture and forestry in mediating climate change. In Current Opinion in Environmental Sustainability (Vol. 13, pp. 68–73). Elsevier B.V. https://doi.org/10.1016/j.cosust.2015.02.003

Lyle, G., Bryan, B. A., & Ostendorf, B. (2015). Identifying the spatial and temporal variability of economic opportunity costs to promote the adoption of alternative land uses in grain growing agricultural areas: An Australian example. Journal of Environmental Management, 155, 123–135. https://doi.org/10.1016/j.jenvman.2015.02.006

Mackay, H. (2018). Mapping and characterising the urban agricultural landscape of two intermediate-sized Ghanaian cities. Land Use Policy, 70, 182–197. https://doi.org/10.1016/j.landusepol.2017.10.031

Malek, Ž., Douw, B., Van Vliet, J., Van Der Zanden, E. H., & Verburg, P. H. (2019). Local land-use decision-making in a global context. In Environmental Research Letters (Vol. 14, Issue 8, p. 083006). Institute of Physics Publishing. https://doi.org/10.1088/1748-9326/ab309e

Manners, R., & van Etten, J. (2018). Are agricultural researchers working on the right crops to enable food and nutrition security under future climates? Global Environmental Change, 53, 182–194. https://doi.org/10.1016/j.gloenvcha.2018.09.010

Marsden, T., & Morley, A. (2015). Sustainable Food Systems: Building a New Paradigm (T. Marsden & A. Morley (Eds.)). Routledge.

Martellozzo, F., Amato, F., Murgante, B., & Clarke, K. C. (2018). Modelling the impact of urban growth on agriculture and natural land in Italy to 2030. Applied Geography, 91, 156–167. https://doi.org/10.1016/j.apgeog.2017.12.004

Mason-D'Croz, D., Sulser, T. B., Wiebe, K., Rosegrant, M. W., Lowder, S. K., Nin-Pratt, A., Willenbockel, D., Robinson, S., Zhu, T., Cenacchi, N., Dunston, S., & Robertson, R. D. (2019). Agricultural investments and hunger in Africa modeling potential contributions to SDG2 – Zero Hunger. World Development, 116, 38–53. https://doi.org/10.1016/j.worlddev.2018.12.006

Mbata, J. N. (2001). Land use practices in Lesotho: Implications for sustainability in agricultural production. Journal of Sustainable Agriculture, 18(2–3), 5–24. https://doi.org/10.1300/J064v18n02_03

Mendas, A., & Delali, A. (2012). Integration of MultiCriteria Decision Analysis in GIS to develop land suitability for agriculture: Application to durum wheat cultivation in the region of Mleta in Algeria. Computers and Electronics in Agriculture, 83, 117–126. https://doi.org/10.1016/j.compag.2012.02.003

Mesgaran, M. B., Madani, K., Hashemi, H., & Azadi, P. (2017). Iran's Land Suitability for Agriculture. Scientific Reports, 7(1), 1–12. https://doi.org/10.1038/s41598-017-08066-y

Michel-Villarreal, R., Hingley, M., Canavari, M., & Bregoli, I. (2019). Sustainability in Alternative Food Networks: A Systematic Literature Review. Sustainability, 11(3), 859. https://doi.org/10.3390/su11030859

Millennium Ecosystem Assessment. (2005). Ecosystems and Human Well-being: Synthesis. Island Press.

Moher, D. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Annals of Internal Medicine, 151(4), 264. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Monteleone, M., Cammerino, A. R. B., & Libutti, A. (2018). Agricultural "greening" and cropland diversification trends: Potential contribution of agroenergy crops in Capitanata (South Italy). Land Use Policy, 70, 591–600. https://doi.org/10.1016/j.landusepol.2017.10.038

Montgomery, B., Dragićević, S., Dujmović, J., & Schmidt, M. (2016). A GIS-based Logic Scoring of Preference method for evaluation of land capability and suitability for agriculture. Computers and Electronics in Agriculture, 124, 340–353. https://doi.org/10.1016/j.compag.2016.04.013

Mori, S., Kato, M., & Ido, T. (2010). GISELA - GIS-based evaluation of land use and agriculture market analysis under global warming. Applied Energy, 87(1), 236–242. https://doi.org/10.1016/j.apenergy.2009.06.013

Mougeot, L. J. A. (2011). International support to Research and Policy on Urban Agriculture (1996-2010): achievements and challenges. Urban Agriculture Magazine, 12–17.

Moyer, J. D., & Bohl, D. K. (2019). Alternative pathways to human development: Assessing trade-offs and synergies in achieving the Sustainable Development Goals. Futures, 105, 199–210. https://doi.org/10.1016/J.FUTURES.2018.10.007

Müller, B., Hoffmann, F., Heckelei, T., Müller, C., Hertel, T. W., Polhill, J. G., van Wijk, M., Achterbosch, T., Alexander, P., Brown, C., Kreuer, D., Ewert, F., Ge, J., Millington, J. D. A., Seppelt, R., Verburg, P. H., & Webber, H. (2020). Modelling food security: Bridging the gap between the micro and the macro scale. Global Environmental Change, 63, 102085. https://doi.org/10.1016/j.gloenvcha.2020.102085

Murrell, E. G. (2017). Can agricultural practices that mitigate or improve crop resilience to climate change also manage crop pests? Current Opinion in Insect Science, 23, 81–88. https://doi.org/10.1016/J.COIS.2017.07.008

Musakwa, W. (2018). Identifying land suitable for agricultural land reform using GIS-MCDA in South Africa. Environment, Development and Sustainability, 20(5), 2281–2299. https://doi.org/10.1007/s10668-017-9989-6

Nakalembe, C., Dempewolf, J., & Justice, C. (2017). Agricultural land use change in Karamoja Region, Uganda. Land Use Policy, 62, 2–12. https://doi.org/10.1016/j.landusepol.2016.11.029

Nayak, M., Swain, D. K., & Sen, R. (2019). Strategic valorization of de-oiled microalgal biomass waste as biofertilizer for sustainable and improved agriculture of rice (Oryza sativa L.)crop. Science of the Total Environment, 682, 475–484. https://doi.org/10.1016/j.scitotenv.2019.05.123

Nicholls, E., Ely, A., Birkin, L., Basu, P., & Goulson, D. (2020). The contribution of small-scale food production in urban areas to the sustainable development goals: a review and case study. Sustainability Science 2020 15:6, 15(6), 1585–1599. https://doi.org/10.1007/S11625-020-00792-Z

Nkomoki, W., Bavorová, M., & Banout, J. (2018). Adoption of sustainable agricultural practices and food security threats: Effects of land tenure in Zambia. Land Use Policy, 78, 532–538. https://doi.org/10.1016/j.landusepol.2018.07.021

OECD/FAO. (2009). OECD-FAO Agricultural Outlook 2009-2018. OECD Publishing and FAO.

OECD. (2020). Towards Sustainable Land Use: Aligning Biodiversity, Climate and Food Policies. OECD Publishing.

Oerke, E. C. (2006). Crop losses to pests. In Journal of Agricultural Science (Vol. 144, Issue 1, pp. 31–43). Cambridge University Press. https://doi.org/10.1017/S0021859605005708

Ogle, J., Delparte, D., & Sanger, H. (2017). Quantifying the sustainability of urban growth and form through time: An algorithmic analysis of a city's development. Applied Geography, 88, 1–14. https://doi.org/10.1016/J.APGEOG.2017.08.016

Opitz, I., Berges, R., Piorr, A., & Krikser, T. (2016). Contributing to food security in urban areas: differences between urban agriculture and peri-urban agriculture in the Global North. Agriculture and Human Values, 33(2), 341–358. https://doi.org/10.1007/s10460-015-9610-2

Pereira, P., Bogunovic, I., Muñoz-Rojas, M., & Brevik, E. C. (2018). Soil ecosystem services, sustainability, valuation and management. In Current Opinion in Environmental Science and Health (Vol. 5, pp. 7–13). Elsevier B.V. https://doi.org/10.1016/j.coesh.2017.12.003

Pereira, P., Brevik, E., & Trevisani, S. (2018). Mapping the environment. In Science of the Total Environment (Vols 610–611, pp. 17–23). Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2017.08.001

Piiroinen, R., Heiskanen, J., Mõttus, M., & Pellikka, P. (2015). Classification of crops across heterogeneous agricultural landscape in Kenya using AisaEAGLE imaging spectroscopy data. International Journal of Applied Earth Observation and Geoinformation, 39, 1–8. https://doi.org/10.1016/j.jag.2015.02.005

Piquer-Rodríguez, M., Butsic, V., Gärtner, P., Macchi, L., Baumann, M., Gavier Pizarro, G., Volante, J. N., Gasparri, I. N., & Kuemmerle, T. (2018). Drivers of agricultural land-use change in the Argentine Pampas and Chaco regions. Applied Geography, 91, 111–122. https://doi.org/10.1016/j.apgeog.2018.01.004

Popp, J., Lakner, Z., Harangi-Rákos, M., & Fári, M. (2014). The effect of bioenergy expansion: Food, energy, and environment. Renewable and Sustainable Energy Reviews, 32, 559–578. https://doi.org/10.1016/j.rser.2014.01.056

Popp, József, Pető, K., & Nagy, J. (2013). Pesticide productivity and food security. A review. Agronomy for Sustainable Development, 33(1), 243–255. https://doi.org/10.1007/s13593-012-0105-x

Pothukuchi, K. (2018). Vacant land disposition for agriculture in Cleveland, Ohio: Is community development a mixed blessing? Journal of Urban Affairs, 40(5), 657–678. https://doi.org/10.1080/07352166.2017.1403855

Potter, C., & Urquhart, J. (2017). Tree disease and pest epidemics in the Anthropocene: A review of the drivers, impacts and policy responses in the UK. Forest Policy and Economics, 79, 61–68. https://doi.org/10.1016/j.forpol.2016.06.024

Radwan, T. M., Blackburn, G. A., Whyatt, J. D., & Atkinson, P. M. (2019). Dramatic Loss of Agricultural Land Due to Urban Expansion Threatens Food Security in the Nile Delta, Egypt. Remote Sensing, 11(3), 332. https://doi.org/10.3390/rs11030332

Raman, R. (2017). The impact of Genetically Modified (GM) crops in modern agriculture: A review. Https://Doi.Org/10.1080/21645698.2017.1413522, 8(4), 195–208. https://doi.org/10.1080/21645698.2017.1413522

Ricart, S., & Rico, A. M. (2019). Assessing technical and social driving factors of water reuse in agriculture: A review on risks, regulation and the yuck factor. Agricultural Water Management, 217, 426–439. https://doi.org/10.1016/J.AGWAT.2019.03.017

Richardson, K. J., Lewis, K. H., Krishnamurthy, P. K., Kent, C., Wiltshire, A. J., & Hanlon, H. M. (2018). Food security outcomes under a changing climate: impacts of mitigation and adaptation on vulnerability to food insecurity. Climatic Change 2018 147:1, 147(1), 327–341. https://doi.org/10.1007/S10584-018-2137-Y

Rosenzweig, C., Iglesias, A., Yang, X. B., Epstein, P. R., & Chivian, E. (2001). Climate Change and Extreme Weather Events; Implications for Food Production, Plant Diseases, and Pests. Global Change and Human Health 2001 2:2, 2(2), 90–104. https://doi.org/10.1023/A:1015086831467

Rosenzweig, C., Mbow, C., Barioni, L. G., Benton, T. G., Herrero, M., Krishnapillai, M., Liwenga, E. T., Pradhan, P., Rivera-Ferre, M. G., Sapkota, T., Tubiello, F. N., Xu, Y., Mencos Contreras, E., & Portugal-Pereira, J. (2020). Climate change responses benefit from a global food system approach. Nature Food 2020 1:2, 1(2), 94–97. https://doi.org/10.1038/s43016-020-0031-z

Ruben, R., Verhagen, J., & Plaisier, C. (2018). The Challenge of Food Systems Research: What Difference Does It Make? Sustainability, 11(1), 171. https://doi.org/10.3390/su11010171

Saha, M., & Eckelman, M. J. (2017). Growing fresh fruits and vegetables in an urban landscape: A geospatial assessment of ground level and rooftop urban agriculture potential in Boston, USA. Landscape and Urban Planning, 165, 130–141. https://doi.org/10.1016/j.landurbplan.2017.04.015

Samasse, K., Hanan, N., Tappan, G., & Diallo, Y. (2018). Assessing Cropland Area in West Africa for Agricultural Yield Analysis. Remote Sensing, 10(11), 1785. https://doi.org/10.3390/rs10111785

Sánchez-Bayo, F., & Wyckhuys, K. A. G. (2019). Worldwide decline of the entomofauna: A review of its drivers. In Biological Conservation (Vol. 232, pp. 8–27). Elsevier Ltd. https://doi.org/10.1016/j.biocon.2019.01.020

Santo, R., Palmer, A., & Kim, B. (2016). Vacant Lots to Vibrant Plots: A Review of the Benefits and Limitations of Urban Agriculture.

Sattler, C., Gianuca, A. T., Schweiger, O., Franzén, M., & Settele, J. (2020). Pesticides and land cover heterogeneity affect functional group and taxonomic diversity of arthropods in rice agroecosystems. Agriculture, Ecosystems & Environment, 297, 106927. https://doi.org/10.1016/J.AGEE.2020.106927

Scherer, L. A., Verburg, P. H., & Schulp, C. J. E. (2018). Opportunities for sustainable intensification in European agriculture. Global Environmental Change, 48, 43–55. https://doi.org/10.1016/j.gloenvcha.2017.11.009

Schiefer, J., Lair, G. J., & Blum, W. E. H. (2016). Potential and limits of land and soil for sustainable intensification of European agriculture. Agriculture, Ecosystems and Environment, 230, 283–293. https://doi.org/10.1016/j.agee.2016.06.021

Schmidhuber, J., & Tubiello, F. N. (2007). Global food security under climate change. In Proceedings of the National Academy of Sciences of the United States of America (Vol. 104, Issue 50, pp. 19703–19708). National Academy of Sciences. https://doi.org/10.1073/pnas.0701976104

Schneider, U. A., Havlík, P., Schmid, E., Valin, H., Mosnier, A., Obersteiner, M., Böttcher, H., Skalský, R., Balkovič, J., Sauer, T., & Fritz, S. (2011). Impacts of population growth, economic development, and technical change on global food production and consumption. Agricultural Systems, 104(2), 204–215. https://doi.org/10.1016/j.agsy.2010.11.003

Schreinemachers, P., & Tipraqsa, P. (2012). Agricultural pesticides and land use intensification in high, middle and low income countries. Food Policy, 37(6), 616–626. https://doi.org/10.1016/j.foodpol.2012.06.003

Schütte, G., Eckerstorfer, M., Rastelli, V., Reichenbecher, W., Restrepo-Vassalli, S., Ruohonen-Lehto, M., Saucy, A.-G. W., & Mertens, M. (2017). Herbicide resistance and biodiversity: agronomic and environmental aspects of genetically modified herbicide-resistant plants. Environmental Sciences Europe 2017 29:1, 29(1), 1–12. https://doi.org/10.1186/S12302-016-0100-Y

Scown, M. W., Winkler, K. J., & Nicholas, K. A. (2019). Aligning research with policy and practice for sustainable agricultural land systems in Europe. Proceedings of the National Academy of Sciences of the United States of America, 116(11), 4911–4916. https://doi.org/10.1073/pnas.1812100116

She, X., Zhang, L., Cen, Y., Wu, T., Huang, C., & Baig, M. H. A. (2015). Comparison of the continuity of vegetation indices derived from Landsat 8 OLI and Landsat 7 ETM+ data among different vegetation types. Remote Sensing, 7(10), 13485–13506. https://doi.org/10.3390/rs71013485

Shen, J., Cui, Z., Miao, Y., Mi, G., Zhang, H., Fan, M., Zhang, C., Jiang, R., Zhang, W., Li, H., Chen, X., Li, X., & Zhang, F. (2013). Transforming agriculture in China: From solely high yield to both high yield and high resource use efficiency. In Global Food Security (Vol. 2, Issue 1, pp. 1–8). Elsevier. https://doi.org/10.1016/j.gfs.2012.12.004

Shi, K., Chen, Y., Yu, B., Xu, T., Li, L., Huang, C., Liu, R., Chen, Z., & Wu, J. (2016). Urban Expansion and Agricultural Land Loss in China: A Multiscale Perspective. Sustainability, 8(8), 790. https://doi.org/10.3390/su8080790

Siegner, A., Sowerwine, J., & Acey, C. (2018). Does urban agriculture improve food security? Examining the nexus of food access and distribution of urban produced foods in the United States: A systematic review. In Sustainability (Switzerland) (Vol. 10, Issue 9, p. 2988). MDPI AG. https://doi.org/10.3390/su10092988

Sioen, G., Terada, T., Sekiyama, M., & Yokohari, M. (2018). Resilience with Mixed Agricultural and Urban Land Uses in Tokyo, Japan. Sustainability, 10(2), 435. https://doi.org/10.3390/su10020435

Spence, N., Hill, L., & Morris, J. (2020). How the global threat of pests and diseases impacts plants, people, and the planet. Plants, People, Planet, 2(1), 5–13. https://doi.org/10.1002/PPP3.10088

Steinfeld, H., Gerber, P., Wassenaar, T., Castel, V., Rosales, M., & de Haan, C. (2006). Livestock's Long Shadow: Environmental Issues and Options. Food and Agriculture Organization of the United Nations (FAO).

Stephens, E. C., Jones, A. D., & Parsons, D. (2018). Agricultural systems research and global food security in the 21st century: An overview and roadmap for future opportunities. Agricultural Systems, 163, 1–6. https://doi.org/10.1016/j.agsy.2017.01.011

Stoll-Kleemann, S., & Schmidt, U. J. (2017). Reducing meat consumption in developed and transition countries to counter climate change and biodiversity loss: a review of influence factors. Regional Environmental Change, 17(5), 1261–1277. https://doi.org/10.1007/S10113-016-1057-5

Struik, P. C., & Kuyper, T. W. (2017). Sustainable intensification in agriculture: the richer shade of green. A review. In Agronomy for Sustainable Development (Vol. 37, Issue 5, pp. 1–15). Springer-Verlag France. https://doi.org/10.1007/s13593-017-0445-7

Sun, Z., You, L., & Müller, D. (2018). Synthesis of agricultural land system change in China over the past 40 years. Journal of Land Use Science, 13(5), 473–479. https://doi.org/10.1080/1747423X.2019.1571120

TerAvest, D., Wandschneider, P. R., Thierfelder, C., & Reganold, J. P. (2019). Diversifying conservation agriculture and conventional tillage cropping systems to improve the wellbeing of smallholder farmers in Malawi. Agricultural Systems, 171, 23–35. https://doi.org/10.1016/j.agsy.2019.01.004

The Royal Society. (2009). Reaping the benefits. Science and the sustainable intensification of global agriculture.

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences of the United States of America, 108(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

Tomlinson, I. (2013). Doubling food production to feed the 9 billion: A critical perspective on a key discourse of food security in the UK. Journal of Rural Studies, 29, 81–90. https://doi.org/10.1016/j.jrurstud.2011.09.001

Torbick, N., Chowdhury, D., Salas, W., & Qi, J. (2017). Monitoring Rice Agriculture across Myanmar Using Time Series Sentinel-1 Assisted by Landsat-8 and PALSAR-2. Remote Sensing, 9(2), 119. https://doi.org/10.3390/rs9020119

Tornaghi, C. (2014). Critical geography of urban agriculture. Progress in Human Geography, 38(4), 551–567. https://doi.org/10.1177/0309132513512542

Tóth, G., Hermann, T., da Silva, M. R., & Montanarella, L. (2018). Monitoring soil for sustainable development and land degradation neutrality. Environmental Monitoring and Assessment 2018 190:2, 190(2), 1–4. https://doi.org/10.1007/S10661-017-6415-3

Tsatsakis, A. M., Nawaz, M. A., Kouretas, D., Balias, G., Savolainen, K., Tutelyan, V. A., Golokhvast, K. S., Lee, J. D., Yang, S. H., & Chung, G. (2017). Environmental impacts of genetically modified plants: A review. Environmental Research, 156, 818–833. https://doi.org/10.1016/J.ENVRES.2017.03.011

Tsuchiya, K., Hara, Y., & Thaitakoo, D. (2015). Linking food and land systems for sustainable peri-urban agriculture in Bangkok Metropolitan Region. Landscape and Urban Planning, 143, 192–204. https://doi.org/10.1016/j.landurbplan.2015.07.008

Tyson, P., Steffen, W., Mitra, A., Fu, C., & Lebel, L. (2001). The earth system: regional–global linkages. Regional Environmental Change, 2, 128–140.

UNDP. (1996). Urban Agriculture: Food, Jobs, and Sustainable Cities. In UNDP Publication Series for Habitat II (Vol. 1). United Nations Development Programme.

United Nations. (2012). The Future We Want. Outcome document of the United Nations Conference on Sustainable Development.

United Nations. (2015a). The Millennium Development Goals Report 2015.

United Nations. (2015b). Transforming our world: the 2030 agenda for sustainable development. A/RES/70/1.

United Nations. (2019). World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420).

van de Kamp, M. E., van Dooren, C., Hollander, A., Geurts, M., Brink, E. J., van Rossum, C., Biesbroek, S., de Valk, E., Toxopeus, I. B., & Temme, E. H. M. (2018). Healthy diets with reduced environmental impact? – The greenhouse gas emissions of various diets adhering to the Dutch food based dietary guidelines. Food Research International, 104, 14–24. https://doi.org/10.1016/j.foodres.2017.06.006

van Esse, H. P., Reuber, T. L., & Does, D. van der. (2020). Genetic modification to improve disease resistance in crops. New Phytologist, 225(1), 70–86. https://doi.org/10.1111/NPH.15967

Van Ittersum, M. K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., & Hochman, Z. (2013). Yield gap analysis with local to global relevance-A review. Field Crops Research, 143, 4–17. https://doi.org/10.1016/j.fcr.2012.09.009

Van Meijl, H., Havlik, P., Lotze-Campen, H., Stehfest, E., Witzke, P., Domínguez, I. P., Bodirsky, B. L., Van Dijk, M., Doelman, J., Fellmann, T., Humpenöder, F., Koopman, J. F. L., Müller, C., Popp, A., Tabeau, A., Valin, H., & Van Zeist, W. J. (2018). Comparing impacts of climate change and mitigation on global agriculture by 2050. Environmental Research Letters, 13(6), 064021. https://doi.org/10.1088/1748-9326/aabdc4

van Noordwijk, M., Duguma, L. A., Dewi, S., Leimona, B., Catacutan, D. C., Lusiana, B., Öborn, I., Hairiah, K., & Minang, P. A. (2018). SDG synergy between agriculture and forestry in the food, energy, water and income nexus: reinventing agroforestry? Current Opinion in Environmental Sustainability, 34, 33–42. https://doi.org/10.1016/J.COSUST.2018.09.003

van Vliet, J., de Groot, H. L. F., Rietveld, P., & Verburg, P. H. (2015). Manifestations and underlying drivers of agricultural land use change in Europe. In Landscape and Urban Planning (Vol. 133, pp. 24–36). Elsevier B.V. https://doi.org/10.1016/j.landurbplan.2014.09.001

Vasile, A. J., Popescu, C., Ion, R. A., & Dobre, I. (2015). From conventional to organic in Romanian agriculture - Impact assessment of a land use changing paradigm. Land Use Policy, 46, 258–266. https://doi.org/10.1016/j.landusepol.2015.02.012

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J. F., & Ceschia, E. (2017). Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. Remote Sensing of Environment, 199, 415–426. https://doi.org/10.1016/j.rse.2017.07.015

Viana, C. M., & Rocha, J. (2020). Evaluating Dominant Land Use/Land Cover Changes and Predicting Future Scenario in a Rural Region Using a Memoryless Stochastic Method. Sustainability, 12(10), 4332. https://doi.org/10.3390/su12104332

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. Remote Sensing of Environment, 236, 111402. https://doi.org/10.1016/j.rse.2019.111402

World Bank Group. (2015). Ending Poverty and Hunger by 2030 An Agenda for the global food system.

Wu, W. Bin, Yu, Q. Y., Peter, V. H., You, L. Z., Yang, P., & Tang, H. J. (2014). How could agricultural land systems contribute to raise food production under global change? Journal of Integrative Agriculture, 13(7), 1432–1442. https://doi.org/10.1016/S2095-3119(14)60819-4

Yan, T., Wang, J., & Huang, J. (2015). Urbanization, agricultural water use, and regional and national crop production in China. Ecological Modelling, 318, 226–235. https://doi.org/10.1016/j.ecolmodel.2014.12.021

Yang, L., Liu, M., Lun, F., Yuan, Z., Zhang, Y., & Min, Q. (2017). An Analysis on Crops Choice and Its Driving Factors in Agricultural Heritage Systems—A Case of Honghe Hani Rice Terraces System. Sustainability, 9(7), 1162. https://doi.org/10.3390/su9071162

Yang, S., Zhao, W., Liu, Y., Cherubini, F., Fu, B., & Pereira, P. (2020a). Prioritizing sustainable development goals and linking them to ecosystem services: A global expert's knowledge evaluation. Geography and Sustainability, 1(4), 321–330. https://doi.org/10.1016/J.GEOSUS.2020.09.004

Yu, D., Wang, D., Li, W., Liu, S., Zhu, Y., Wu, W., & Zhou, Y. (2018). Decreased Landscape Ecological Security of Peri-Urban Cultivated Land Following Rapid Urbanization: An Impediment to Sustainable Agriculture. Sustainability, 10(2), 394. https://doi.org/10.3390/su10020394

Yu, M., Bambacus, M., Cervone, G., Clarke, K., Duffy, D., Huang, Q., Li, J., Li, W., Li, Z., Liu, Q., Resch, B., Yang, J., & Yang, C. (2020). Spatiotemporal event detection: a review. International Journal of Digital Earth, 13(12), 1339–1365. https://doi.org/10.1080/17538947.2020.1738569

Yu, Q., Wu, W., Yang, P., Li, Z., Xiong, W., & Tang, H. (2012). Proposing an interdisciplinary and cross-scale framework for global change and food security researches. Agriculture, Ecosystems and Environment, 156, 57–71. https://doi.org/10.1016/j.agee.2012.04.026

Yuan, M. (2021). Geographical information science for the United Nations' 2030 agenda for sustainable development. International Journal of Geographical Information Science, 35(1), 1–8. https://doi.org/10.1080/13658816.2020.1766244

Zabel, F., Putzenlechner, B., & Mauser, W. (2014). Global agricultural land resources - A high resolution suitability evaluation and its perspectives until 2100 under climate change conditions. In PLoS ONE (Vol. 9, Issue 9). Public Library of Science. https://doi.org/10.1371/journal.pone.0107522

Zezza, A., & Tasciotti, L. (2010). Urban agriculture, poverty, and food security: Empirical evidence from a sample of developing countries. Food Policy, 35(4), 265–273. https://doi.org/10.1016/j.foodpol.2010.04.007

Zuo, L., Zhang, Z., Carlson, K. M., MacDonald, G. K., Brauman, K. A., Liu, Y., Zhang, W., Zhang, H., Wu, W., Zhao, X., Wang, X., Liu, B., Yi, L., Wen, Q., Liu, F., Xu, J., Hu, S., Sun, F., Gerber, J. S., & West, P. C. (2018). Progress towards sustainable intensification in China challenged by land-use change. Nature Sustainability, 1(6), 304–313. https://doi.org/10.1038/s41893-018-0076-2

Part III. Dynamics of agricultural land: exploring remotely-sensed and volunteered geographic information data

Chapter 3. Long-Term Satellite Image Time-Series for Land Use/Land Cover Change Detection Using Refined Open Source Data in a Rural Region

Cláudia M. Viana, Inês Girão, Jorge Rocha

Article published at the Remote Sensing (2019), Volume 11, Part 9, 1104.

https://doi.org/10.3390/rs11091104

The structure and formatting are slightly adapted.

Abstract

The increasing availability and volume of remote sensing data, such as Landsat satellite images, have allowed the multidimensional analysis of land use/land cover (LULC) changes. However, the performance of image classification is highly dependent on the quality and quantity of the training set and its temporal continuity, which may affect the accuracy of the classification and bias the analysis of the LULC changes. In this study, we intended to apply a long-term LULC analysis in a rural region based on a Landsat time series of 21 years (1995 to 2015). Here, we investigated the use of open LULC source data to provide training samples and the application of the K-means clustering technique to refine the broad range of spectral signatures for each LULC class. Experiments were conducted on a predominantly rural region characterized by a mixed agro-silvo-pastoral environment. The open source data of the official Portuguese LULC map (Carta de Uso e Ocupação do Solo, COS) from 1995, 2007, 2010, and 2015 were integrated to generate the training samples for the entire period of analysis. The time series was computed from Landsat data based on the normalized difference vegetation index and normalized difference water index, using 221 Landsat images. The Time-Weighted Dynamic Time Warping (TWDTW) classifier was used, since it accounts for LULC-type seasonality and has already achieved promising overall accuracy values for classifications based on time series. The results revealed that the proposed method was efficient in classifying a long-term satellite time-series with an overall accuracy of 76%, providing insights into the main LULC changes that occurred over 21 years.

Keywords: LULC change; classification; Landsat; TWDTW; cropland mapping; remote sensing

3.1. Introduction

The increased availability of free and analysis-ready remote sensing (RS) data, such as Landsat, supports dynamic analysis in time and space (Hansen & Loveland, 2012; Wulder et al., 2012), since it allows researchers to perform multidimensional classifications of land use/land cover (LULC) based on satellite image time series (Lunetta et al., 2006; Schmidt et al., 2015; Xiao et al., 2005). In fact, the classification methods involving multitemporal remotely sensed images are becoming a trend in LULC change detection (Zhu & Woodcock, 2014) and also in detailed LULC classification approaches (Yan & Roy, 2015), since these are reported to achieve better results than single-date classification methods (Carrão et al., 2008; Gómez et al., 2016; Khatami et al., 2016). Particularly due to the developed and improved tools and algorithms that give priority to time dimension in satellite data analysis (Rufin et al., 2015; Seto & Fragkias, 2005), the changes and disturbances of LULC classes can now be more easily detected.

As Zhu & Woodcock (2014) emphasized, the ability to accurately identify LULC changes using remote sensing data depends on an algorithm that can use high spatial resolution data and is based on a multitemporal analysis. However, the multitemporal analysis application can have some limitations—for example, vegetation phenology, sun angles, clouds, sensor errors, among others (Zhu & Woodcock, 2014). Likewise, concerning the well-established, advanced, nonparametric classifiers in remote sensing literature (Gómez et al., 2016; Phiri & Morgenroth, 2017), such as the support vector machine (SVM) and the random forest (RF) (Raczko & Zagajewski, 2017), it has been recognized that some challenges still remain regarding LULC classifications based on time series (Petitjean et al., 2012). These challenges are: (i) the inexistence/existence and the quality of the samples required to train the algorithm, (ii) the irregular phenological signatures of different LULC types through time, and (iii) the absence of a data temporal continuum (Petitjean et al., 2012).

The Dynamic Time Warping (DTW) classifier has been demonstrated to be a capable solution to deal with some of these challenges (Petitjean et al., 2012; Petitjean & Weber, 2014). Since it can “fill-in” temporal gaps in the remote sensing time series (e.g., cloudy images), it has been successfully applied in satellite imagery time-series analysis (Guan et al., 2016; Maus, Camara, et al., 2016). However, the distinctive phenological cycle of each LULC class requires an equilibrium between shape matching and temporal alignment (Reed et al., 1994; Zhang et al., 2003), which is why Maus et al. (Maus, Câmara, et al., 2016) improved the DTW algorithm. Maus et al. (Maus, Câmara, et al., 2016) proposed the Time-Weighted Dynamic Time Warping (TWDTW) method that includes time-weighting to account for seasonality. This yields better time series to benefit the LULC mapping. In the research of Maus et al. (Maus, Camara, et al., 2016), this method was applied to the LULC classification of a tropical forest area from MODIS enhanced vegetation index (EVI) data. It had the highest accuracy for the forest, pastures, single cropping and double cropping classes' classification and an overall accuracy of about 87%. Belgiu and Csillik (Belgiu & Csillik, 2018) evaluated the TWDTW method for mapping different cropland types in two European cities (in Italy and Romania) and one American city (in California) from Sentinel-2 normalized difference vegetation index (NDVI) time series. In their research, this method achieved a higher overall accuracy when classifying different cropland classes in the two European study areas compared to the classification results of the RF method (Belgiu & Csillik, 2018). Also, the TWDTW method proved it could produce good output in terms of mapping accuracy, even with few training samples (Csillik & Belgiu, 2017; Petitjean et al., 2012).

In addition, it is not only the absence of LULC reference data to train the algorithm that can be detrimental to the image classification results but also the lack of representative samples for each LULC class (Lu & Weng, 2007). If we look into the literature regarding the acquisition of training samples, they are usually acquired from field surveys or expert knowledge through the visual interpretation of other products—for example, high-resolution images from Google Earth and aerial photographs (Lu & Weng, 2007). However, collecting training samples from fieldwork involves high associated costs, in terms of money and time, and collecting training samples through visual interpretations can be difficult, due to the lack of available products, causing biases in the representation of the samples (Usman, 2013). Various approaches have been used to overcome the lack insufficient training samples, such as semi-supervised learning or, more recently, active learning (Huang et al., 2015; Lu et al., 2017). However, both these approaches are dependent on preexisting labelled samples, which require user expertise and proprietary software. In addition, in a long-term multitemporal analysis approach, it would be very unlikely to have samples for each of the years under study (especially, if the analysis had just begun in the current year). Accordingly, it is essential to explore ways to obtain numerous quality training samples to allow more accurate remote sensing classification.

3.1.1. Context and Background

High economic pressures have been strongly influencing agricultural practices over the years (EEA, 2000). Considering the importance of agricultural land, it is urgent to have information regarding the dynamics of LULC changes at regional levels over time to promote parsimonious use of the available resources (Baessler & Klotz, 2006; FAO, 2010; Noszczyk et al., 2019). However, the ability to identify LULC changes accurately depends on a multitemporal analysis (Gómez et al., 2016; Zhu & Woodcock, 2014). Nevertheless, how can we evaluate the LULC dynamics over time when there is scarce reference LULC data for certain regions?

In addition, the agricultural areas can be characterized by different cropping calendars (due to different agricultural practices, intensive and super-intensive) and field geometries (with the existence of both fragmented parcels and larger and compact parcels) and can be influenced by the climate conditions and LULC management (Maus, Camara, et al., 2016; Reed et al., 1994; Zhang et al., 2003). These characteristics, together with the influences of the complex spectral properties of the different vegetation types (Allen et al., 2018), point to some of the challenges present in a multitemporal remote sensing application.

In this study, we address some of these challenges. We start by establishing a methodology to develop a multitemporal analysis of the LULC changes and thereby provide essential information regarding the extent of the changes. The goal of this study was twofold: (i) access free LULC data to be used as training samples for a long-term satellite image time-series classification; and (ii) identify the LULC changes that occurred from 1995 to 2015 in a rural region characterized by a mixed agro-silvo-pastoral environment in the municipality of Beja, Portugal. This region is characterized by a set of LULC complex patterns typical of the Iberian Peninsula. It is a Mediterranean agro-forestry system (Correia, 1993; Russo, 1996) that includes a vast landscape of intermingling cultures, such as wheat, olive groves, vineyards, cork oak forests and pastures, which have a high economic importance in the Portuguese agricultural industry (Godinho, Guiomar, et al., 2016). Therefore, we first sought to integrate the free official Portuguese LULC maps (Carta de Uso e Ocupação do Solo, COS) to produce a data source for training purposes. Second, we sought to investigate the potential of a k-means clustering technique (Cuesta-Albertos et al., 1997; Usman, 2013; Viana, Girão, et al., 2019) to refine the broad range of spectral signatures for each LULC class of the training data. Third, we tested the TWDTW algorithm on a Landsat imagery time-series classification, using the refined sample source. Finally, we evaluated the extent of the changes over 21 years.

Our research is organized as follows. Chapter 3.2 describes the study area and our data. The applied methodology is detailed in Chapter 3.3. The multitemporal satellite imagery classification and the results of the LULC changes are described in Chapter 3.4. Chapter 3.5 elaborates on the discussion. Finally, the main conclusions of our research are described in Chapter 3.6.

3.2. Study area and data

3.2.1. Brief Introduction to the Study Area

The municipality of Beja, in the Alentejo region of Portugal, with an area of about 1,145 km^2, was selected for the study (see Figure 3.1). Different crops characterize this region, but the dominant land use is a mixed agro-silvo-pastoral environment (Correia, 1993; Russo, 1996). The different agricultural practices require different cropping calendars and field geometries, making the landscape of this region distinctive, with the existence of both fragmented land parcels and larger and more compact ones. This region has a Mediterranean climate, influenced by its distance to the coast (average high temperatures are around 33 °C in August and 14 °C in January) (IPMA, 2019). The mean annual precipitation is around 558 mm, concentrated in the months of November to January. Beja is characterized by a low population density (31 inhabitants per km^2). The Guadiana River is the most important watercourse in this region; it crosses the west border of the municipality. The Alqueva Dam was built in 2002 on the Guadiana River, creating the largest artificial lake in Europe. However, it was only fully operational in 2012 in providing water for agriculture services (Allen et al., 2018).

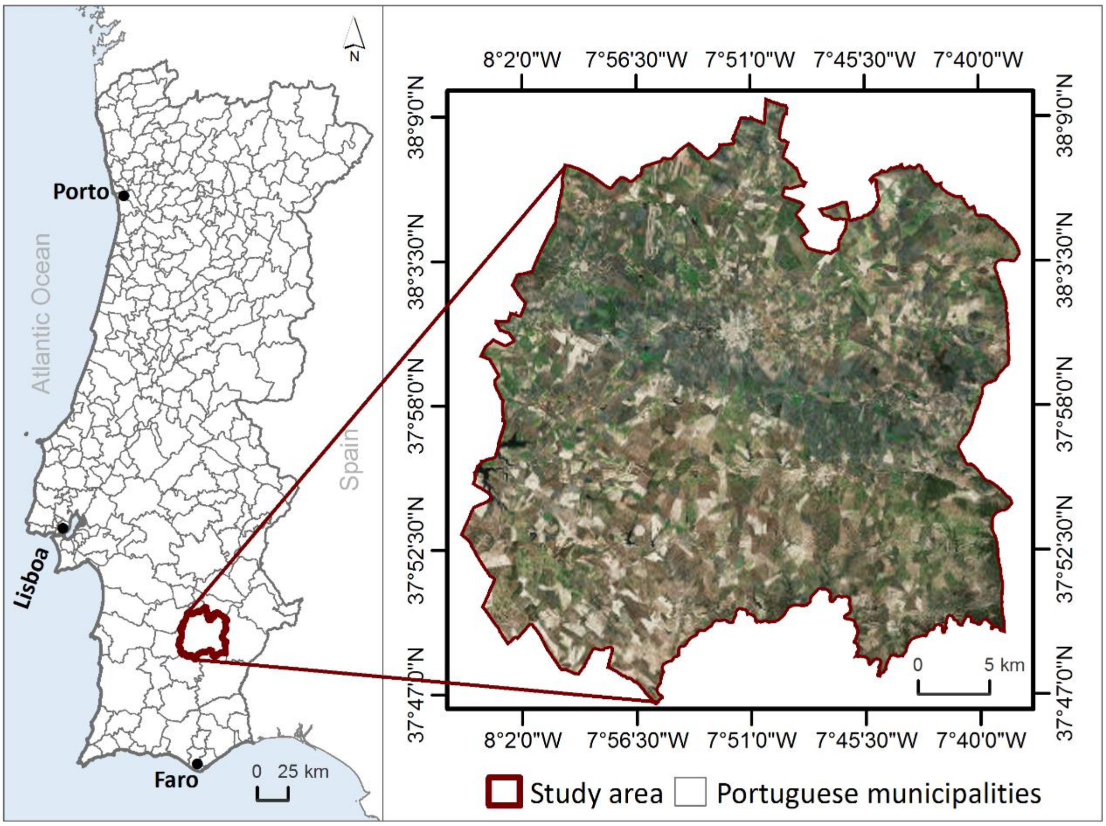


Figure 3.1. Study area (Basemap: true colour orthophotomap provided by ESRI/Digital Globe)

3.2.2. Satellite Image Collection and Preprocessing

Landsat data is the longest and most informative temporal report of the earth. It is the only data source that gives a continuous image of the earth's resources with a high spatial resolution (30 m) spanning nearly 40 years (Woodcock et al., 2008). To obtain vast temporal continuity data and compose our long time series, we downloaded available Landsat data (with less than 50% of cloud cover) from 1995 through 2015 from the official website of the United States Geological Survey (https://earthexplorer.usgs.gov/). This included Landsat 4–5 Level-2 and Landsat 8 Level-2 data (path 203, row 34). We were working with data that spanned 21 years, for a total of 221 scenes (see Table 3.1). As shown in Table 1, there were no available Landsat images for 2012, either from Landsat 4–5 Level-2 or Landsat 8 Level-2. Only the Landsat 7 sensor had information for 2012, but unfortunately, as is commonly known, all Landsat 7 scenes have data gaps for this area (due to a technical problem with the satellite) (Hossain et al., 2015).

As we were using Landsat Level-2 Surface Reflectance images, no atmospheric correction was needed. Also, no further normalization was required since the TWDTW uses a Generalized Additive Model (GAM) with a spline smooth function and a logistic regression for time weight (see Chapter 3.4). Furthermore, no masking for clouds or shadows was performed since the TWDTW classifier can “fill in” temporal gaps (Guan et al., 2016; Maus, Camara, et al., 2016). The ETRS_1989_Portugal_TM06 coordinate system was applied to the image datasets, respecting the 30 x 30 m spatial resolution.

Table 3.1. Frequency by year of all available Landsat Level 2 data between 1995 and 2015 years

Year	Number of images	Year	Number of images
1995	14	2006	8
1996	8	2007	13
1997	13	2008	8
1998	15	2009	16
1999	12	2010	12
2000	10	2011	11
2001	12	2012	0*
2002	5	2013	8
2003	6	2014	14
2004	11	2015	17
2005	8		
*No available images			

3.2.3. Official Portuguese Land Cover Map (COS) Data

The COS maps of 1995, 2007, 2010, and 2015 used to derive the training samples were produced by the Portuguese General Directorate for Territorial Development (DGT). These datasets are freely available for download on the DGT website (http://mapas.dgterritorio.pt/geoportal/catalogo.html). The COS maps are in a vector format (polygons) with a cartographic accuracy of 0.5 m and a minimum mapping unit (MMU) of 1 hectare and are produced by photointerpretation (i.e., over orthophoto maps) (DGT, 2018). COS uses a hierarchical, a priori nomenclature system, having five disaggregation levels. A DGT report (DGT, 2018) describes in detail the COS nomenclature characteristics.

3.3. Methods

The first step of this research was to download the Landsat Level-2 imagery data from 1995 to 2015. Second, we used the data to compute the normalized difference vegetation index (NDVI) and the normalized difference water index (NDWI). Then, we combined all the scenes to create one raster time-series file for each index. Third, using the COS maps, we identified the most representative LULC classes in the study area and randomly generated sample points for each year to be used in the training. Fourth, we refined the training samples by class and by year using the k-means clustering technique. Finally, in the fifth step, we classified the Landsat imagery time series using the TWDTW algorithm and evaluated the classification accuracy (see Figure 3.2).

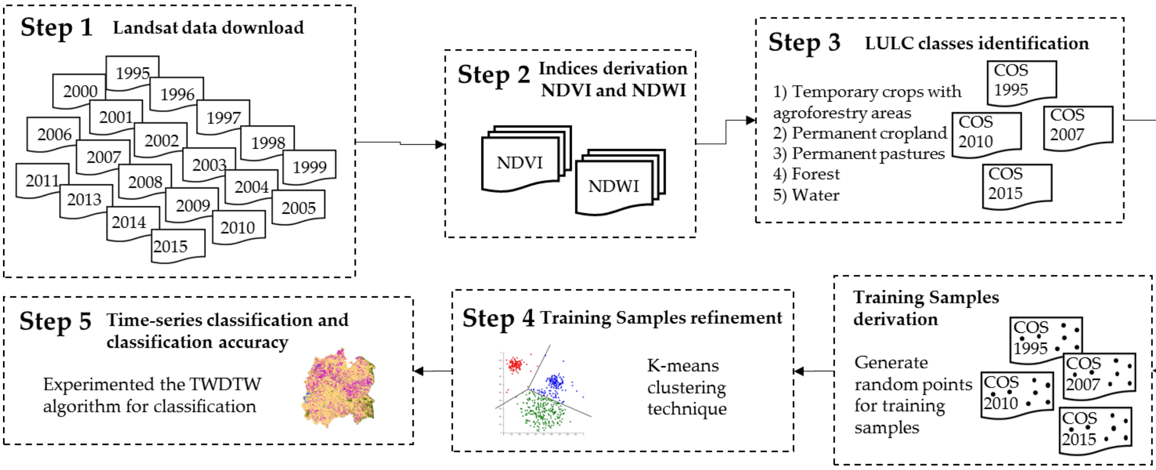


Figure 3.2. Workflow for the long-term time series LULC methodology

3.3.1. Derivation of NDVI and NDWI from Landsat Imagery

The municipality of Beja has a heterogeneous biophysical environment, so to improve the identification of the different vegetation classes using Landsat imagery seasonal time series (Carrão et al., 2008; Lobo et al., 2004; Vicente-Serrano & Heredia-Laclaustra, 2004), we computed two indexed time series, NDVI and NDWI. To derive these indices, we used the equations in Table 3.2.

The normalized difference vegetation index (NDVI) is the most commonly used satellite-based measure to globally monitor vegetation (Rouse et al., 1973) due to its robust and easy interpretation. This index considers the distinctive shape of the vegetation reflectance curve and is able to screen dynamic specifications of vegetation and/or vegetation health (Azzali & Menenti, 2000). Despite the existence of several other vegetation indices, the ability of the NDVI index to monitor vegetation for seasonal and inter-annual changes has already been sufficiently substantiated in prior studies (Rouse et al., 1973; Yuan & Bauer, 2007). The normalized difference water index (NDWI) was used as an additional aid, since it was conceived to outline open water features and improve their identification through RS imagery (Gao, 1996; McFeeters, 1996). According to Gao (Gao, 1996), this index complements NDVI by improving in the identification of liquid water content of vegetation, whereas NDVI just accounts for the vegetation's greenness (Ceccato et al., 2002).

Table 3.2. Normalized difference vegetation index (NDVI) and normalized difference water index (NDWI) equations

Vegetation index name	Landsat 4-5 images	Landsat 8 images	Reference
NDVI	$\frac{Band\ 4\ (\text{NIR}) - Band\ 3(\text{Red})}{Band\ 4\ (\text{NIR}) + \text{Band } 3(\text{Red})}$	$\frac{Band\ 5\ (\text{NIR}) - Band\ 4\ (\text{Red})}{Band\ 5\ (\text{NIR}) + Band\ 4\ (\text{Red})}$	(Rouse et al., 1973)
NDWI	$\frac{Band\ 2\ (Green) - Band\ 4(NIR)}{Band\ 2\ (Green) + Band\ 4(NIR)}$	$\frac{Band\ 3\ (Green) - Band\ 5\ (NIR)}{Band\ 3\ (Green) + Band\ 5\ (NIR)}$	(Gao, 1996)

3.3.2. Training Sample Derivation

The COS maps for 1995, 2007, 2010, and 2015 were used to produce a sample source for training purposes. As the first step, we statistically analyzed the COS maps to understand what were the main LULC classes in the study area and how they changed over time. The identification of the major LULC classes was based on the second level of nomenclature of the COS maps. We found the predominant temporary crops (51%) in 1995 (Table 3.3). However, in later years (2007, 2010 and 2015), the coverage of temporary crops declined by almost 8%, though they still represented the largest cropland use (43%). The heterogeneous agricultural areas represented the second largest LULC class over the 21-year period (19% in 1995 and 2015; 20% in 2007 and 2010). Indeed, during this time interval, there was an increase in forested areas (11% in 1995; 12% in 2007; and 13% in 2010 and 2015). Despite the decrease in pasture coverage between 2010 and 2015 by 2%, this LULC class covered about 10% of the study area in 1995 and 12% in 2007. Although permanent crops represent a small amount of LULC in the Beja municipality, they increased substantially (by 7%) between 1995 and 2015. Artificial surfaces and wetlands represent minor LULC classes in this municipality.

The results revealed the predominance of agricultural land (>82%) in the Beja municipality over the 21-year period. Nevertheless, the second largest class (heterogeneous agricultural areas) encompassed different mosaics: i) areas of temporary crops and/or permanent crops on the same parcel; ii) temporary crops cultivated under agroforestry systems; iii) meadows and/or permanent crops, and iv) landscapes in which crops and pastures intermix with natural vegetation [41]. Indeed, this was a highly heterogeneous class that may include each of the other agricultural classes that predominate in the Beja municipality. When analyzing this class in more detail (looking at the COS maps' third level of nomenclature), we found a predominance of temporary crops cultivated under agroforestry systems (>85%) over the 21-year period.

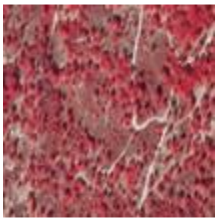
Therefore, we decided to mix the temporary crops and the heterogeneous agricultural areas classes. As such, five major LULC classes were identified as the most representative of the selected study area: 1) temporary crops with agroforestry areas (including the temporary crops and the heterogeneous agricultural areas); 2) permanent cropland; 3) permanent pastures; 4) forest; and 5) water. Also, the temporary crops with agroforestry areas could encompass impervious surfaces, since in Beja, the dispersed settlements had relatively small dimensions with houses far apart from each other and surrounded by bare soils and/or croplands. The impervious surfaces were less than 2% of the whole Beja municipality area and have not significantly changed throughout the years.

Table 3.3. Descriptive statistics of the COS maps

<table>
<tr><th>Nomenclature Level 1</th><th>Nomenclature Level 2</th><th>Class Coverage (1995)</th><th>Class Coverage (2007)</th><th>Class Coverage (2010)</th><th>Class Coverage (2015)</th></tr>
<tr><td rowspan="4">1. Artificial surfaces</td><td>1.1. Urban fabric</td><td rowspan="4">1.4%</td><td rowspan="4">1.5%</td><td rowspan="4">1.6%</td><td rowspan="4">1.6%</td></tr>
<tr><td>1.2. Industrial, commercial and transport units</td></tr>
<tr><td>1.3. Mine, dump and construction sites</td></tr>
<tr><td>1.4 Artificial, non-agricultural vegetated areas</td></tr>
<tr><td rowspan="4">2. Agricultural areas</td><td>2.1 Temporary crops</td><td>51.1%</td><td>42.8%</td><td>42.6%</td><td>42.5%</td></tr>
<tr><td>2.2 Permanent crops</td><td>5.7%</td><td>9.1%</td><td>11.9%</td><td>13.1%</td></tr>
<tr><td>2.3 Pastures</td><td>9.7%</td><td>12.3%</td><td>8.1%</td><td>8.1%</td></tr>
<tr><td>2.4 Heterogeneous agricultural areas</td><td>19.1%</td><td>19.8%</td><td>20.3%</td><td>19.2%</td></tr>
<tr><td rowspan="3">3. Forest and semi natural areas</td><td>3.1 Forests</td><td rowspan="3">11.3%</td><td rowspan="3">12.4%</td><td rowspan="3">13.1%</td><td rowspan="3">13.0%</td></tr>
<tr><td>3.2 Scrub and/or herbaceous vegetation associations</td></tr>
<tr><td>3.3 Open spaces with little or no vegetation</td></tr>
<tr><td rowspan="2">4. Wetlands</td><td>4.1 Inland wetlands</td><td rowspan="2">0%</td><td rowspan="2">0%</td><td rowspan="2">0%</td><td rowspan="2">0%</td></tr>
<tr><td>4.2 Maritime wetlands</td></tr>
<tr><td rowspan="2">5. Water bodies</td><td>5.1 Inland waters</td><td rowspan="2">1.7%</td><td rowspan="2">2.1%</td><td rowspan="2">2.4%</td><td rowspan="2">2.5%</td></tr>
<tr><td>5.2 Marine waters</td></tr>
<tr><td colspan="2">Total</td><td>100%</td><td>100%</td><td>100%</td><td>100%</td></tr>
</table>

As the second step, we randomly generated the training sample points for each of the five LULC classes previously identified (see the COS nomenclature codes in Table 3.4). However, we decided to a priori reduce the polygon areas of the COS maps, using a QGIS Python plugin called “Buffer by Percentage”. We followed this approach, because we knew training samples had low accuracies when their proximity to the polygon’s boundaries increased (Viana, Encalada, Rocha, et al., 2019).

Table 3.4. Land use/land cover classes' characteristics

LULC class	COS nomenclature code	Description	Image example	Field example
Temporary crops with agroforestry areas	2.1., 2.4.4	Temporary crops (non-irrigated and permanently irrigated crops); complex cultivation patterns (herbaceous understory); land principally occupied by agriculture, with significant areas of natural vegetation; agroforestry areas.		
Permanent cropland	2.2.	Crops occupied the land for a long period and had a nonrotating regime (olive groves, vineyards)		
Permanent pastures	2.3.	Areas permanently occupied (≥ 5 years), with mainly herbaceous vegetation		
Forest	3.1., 3.2., 3.3	Areas occupied by tree clusters resulting from natural regeneration or planting		
Water	5.1., 5.2.	Freshwater surfaces, including watercourses and water plans (both natural and artificial)		

As we were using the shape and area of each COS map polygon to produce the training samples, we sought to ensure a relationship of magnitude between the MMUs of both data sources (the COS map and the Landsat imagery) (García-Álvarez, 2018). This ensured the generated samples related to one cell superimposed to just one polygon. Since the COS has an MMU of 1 hectare and the Landsat imagery cell size was 900 m^2, we reduced the COS polygon areas to 11%

of the original. Consequently, 2000 training samples were generated for the heterogeneous agriculture class (since this area occupied more than 50% of the study area) and 1000 training samples were generated for each of the remaining LULC classes of the four COS maps (1995, 2007, 2010, and 2015) making 6000 samples per year.

3.3.3. Training Sample Refinement

The created sample source had the following limitations: (1) the sample source was acquired from vector-based LULC data (the COS maps); (2) no training samples acquired from fieldwork were integrated; and (3) there was a high probability of spectral-temporal signature confusion among LULC classes, due to the heterogeneous biophysical environment of the study area. As such, we decided to apply the k-means clustering technique (Cuesta-Albertos et al., 1997; Usman, 2013; Viana, Girão, et al., 2019) to improve the sample source.

Cluster analysis methods are normally used in an attempt to detect homogeneous groups. The rationale behind implementing a clustering method to refine the training set is that each generated training sample per class should be grouped to either one or, at most, two clusters. This is due to the fact that each generated training sample should have a similar spectral signature, such as the one belonging to their LULC class. To verify this hypothesis, we used the R package tclust (Fritz et al., 2012).

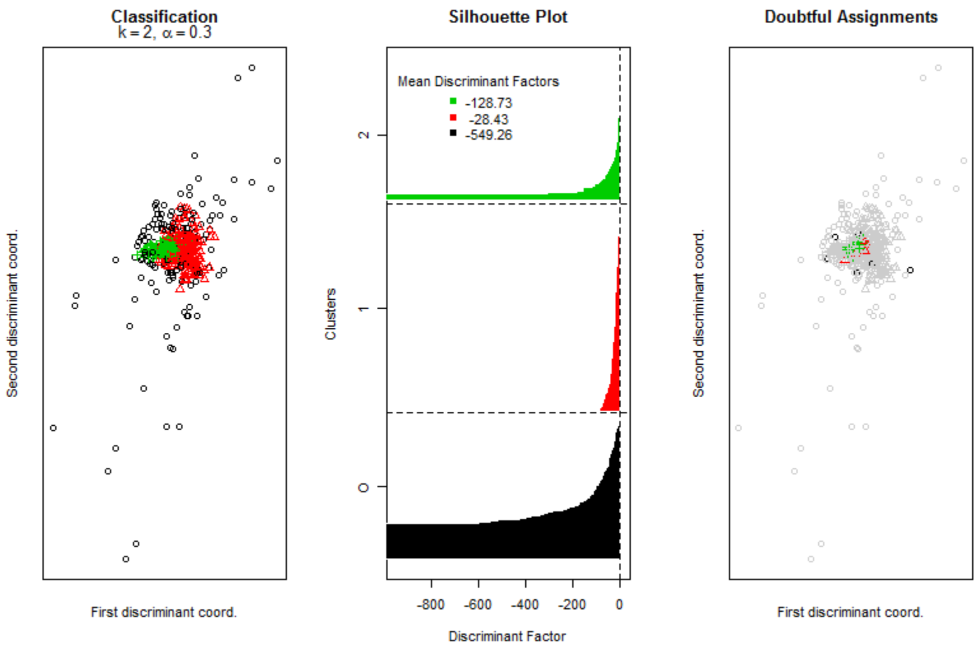


Figure 3.3. Example of graphical display based on the discriminant factors (Df) values from the tclust cluster solution obtained with k = 2, α = 0.25 for one of the LULC classes. Larger Df values indicate not very "well-determined" clusters

The R package tclust offers an adequate cluster scatter matrix constraint, leading to a wide range of clustering procedures and avoiding the occurrence of spurious non-interesting clusters. Also, it deals with noisy data and provides graphical exploratory tools (see Figure 3.3) to help the user make sensible choices regarding trimming the outliers and choosing the number of clusters (Fritz et al., 2012). These options were the main reason why this package was chosen.

The clustering process is schematically portrayed in Figure 3.4 and is essentially divided into three distinct processes: 1) classification-trimmed likelihood calculation (ctlcurves); 2) cluster computation; and 3) mean discriminant factor value calculation (DiscrFact). The variables used for the creation of clusters were the NDVI and the NDWI indices (the same variables that would be used for the image classification).

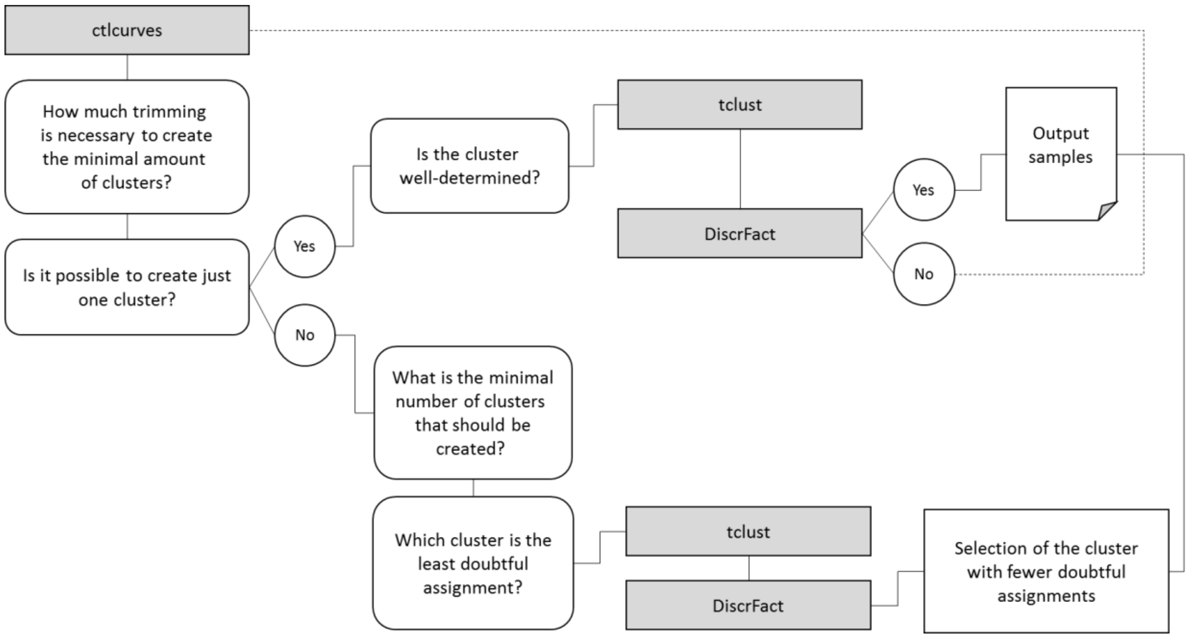


Figure 3.4. Cluster process workflow

The ctlcurves function approximates the classification-trimmed likelihoods by successively applying the tclust function to a sequence of values, k (number of clusters) and α (fraction of the most outlying data). The calculation of ctlcurves shows the implications of increasing/decreasing k-values in the fraction of the most outlying samples (limited, in this case, to 30% of the sample universe).

To determine the most adequate parameters to include in the tclust function, the following logic was employed:

1. Was it possible to create only one cluster that could well represent each LULC class reality? If so, then only one cluster was created.

2. When more than one cluster had to be created, what was the minimal number of clusters that should be created? In this case, we selected the minimum number of clusters possible,

trimming up to the maximum of 30% of the sample universe. When this strategy was put into practice, two and, in rare cases, three clusters were created.

To make sure that in both hypotheses we were left with well-determined clusters, the quality of the cluster assignments and the trimming was evaluated by applying the function DiscrFact. This function helped: 1) to discover if the clusters were well-determined, and 2) to select the cluster with the least doubtful assignments to represent the reality of the respective class (Rousseeuw, 1987).

The parameters k and α for the tclust function that provided (among other information) the cluster assignment to each observation, differed according to the results generated by the ctlcurves and DiscrFact functions. The remaining parameters were left at their default values, with the exception of restr that was set to "eigen" to simultaneously control the relative group sizes and the deviation from sphericity in each cluster. Also, the equal.weights parameter was set to "TRUE" to avoid the creation of a well-determined cluster that actually does not represent the LULC class. This way, we achieved heterogeneous clusters as much as possible, since no LULC class had a pure and unique spectral signature. For a correct classification, we needed some variability in the range of values to avoid eliminating potentially important information (Fritz et al., 2012). The restr.fact parameter was increased until no artificial restriction was necessary to create the clusters (Rousseeuw, 1987) (1 being the most restrictive case).

Finally, the tclust package was used to refine the training points generated for each class and for each of the four years. In Table 3.5, we present the final number of training samples per class for each year after the clustering analysis.

Table 3.5. Final number of samples per class

		LULC class				
		Temporary crops with agroforestry areas	Permanent pastures	Permanent cropland	Forest	Water
Number of training samples	1995	1158	474	684	399	648
	2007	1062	384	474	486	456
	2010	786	600	630	684	402
	2015	828	540	636	787	588

3.3.4. Dynamic Time Warping and Time-Weighted Dynamic Time Warping Methods

Sakoe and Chiba (1978) developed the Dynamic Time Warping (DTW) method for speech recognition applications, and it was later introduced for use in satellite imagery time series analysis (Guan et al., 2016; Maus, Camara, et al., 2016; Petitjean et al., 2012; Petitjean & Weber, 2014). However, DTW proved to be unfit for satellite imagery time-series analysis, because "it disregards the temporal range when trying to find the best pair matches between time series" (Maus, Câmara, et al., 2016). The distinctive cycle of each LULC class requires an equilibrium between shape matching and temporal alignment to be employed in LULC classification using remote sensing time-series images (Reed et al., 1994; Zhang et al., 2003).

To build the time-series analysis, the TWDTW algorithm maps the derived indices (NDVI and NDWI) to a three-dimensional array in space and time, where each pixel location is associated with a sequence of measurements. To match LULC classes to the subintervals (i.e., years) of a long-term satellite image time series and classify phenological cycles, this method uses both cycle amplitude and phase information. The spectral-temporal patterns are defined by applying a GAM (Maus, Camara, et al., 2016; Maus, Câmara, et al., 2016) with a smoothing function of y = s(x), where the function s corresponds to a spline model, x is time, and y is each satellite band.

To set the temporal constraints of the TWDTW algorithm, we applied a logistic weight function, since this provides more accurate results than a linear function (Maus, Camara, et al., 2016; Maus, Câmara, et al., 2016). This logistic function uses a small weight for short time warps and substantial weight for larger time warps, with a midpoint (beta) and steepness (alpha). It is formally given by (Eq. 3.1) (Maus, Camara, et al., 2016; Maus, Câmara, et al., 2016):

$$\mathrm{w} = \frac{1}{1 + \mathrm{e}^{-\alpha(\mathrm{g}(t_1, t_2) - \beta)\cdot}} \tag{3.1}$$

where $\mathrm{g}(t_1, t_2)$ is the elapsed time (in days) between date t_1 (in the pattern) and date t_2 (in the time series).

The purpose of this function is to control the time warp. If a large seasonal difference exists between the pattern and its matching point in the time series, an additional cost is added to the distance measure (DTW). The g function constraint drives time warping and ensures that the time-series alignment is reliant on the seasons. This is particularly valuable for identifying temporary crops and in differentiating pasture from agriculture. To achieve an LULC classification, and as recommended in Maus et al. (Maus, Câmara, et al., 2016), we assign the logistic function $\alpha = -0.1$ and $\beta = 100$ to weight the function. This implies adding a time weight to the DTW with a low penalty for time warps smaller than 60 days and a higher penalty for larger time warps. We also set the temporal pattern frequency to eight days to ensure, at least, one Landsat image was considered (i.e., a temporal resolution of eight days).

Since LULC classifications are extremely dependent on the quality of spectro-temporal patterns, we performed a k-fold cross-validation (Ramezan et al., 2019). This function uses the training samples to generate the spectral-temporal patterns. The results of this classification are used in the accuracy calculation. In this study, we obtained the accuracy for each data partition using 100 different data partitions, each of them with 10% of the training samples randomly selected for training and 90% for validation. All calculations and analyses were accomplished using the R programming language and operating environment (Team, 2017). Maus et al. (Maus, Câmara, et al., 2016) implemented the TWDTW method for LULC classifications using satellite imagery time series on an open-source R package (dtwSat, for a package overview see (Maus, Câmara, et al., 2016)).

3.3.5. Classification Accuracy Assessment

Several metrics commonly used in remote sensing were calculated, including the kappa index (Landis & Koch, 1977), the overall and individual user and producer accuracies (Arsanjani et al., 2015), the error matrix (confusion matrix), and the confidence intervals for the accuracy measures (error tolerances) (Baraldi et al., 2006). The kappa index estimates how good the classification is when compared to a randomly generated image. The overall accuracy measure indicates total classification performance as a percentage by dividing the number of samples correctly classified by the total number of samples. Individual user accuracy measures the probability that any classified pixel will actually match the samples, while individual producer accuracy provides the probability that a particular LULC class is classified as such in the samples. The error matrix indicates the samples correctly (diagonal) and incorrectly classified for each LULC class.

3.4. Results

3.4.1. LULC Classification and Analysis

The distance measure is employed to produce categorical land cover maps, where the classification result is based on the most similar pattern for each period. Classification results for each period can be visualized in Figure 3.5. The results reveal the predominance of temporary crops with agroforestry areas in the Beja municipality. Beja is, in fact, characterized by extensive agriculture with complex cultivation patterns, where significant areas are occupied by temporary crops with natural vegetation and agroforestry areas. The substantial coverage of temporary crops with agroforestry areas confirm the importance of cereal production (particularly wheat) in this municipality, as well as the relevance of cork oak exploitation in the vast landscape of agroforestry areas (named Montado in Portugal and Dehesa in Spain) (Correia, 1993; Godinho, Gil, et al., 2016).

Over the 21-year period, the biggest change noticed was the substantial increase of the permanent cropland through scattered parcels, especially after the 2010–2011 period. Although this class represents a small coverage in Beja, this considerable increase suggests a change focused on intensive farming without fallow.

We found that permanent pastures represented a minor LULC class and were mainly concentrated near forested areas. Indeed, the forested areas remain mainly in the southeastern and northeastern part of the municipality, possibly due to the topographic conditions (plateaus) and the proximity to the Guadiana River. Figure 3.6 helps to visualize in detail the main significant changes throughout the 21 years under analysis.

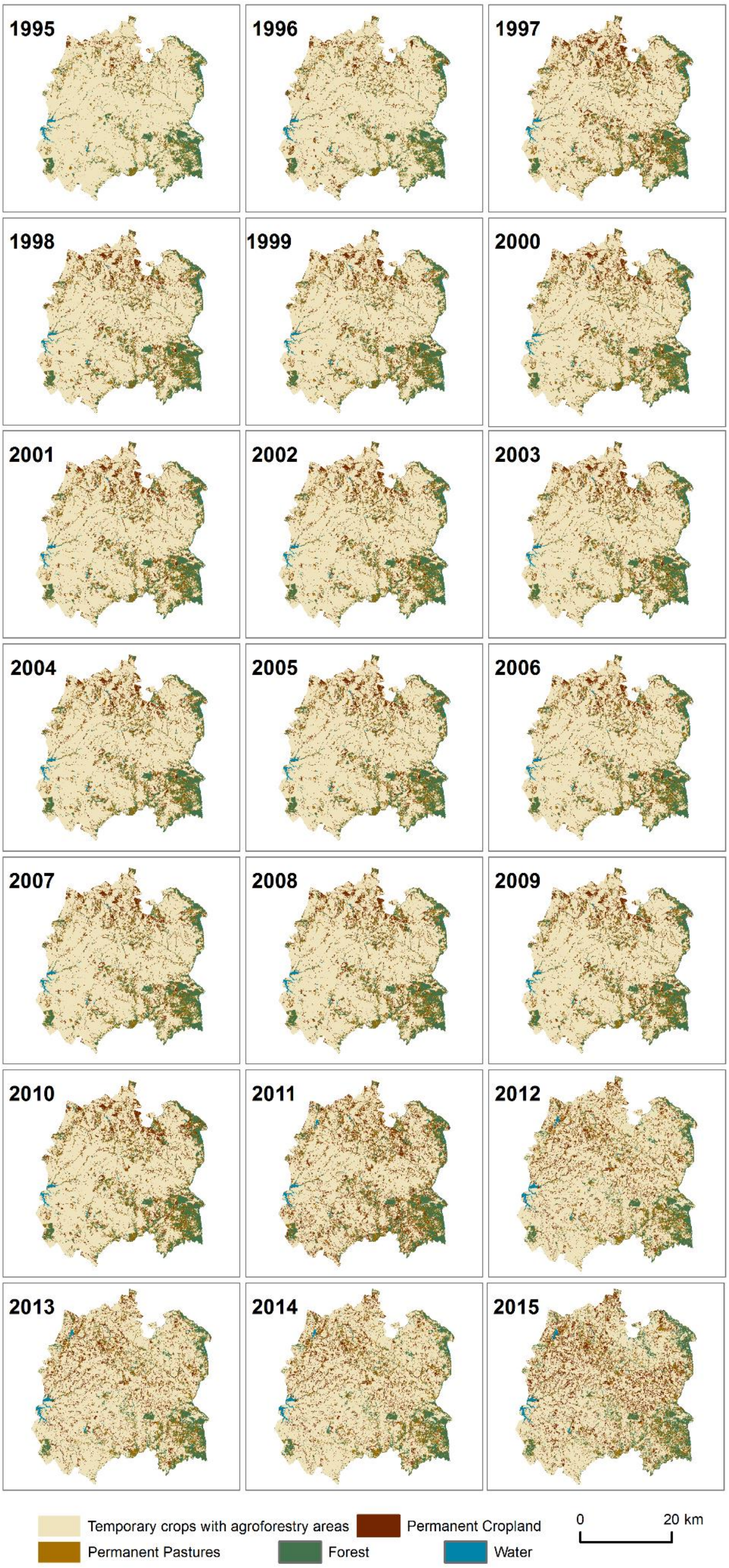


Figure 3.5. LULC classification maps for each period (1995–2015)

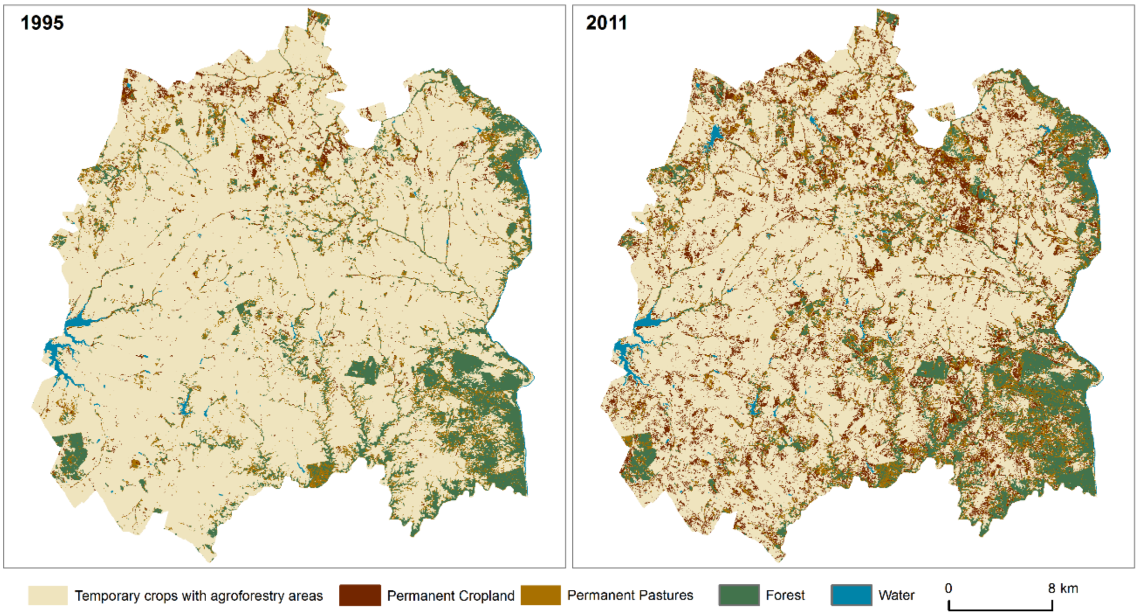


Figure 3.6. LULC classification maps for 1995 and 2011

Different occupation areas for each class were observed, which was in agreement with the LULC class spatial distribution. The increase of the permanent cropland was clear from 2011 forward, occupying an area of about 16% in 2015 in contrast with the 1995 occupation area of 8% (Figure 3.7). The water areas had a similar pattern of consistent increases, up from about 7% of the total area in 1995 to 11% in 2015, with a significant increase in the 2010–2011 period. Regarding the agricultural classes, the temporary crops with agroforestry areas class was clearly prevalent, occupying a total area close to 63% in 1995. However, this class had a significant area decrease from 1995 to 2015 (about 11%). The permanent pastures group did not vary greatly from 1995 to 2015; it always occupied an area of about 12%–13%. Forested areas did vary more over the 21 years analyzed, the area occupied ranged between 11% (maximum area) and 8% (minimum area). In particular, from 2003 to 2010-2011, little or no change in the LULC was detected, perhaps due to the measures brought by the 2003 Common Agriculture Policy (CAP) that more driven by social integration and less to supporting production (Allen et al., 2018).

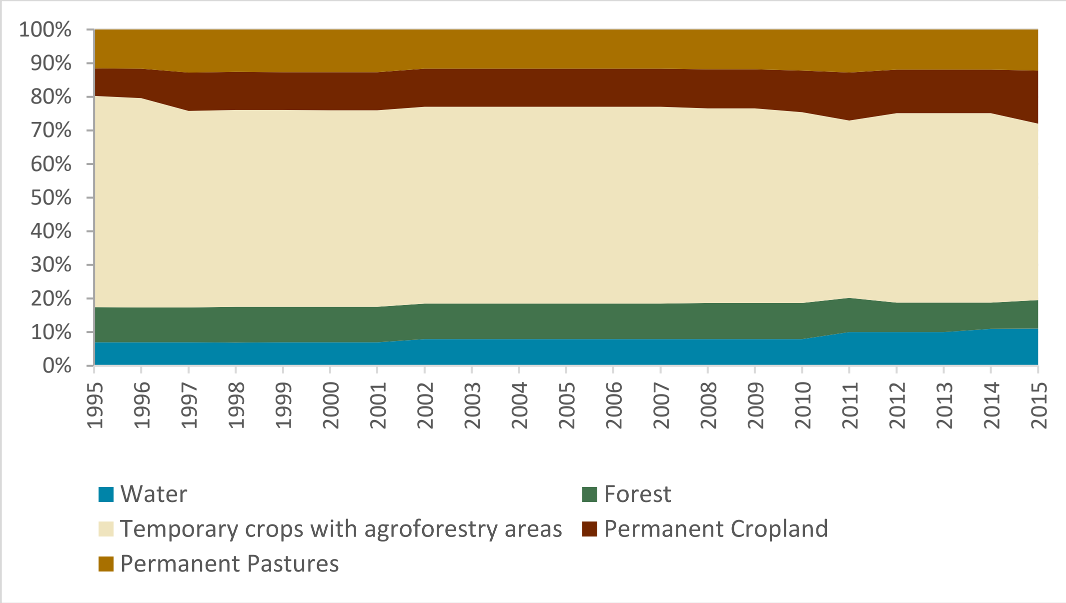


Figure 3.7. LULC classes (%) between 1995 and 2015

3.4.2. LULC Classification Accuracy Summary

A set of metrics—including overall user and producer accuracy, the kappa index, and an error matrix (confusion matrix) were calculated. The results of the assessment were gathered for the full period (i.e., the total LULC mapping area was the same as the surface area times the number of maps), but it is also possible to evaluate and visualize each period separately. The accuracy assessment was computed using the validation samples (90% of the total samples, with a 95% confidence level), obtaining an overall classification accuracy of 75.96%.

Table 3.6 presents the error matrix in proportion to the classification map. Only the water class had almost all the samples correctly classified. It is noteworthy that the misclassified samples were mistaken as forest. In this region, forested areas are characterized by hardwood forest that grows along water lines and open spaces with little or no vegetation. Regarding forested areas, there was some confusion between this class and the agricultural classes. Still, only about 4% of the total samples were misclassified. However, between the agricultural classes, the temporary crops with agroforestry areas samples that were misclassified were mostly mistaken as permanent cropland or permanent pastures (about one-quarter of this class's total sample size). In contrary, the permanent cropland was mostly classified as temporary crops with agroforestry areas (almost 40% of the total sample size was wrongly classified), but it also presented some confusion with the forest class (about 5% of the samples). Similarly, there was some confusion in the permanent pastures class with the temporary crops with agroforestry areas (about 27% of the samples) and also with forest (6% of the samples).

In the class-based analysis, we observed a wide range of accuracy values. The water class was the most accurate, with both use and producer accuracy above 99%, suggesting high confidence and sensitivity of the method for this class. The forest class followed the same pattern as the water class, with high user accuracy (96%) and producer accuracy (88%). However, regarding the agricultural classes, they did not present consistent accuracy values. The temporary crops with agroforestry areas class was the most accurate with a user rate above 74% and a producer rate above 87%, which suggests the pixels classified as temporary crops with agroforestry areas closely matched those in the COS. In contrast, the permanent cropland and permanent pastures classes presented user accuracy values around 56% and 66%, respectively, while the producer accuracy values were even lower (43% for permanent cropland and 39% for permanent pastures). Overall, a kappa coefficient of 0.62 indicated a substantial agreement between the classification and the samples.

Table 3.6. Error matrix of the map classification task (full period)

Map class	Reference class: Temporary crops with agroforestry areas	Permanent cropland	Permanent pastures	Forest	Water	Total	User Accuracy ± Error tolerance
Temporary crops with agroforestry areas	9343	1726	1494	53	0	12616	74.06 ± 0.1%
Permanent cropland	940	1338	4	108	0	2390	55.98 ± 0.3%
Permanent pastures	395	10	958	87	0	1450	66.07 ± 0.4%
Forest	24	19	22	1751	0	1823	96.42 ± 0.1%
Water	0	0	0	7	2055	2055	99.66 ± 0.0%
Total	10702	3093	2478	1999	2062	20334	
Producer Accuracy ± Error tolerance	87.30 ± 0.1%	43.25 ± 0.3%	38.66 ± 0.3%	87.28 ± 0.2%	100.0 ± 0.0%		
Overall Accuracy	75.96 ± 0.1%						
Kappa index	0.62						

3.5. Discussion

Over the years, the remote sensing applications for rural areas have become extremely important (Atzberger, 2013), since they support the monitoring and mapping of the LULC changes of agricultural land. Some studies have emphasized the usefulness of using remote sensing data for agricultural land change classification (Ross et al., 2017; Wardlow et al., 2007), cropland estimation/scenario forecast (Lyle et al., 2013; Meroni et al., 2013; Mulianga et al., 2013), and vegetation health and crop production monitoring (Gumma et al., 2016; Teluguntla et al., 2015). The ongoing abandonment of agricultural land (Fuchs et al., 2015; Radwan et al., 2019; Shi et al., 2016), and the fact that these areas have been threatened by socioeconomic and biophysical factors (Feranec et al., 2017; Serra et al., 2008) demonstrates the need to identify and characterize the spatiotemporal changes. Accordingly, in this study, we applied a long-term

LULC analysis in a rural region based on Landsat time series of 21 years. However, it involved overcoming a number of challenges.

3.5.1. The Implications of a Long-Term Landsat Time Series

For case studies at a regional scale, Landsat TM/ETM+ has been one of the most frequently used products due to its medium/high spatial resolution (30 m), large temporal extent (since the 1970s), and free and analysis-ready availability (Hansen & Loveland, 2012; Wulder et al., 2012). However, in this study there were some limitations that were identified regarding the use of the Landsat data.

Having a long time-series span without interruptions was not possible, since there were no images available for the year 2012. Only the Landsat 7 sensor had information relative to 2012, but unfortunately, as is commonly known, many Landsat 7 scenes have data gaps (the result of a technical problem with the satellite) (Hossain et al., 2015). Although there are several approaches to handling the Landsat 7 sensor error problem, all the processes are time-consuming, and the gaps may not be completely eliminated (Santos et al., 2017). In addition, to have a large temporally continuous dataset and create our long time series, we had to combine a set of images of Landsat 4–5 Level-2 and Landsat 8 Level-2 data. As such, the absence of images for 2012 and the use of two different sensors were limitations of this study.

Clouds are often a problem in processing satellite imagery. In our study, the temporal continuity was not affected, as we could use all scenes regardless of the presence of clouds. However, when using other classifiers, a long-term Landsat time series analysis can be difficult, due to the high number of scenes with the presence of clouds, which can affect the temporal continuity and the integrity of the classification.

3.5.2. The Generation of Training Samples from Official LULC Maps

One challenge associated with a long-term application is the lack of a sufficient and representative training sample. In fact, the accuracy of an image classification result is highly dependent on the quality and quantity of the training set used to train the classification model (Brodley & Friedl, 1999; Lippitt et al., 2008). For most supervised classifiers (such as a maximum likelihood classification, multilayer perceptron, SVM, or RF) not having sufficient and representative training data can be detrimental to the image classification results (Lu & Weng, 2007). Furthermore, in a long-term multitemporal image classification result, it would be very unlikely to have samples for each of the years under study.

Most studies about image classification usually mention the use of field surveys or visual interpretations to obtain training samples (Lu & Weng, 2007). How can a long-term classification be done when both options are not feasible? The approach presented in this study provides a path toward addressing this problem by exploring a way to generate a set of samples for training purposes using free LULC data (the COS maps) from different years. Even if money and time were available to collect training samples from fieldwork, the development of an acceptable automated alternative would allow those resources to be used in other ways.

In this exploratory study, we tested the use of the COS maps to produce a sample source for training purposes for various reasons: (i) the LULC data was produced by a governmental institution (DGT); (ii) it had a medium spatial resolution; (iii) it was the data with most temporal continuity in Portugal; and (iv) it was one of the most used data sources for LULC analysis in Portugal (Meneses et al., 2018). However, generating training samples from a LULC map (such as the COS) produced by a governmental institution that we considered a reliable source of information still had associated problems. For example, following a previous work (Viana, Encalada, & Rocha, 2019) we decided to reduce the COS map polygon areas before the generation of random points since the authors of the cited work concluded that the boundaries of the polygons of such sources can have low accuracy. In addition, bias in the representation of the samples can exist if there is no prior knowledge regarding the major LULC types present in the study area. This makes it very difficult to know which LULC classes from the COS should be used to generate the training samples associated with the most representative LULC classes of the study area. We had to support the identification of the major LULC classes present in our study area by a statistical analysis of the coverage of the COS LULC classes.

In addition, even though we were able to generate 6000 training samples for each year in the study, this set was irregularly distributed over time, since it only represented the information of the LULC for four of the 21 years under analysis. This was, in fact, the major constraint we had, and, once again, it depicted the limitations associated with this type of remote sensing application.

3.5.3. The LULC Types and Temporal Phenological Signatures Diversity

In our study, we decided to apply the k-means clustering technique to refine the training samples, since they were acquired using an explorative approach without integrating any ground-truth data and there was mixed agro-silvo-pastoral ecosystem present in our study area. Moreover, Viana et al. (2019) had already demonstrated the training samples refinement using the k-means clustering technique improves the image classification’s overall accuracy (+8%).

Interestingly, Allen et al. (2018) and Senf et al. (2015) had already mentioned the difficulty of classifying the different LULC types of a region with a heterogeneous biophysical environment due to the wide range of spectral signatures for each LULC class. Our word with the Beja municipal region was no different; the permanent cropland class is an example of a class with a complex spectral pattern. This class was mostly constituted of olive groves, and the production techniques of olive groves have changed over the past decade. This region has always been characterized by open olive grove crops, but recent foreign direct investment in super-intensive olive grove plantations has resulted in a different landscape. It is easy to find intensive and super-intensive olive grove production, which is a particular type with its own spectral signature characteristics that differ from the "usual" open olive grove production (since the trees are small, have less space between them, and have no grass beneath). As such, the same LULC class can have different spectral signatures, which helps explain the low accuracy values for the permanent cropland class and the confusion between the remaining agricultural classes.

As mentioned before, the Beja municipality has a heterogeneous land cover area (with the existence of both fragmented parcels and larger and compact parcels) with different crop phenological cycles and agricultural practices. In addition, the most misclassification was between the permanent cropland and temporary crops with agroforestry areas presumably due to how olive groves (when not super-intensively produced) and especially the vineyard planting was organized (in line with spaces of bare soil between them). Also, the presence and extent of Montado in the Beja municipality can easily explain the misclassification among each of the three agricultural classes, since this class is mostly characterized by a herbaceous understory, and can, therefore, be easily mistaken with olive groves or pastures (Allen et al., 2018).

In particular, Allen et al. (2018) have emphasized the lack of research focusing on the detection of LULC changes in the Alentejo region, based on identifying the extent of more than one type of LULC at a large temporal scale. The approaches of Calvão and Palmeirim (2011) and Godinho et al. (2016) to identify the LULC changes in Alentejo were focused on single classes (semi-deciduous scrub, and Montado, respectively). As a matter of fact, Allen et al. (2018) were the first to assess the extent of the changes in the different LULC classes present in an agro-silvo-pastoral environment and recognized the limitations of such analysis. Our study reveals similar drawbacks that emerged in the classification confusion among the agricultural classes present in our study area.

Nevertheless, the final result is a starting point for understanding the LULC dynamics for 21 years—confirming the impact some political decisions have had in this region (such as the construction of the Alqueva Dam in 2002 that filled its reservoir until 2012) and the lack of control of the type of agricultural activity that has had a significant impact on soil quality due to the super-intensive production for permanent crops like the case of the olive groves.

3.5.4. The TWDTW Method for Long-Term Time-Series Classification

This study also introduced the use of the TWDTW method for long-term classification using the Landsat time series. The TWDTW had already successfully achieved high accuracy values for satellite time series, using MODIS imagery (Maus, Camara, et al., 2016; Maus, Câmara, et al., 2016) and Sentinel-2 imagery (Belgiu & Csillik, 2018), showing to a certain extent the capacity of this classifier for handling multidimensional analysis. In this study, the assessment revealed an overall good accuracy of about 76%. Although the water and forest classes had contributed to this overall value, the accuracy values obtained for each of the three agricultural classes (the agricultural classes were prevalent in the study area, occupying about 80% of the total area) were problematic. Certainly, the 24% misclassification rate among the LULC classes could have a statistical influence in the results, particularly in relation to the extent of the changes in classes over the 21-year period. However, our results are in accordance with the statistical analysis we performed with the COS maps (Chapter 3.2), which allowed identifying the main LULC classes in the study area and how they changed over time.

It is noteworthy that this misclassification among each of the three agricultural classes was already expected, due to the complex spectral properties of the different vegetation types present in these large agricultural classes. This is especially true for temporary crops with agroforestry areas. They have a complex cultivation pattern encompassing temporary irrigated and rain-fed crops that are produced and harvested in different seasons followed by bare soil afterwards that can be easily confused with the permanent cropland and permanent pastures. Furthermore, the classifications are complicated by the more recent super-intensive production for permanent crops like the case of the olive groves (Viana & Rocha, 2018). Intensification of the permanent cropland is very likely associated with increased exploitation of water resources, following the construction of Alqueva Dam. The increased expansion of water bodies is also apparently attributable to the construction of this Dam (Allen et al., 2018).

In addition, to improve the identification of each LULC type and increase the classification accuracy, we decided to compute two-time series indices (NDVI and NDWI) and perform the cluster analysis using the spectral signatures of the training samples for all LULC classes. However, we recognize that there may be other indices that can be used to further improve the results, such as the enhanced vegetation index (EVI) (Guan et al., 2016; Teluguntla et al., 2015; Wardlow et al., 2007), the generalized soil adjusted vegetation index (SAVI) (Gilabert et al., 2002), or the Modified Soil-adjusted Vegetation Index (MSAVI) (Qi et al., 1994). These should be considered in further studies.

Nevertheless, our results, in line with Maus et al. (Maus, Camara, et al., 2016), and Belgiu and Csillik (Belgiu & Csillik, 2018), highlight the potential of this method to classify complex ecosystems, especially if we take into account our methodology's major constraint, the sample set irregularly distributed over time. Yet, it is important to highlight the processing time associated to create the categorical LULC map for each period when using the dtwSat package. The TWDTW method processing time (using an Intel® Core™ i7-6700K CPU with a 4.00 GHz clock and 32 GB of RAM) was almost 16 hours when processing in parallel (Maus, Câmara, et al., 2016). This can be an issue to take into account, especially if we want to reproduce this approach for larger areas or if the available hardware has limited resources.

3.6. Conclusions

In this study, the TWDTW method was used to classify a heterogeneous area located in the southeast of Portugal, providing insights into the main changes that occurred over 21 years. This time series, derived from Landsat data, facilitated the inter-annual change comparisons in cropland type activity in an area of predominantly mixed cropland. If we consider the above-mentioned limitations (absence of images for 2012, the use of two different satellite sensors, and the training samples irregularly distributed over time) and the overall good classification accuracy value obtained, we are truly convinced that in the near future, this classifier will become one of the most used for time-series classification, especially because it is an open-source and free method for remote sensing time-series analysis.

The cluster analysis performed on R was efficient and straightforward. It proved itself a promising approach capable of detecting homogeneous training sample groups for each class. It guaranteed an acceptable level of quality, as we collected the samples based on vector cartography, the polygons constituted homogeneous entities, and the salt-and-pepper effect proliferated in the satellite images.

Also, we acknowledge that with the COS data we could have already had a notion of the major transformations, but we would not have had seen the trends from year to year and been able to quantify them. In addition, the time gap of the COS data was very large, especially between the first and the second maps (12 years apart). Also, this methodology can now be replicated for other study areas, using other LULC data (such as CORINE Land Cover (CLC), since it is available for most of Europe and for the years 1990, 2000, 2006, 2012, and 2018). Also, for more recent years, the OpenStreetMap data can be a reliable alternative since this source of voluntary geographic information has a level of accuracy similar to the COS.

For future studies, it would be important to try to disaggregate the three major agricultural classes into subclasses to make it possible to describe the type of crop in detail. Ultimately, we conducted a study, using only open data and open software that has allowed gaining more insight regarding a relatively large rural area (>1,000 km2). It is an area where the LULC changes have been dynamic, and the data provide support to improve cropland management with sustainable practices and policies.

References

A. Ramezan, C., A. Warner, T., & E. Maxwell, A. (2019). Evaluation of Sampling and Cross-Validation Tuning Strategies for Regional-Scale Machine Learning Classification. Remote Sensing, 11(2), 185. https://doi.org/10.3390/rs11020185

Allen, H., Simonson, W., Parham, E., Santos, E. de B. e, & Hotham, P. (2018). Satellite remote sensing of land cover change in a mixed agro-silvo-pastoral landscape in the Alentejo, Portugal.

International Journal of Remote Sensing, 00(00), 1–21. https://doi.org/10.1080/01431161.2018.1440095

Arsanjani, J. J., Mooney, P., Zipf, A., & Schauss, A. (2015). Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets (pp. 37–58). Springer, Cham. https://doi.org/10.1007/978-3-319-14280-7_3

Atzberger, C. (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. Remote Sensing, 5(2), 949–981. https://doi.org/10.3390/rs5020949

Azzali, S., & Menenti, M. (2000). Mapping vegetation-soil-climate complexes in southern Africa using temporal Fourier analysis of NOAA-AVHRR NDVI data. International Journal of Remote Sensing, 21(5), 973–996. https://doi.org/10.1080/014311600210380

Baessler, C., & Klotz, S. (2006). Effects of changes in agricultural land-use on landscape structure and arable weed vegetation over the last 50 years. Agriculture, Ecosystems and Environment, 115(1–4), 43–50. https://doi.org/10.1016/j.agee.2005.12.007

Baraldi, A., Puzzolo, V., Blonda, P., Bruzzone, L., & Tarantino, C. (2006). Automatic spectral rule-based preliminary mapping of calibrated landsat TM and ETM+ images. IEEE Transactions on Geoscience and Remote Sensing, 44(9), 2563–2585. https://doi.org/10.1109/TGRS.2006.874140

Belgiu, M., & Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sensing of Environment, 204(January 2017), 509–523. https://doi.org/10.1016/j.rse.2017.10.005

Brodley, C. E., & Friedl, M. A. (1999). Identifying Mislabeled Training Data. Journal of Artificial Intelligence Research, 11, 131–167. https://doi.org/10.1136/ard.2003.010348

Calvão, T., & Palmeirim, J. M. (2011). A comparative evaluation of spectral vegetation indices for the estimation of biophysical characteristics of mediterranean semi-deciduous shrub communities. International Journal of Remote Sensing, 32(8), 2275–2296. https://doi.org/10.1080/01431161003698245

Carrão, H., Gonçalves, P., & Caetano, M. (2008). Contribution of multispectral and multitemporal information from MODIS images to land cover classification. Remote Sensing of Environment, 112(3), 986–997. https://doi.org/10.1016/J.RSE.2007.07.002

Ceccato, P., Gobron, N., Flasse, S., Pinty, B., & Tarantola, S. (2002). Designing a spectral index to estimate vegetation water content from remote sensing data: Part 2. Validation and applications. Remote Sensing of Environment, 82(2–3), 198–207. https://doi.org/10.1016/S0034-4257(02)00036-6

Correia, T. P. (1993). Threatened landscape in Alentejo, Portugal: the ‘montado’ and other ‘agro-silvo-pastoral’ systems. Landscape and Urban Planning, 24(1–4), 43–48. https://doi.org/10.1016/0169-2046(93)90081-N

Csillik, O., & Belgiu, M. (2017). Cropland mapping from Sentinel-2 time series data using object-based image analysis. Agile 2017, 2016.

Cuesta-Albertos, J. A., Gordaliza, A., & Matrán, C. (1997). Trimmed k-means: An attempt to robustify quantizers. Annals of Statistics, 25(2), 553–576. https://doi.org/10.1214/aos/1031833664

DGT. (2018). Especificações Técnicas da Carta de Uso e Ocupação do Solo (COS) de Portugal Continental para 1995, 2007, 2010 e 2015.

EEA. (2000). Down to Earth: Soil Degradation and Sustainable Development in Europe. A challenge for the 21st century. (E. E. Agency (Ed.); Environmen).

FAO. (2010). The State of Food Insecurity in the World: Addressing food insecurity in protracted crises. In Notes. https://doi.org/10.1519/JSC.0b013e3181b8666e

Feranec, J., Soukup, T., Taff, G., Stych, P., & Bicik, I. (2017). Overview of changes in land cover and land use in Eastern Europe. In G. Gutman & V. Radeloff (Eds.), Land-cover and land-use changes in Eastern Europe after the collapse of the Soviet Union in 1991 (Springer).

Fritz, H., García-Escudero, L. A., & Mayo-Iscar, A. (2012). tclust : An R Package for a Trimming Approach to Cluster Analysis. Journal of Statistical Software, 47(12). https://doi.org/10.18637/jss.v047.i12

Fuchs, R., Herold, M., Verburg, P. H., Clevers, J. G. P. W., & Eberle, J. (2015). Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010. Global Change Biology, 21(1), 299–313. https://doi.org/10.1111/gcb.12714

Gao, B. C. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

García-Álvarez, D. (2018). The Influence of Scale in LULC Modeling. A Comparison Between Two Different LULC Maps (SIOSE and CORINE). In M. T. Camacho Olmedo, M. Paegelow, J. F. Mas, & F. Escobar (Eds.), Geomatic Simulations and Scenarios for Modelling LUCC. A Review and Comparison of Modelling Techniques (pp. 187–213). Springer, Cham. https://doi.org/10.1007/978-3-319-60801-3_10

Gilabert, M. ., González-Piqueras, J., García-Haro, F. ., & Meliá, J. (2002). A generalized soil-adjusted vegetation index. Remote Sensing of Environment, 82(2–3), 303–310. https://doi.org/10.1016/S0034-4257(02)00048-2

Godinho, S., Gil, A., Guiomar, N., Neves, N., & Pinto-Correia, T. (2016). A remote sensing-based approach to estimating montado canopy density using the FCD model: a contribution to identifying HNV farmlands in southern Portugal. Agroforestry Systems, 90(1), 23–34. https://doi.org/10.1007/s10457-014-9769-3

Godinho, S., Guiomar, N., Machado, R., Santos, P., Sá-Sousa, P., Fernandes, J. P., Neves, N., & Pinto-Correia, T. (2016). Assessment of environment, land management, and spatial variables on recent changes in montado land cover in southern Portugal. Agroforestry Systems, 90(1), 177–192. https://doi.org/10.1007/s10457-014-9757-7

Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 116, 55–72. https://doi.org/10.1016/J.ISPRSJPRS.2016.03.008

Guan, X., Huang, C., Liu, G., Meng, X., & Liu, Q. (2016). Mapping rice cropping systems in Vietnam using an NDVI-based time-series similarity measurement based on DTW distance. Remote Sensing, 8(1). https://doi.org/10.3390/rs8010019

Gumma, M. K., Thenkabail, P. S., Teluguntla, P., Rao, M. N., Mohammed, I. A., & Whitbread, A. M. (2016). Mapping rice-fallow cropland areas for short-season grain legumes intensification in South Asia using MODIS 250 m time-series data. International Journal of Digital Earth, 9(10), 981–1003. https://doi.org/10.1080/17538947.2016.1168489

Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. Remote Sensing of Environment, 122, 66–74. https://doi.org/10.1016/j.rse.2011.08.024

Hossain, M. S., Bujang, J. S., Zakaria, M. H., & Hashim, M. (2015). Assessment of the impact of Landsat 7 Scan Line Corrector data gaps on Sungai Pulai Estuary seagrass mapping. Applied Geomatics, 7(3), 189–202. https://doi.org/10.1007/s12518-015-0162-3

Huang, X., Weng, C., Lu, Q., Feng, T., & Zhang, L. (2015). Automatic labelling and selection of training samples for high-resolution remote sensing image classification over urban areas. Remote Sensing, 7(12), 16024–16044. https://doi.org/10.3390/rs71215819

Instituto Português do Mar e da Atmosfera (IPMA). Iberian Climate Atlas. Available online: http://www.ipma.pt/resources.www/docs_pontuais/ocorrencias/2011/atlas_clima_iberico.pdf (accessed on 10 January 2019).

Khatami, R., Mountrakis, G., & Stehman, S. V. (2016). A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. Remote Sensing of Environment, 177, 89–100. https://doi.org/10.1016/J.RSE.2016.02.028

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data Published by : International Biometric Society Stable URL : http://www.jstor.org/stable/2529310 Accessed : 02-03-2016 17 : 57 UTC Your use of the JSTOR archive indicates your acceptance of the Ter. Biometrics, 33(1), 159–174.

Lippitt, C. D., Rogan, J., Li, Z., Eastman, J. R., & Jones, T. G. (2008). Mapping Selective Logging in Mixed Deciduous Forest: A Comparison of Machine Learning Algorithms. Photogrammetric Engineering & Remote Sensing, 74(10), 1201–1211.

Lobo, A., Legendre, P., Rebollar, J. L. G., Carreras, J., & Ninot, J. M. (2004). Land cover classification at a regional scale in Iberia: Separability in a multi-temporal and multi-spectral data set of satellite images. International Journal of Remote Sensing, 25(1), 205–213. https://doi.org/10.1080/0143116031000116435

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing, 28(5), 823–870. https://doi.org/10.1080/01431160600746456

Lu, Q., Ma, Y., & Xia, G. S. (2017). Active learning for training sample selection in remote sensing image classification using spatial information. Remote Sensing Letters, 8(12), 1210–1219. https://doi.org/10.1080/2150704X.2017.1375610

Lunetta, R. S., Knight, J. F., Ediriwickrema, J., Lyon, J. G., & Worthy, L. D. (2006). Land-cover change detection using multi-temporal MODIS NDVI data. Remote Sensing of Environment, 105(2), 142–154. https://doi.org/10.1016/j.rse.2006.06.018

Lyle, G., Lewis, M., & Ostendorf, B. (2013). Testing the temporal ability of landsat imagery and precision agriculture technology to provide high resolution historical estimates of wheat yield at the farm scale. Remote Sensing, 5(4), 1549–1567. https://doi.org/10.3390/rs5041549

Maus, V., Câmara, G., Appel, M., & Pebesma, E. (2016). dtwSat: Time-Weighted Dynamic Time Warping for Satellite Image Time Series Analysis in R. Journal of Statistical Software.

Maus, V., Camara, G., Cartaxo, R., Sanchez, A., Ramos, F. M., & Ribeiro, G. Q. (2016). A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(8), 3729–373910. https://doi.org/10.1109/JSTARS.2016.2517118

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens., 17, 1425–1432.

Meneses, B. M., Reis, E., Vale, M. J., & Reis, R. (2018). Modelling the Land Use and Land cover changes in Portugal: A multi-scale and multi-temporal approach. Finisterra, 53(107). https://doi.org/10.18055/Finis12258

Meroni, M., Marinho, E., Sghaier, N., Verstrate, M. M., & Leo, O. (2013). Remote sensing based yield estimation in a stochastic framework - Case study of durum wheat in Tunisia. Remote Sensing, 5(2), 539–557. https://doi.org/10.3390/rs5020539

Mulianga, B., Bégué, A., Simoes, M., & Todoroff, P. (2013). Forecasting regional sugarcane yield based on time integral and spatial aggregation of MODIS NDVI. Remote Sensing, 5(5), 2184–2199. https://doi.org/10.3390/rs5052184

Noszczyk, T., Rutkowska, A., & Hernik, J. (2019). Exploring the land use changes in Eastern Poland: statistics-based modeling. Human and Ecological Risk Assessment: An International Journal, 1–28. https://doi.org/10.1080/10807039.2018.1506254

Petitjean, F., Inglada, J., & Gancarski, P. (2012). Satellite Image Time Series Analysis Under Time Warping. Geoscience and Remote Sensing, IEEE Transactions On, 50(8), 3081–3095. https://doi.org/10.1109/TGRS.2011.2179050

Petitjean, F., & Weber, J. (2014). Efficient Satellite Image Time Series Analysis Under Time Warping. Geoscience and Remote Sensing Letters, IEEE, 11(6), 1143–1147. https://doi.org/10.1109/LGRS.2013.2288358

Phiri, D., & Morgenroth, J. (2017). Developments in Landsat land cover classification methods: A review. Remote Sensing, 9(9). https://doi.org/10.3390/rs9090967

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. Remote Sensing of Environment, 48(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

Raczko, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. European Journal of Remote Sensing, 50(1), 144–154. https://doi.org/10.1080/22797254.2017.1299557

Radwan, T. M., Blackburn, G. A., Whyatt, J. D., & Atkinson, P. M. (2019). Dramatic Loss of Agricultural Land Due to Urban Expansion Threatens Food Security in the Nile Delta, Egypt. Remote Sensing, 11(3), 332. https://doi.org/10.3390/rs11030332

Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., & Ohlen, D. O. (1994). Measuring phenological variability from satellite imagery. Journal of Vegetation Science, 5(5), 703–714. https://doi.org/10.2307/3235884

Ross, C., Fildes, S., & Millington, A. (2017). Land-Use and Land-Cover Change in the Páramo of South-Central Ecuador, 1979–2014. Land, 6(3), 46. https://doi.org/10.3390/land6030046

Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. Third Earth Resources Technology Satellite (ERTS) Symposium, 1, 309–317. https://doi.org/citeulike-article-id:12009708

Rousseeuw, P. J. (1987). Silhouettes : a graphical aid to the interpretation and validation of cluster analysis. 20, 53–65.

Rufin, P., Müller, H., Pflugmacher, D., & Hostert, P. (2015). Land use intensity trajectories on Amazonian pastures derived from Landsat time series. International Journal of Applied Earth Observation and Geoinformation, 41, 1–10. https://doi.org/10.1016/j.jag.2015.04.010

Russo, R. O. (1996). Agrosilvopastoral Systems: A Practical Approach Toward Sustainable Agriculture. Journal of Sustainable Agriculture, 7(4), 5–16. https://doi.org/10.1300/J064v07n04_03

Sakoe, H., & Chiba, S. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing, 26(1), 43–49. https://doi.org/10.1109/TASSP.1978.1163055

Santos, L. A., Simoes, R. E. O., Pletsch, M. A. J. S., Santos, L. A., Camara, G., & Maus, V. (2017). Satellite multisensor spatiotemporal analysis : a TWDTW preview approach. August.

Schmidt, M., Lucas, R., Bunting, P., Verbesselt, J., & Armston, J. (2015). Multi-resolution time series imagery for forest disturbance and regrowth monitoring in Queensland, Australia. Remote Sensing of Environment, 158, 156–168. https://doi.org/10.1016/j.rse.2014.11.015

Senf, C., Leitão, P. J., Pflugmacher, D., van der Linden, S., & Hostert, P. (2015). Mapping land cover in complex Mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. Remote Sensing of Environment, 156, 527–536. https://doi.org/10.1016/J.RSE.2014.10.018

Serra, P., Pons, X., & Saurí, D. (2008). Land-cover and land-use change in a Mediterranean landscape: A spatial analysis of driving forces integrating biophysical and human factors. Applied Geography, 28(3), 189–209. https://doi.org/10.1016/j.apgeog.2008.02.001

Seto, K. C., & Fragkias, M. (2005). Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics. Landscape Ecology, 20(7), 871–888. https://doi.org/10.1007/s10980-005-5238-8

Shi, K., Chen, Y., Yu, B., Xu, T., Li, L., Huang, C., Liu, R., Chen, Z., & Wu, J. (2016). Urban Expansion and Agricultural Land Loss in China: A Multiscale Perspective. Sustainability, 8(8), 790. https://doi.org/10.3390/su8080790

Team, R. C. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. ISBN3-900051-07-0 https://www. R-project. org.

Teluguntla, P., Ryu, D., George, B., Walker, J., & Malano, H. (2015). Mapping Flooded Rice Paddies Using Time Series of MODIS Imagery in the Krishna River Basin, India. Remote Sensing, 7(7), 8858–8882. https://doi.org/10.3390/rs70708858

Usman, B. (2013). Satellite Imagery Land Cover Classification using K-Means Clustering Algorithm Computer Vision for Environmental Information Extraction. Sci. & Engg, 63(October 2013), 18671–18675. https://doi.org/10.1145/1837853.1693485

Viana, C. M., Encalada, L., & Rocha, J. (2019). The value of OpenStreetMap Historical Contributions as a Source of Sampling Data for Multi-temporal Land Use/Cover Maps. ISPRS International Journal of Geo-Information, 8(3), 116. https://doi.org/10.3390/ijgi8030116

Viana, C. M., Encalada, L., Rocha, J., Viana, C. M., Encalada, L., & Rocha, J. (2019). The value of OpenStreetMap Historical Contributions as a Source of Sampling Data for Multi-temporal Land Use/Cover Maps. ISPRS International Journal of Geo-Information, 8(3), 116. https://doi.org/10.3390/ijgi8030116

Viana, C. M., Girão, I., & Rocha, J. (2019). Training samples from open data for satellite imagery classification: using K-means clustering algorithm. In P. Kyriakidis, D. Hadjimitsis, D. Skarlatos, & A. Mansourian (Eds.), Accepted Short Papers and Posters from the 22nd AGILE Conference on Geo-information Science. (pp. 978-90-816960-9–8). Stichting AGILE.

Viana, C. M., & Rocha, J. (2018). Spatiotemporal analysis and scenario simulation of agricultural land use land cover using GIS and a Markov chain model. In A. Mansourian, P. Pilesjö, L. Harrie, & R. von Lammeren (Eds.), Geospatial Technologies for All : short papers, posters and poster abstracts of the 21th AGILE Conference on Geographic Information Science.

Vicente-Serrano, S. M., & Heredia-Laclaustra, A. (2004). NAO influence on NDVI trends in the Iberian peninsula (1982-2000). International Journal of Remote Sensing, 25(14), 2871–2879. https://doi.org/10.1080/01431160410001685009

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. Remote Sensing of Environment, 108(3), 290–310. https://doi.org/10.1016/j.rse.2006.11.021

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S. N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P. S., Vermote, E. F., Vogelmann, J., Wulder, M. A., & Wynne, R. (2008). Free Access to Landsat Imagery. Science, 320(5879), 1011a-1011a. https://doi.org/10.1126/science.320.5879.1011a

Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. Remote Sensing of Environment, 122, 2–10. https://doi.org/10.1016/j.rse.2012.01.010

Xiao, X., Boles, S., Liu, J., Zhuang, D., Frolking, S., Li, C., Salas, W., & Moore, B. (2005). Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. Remote Sensing of Environment, 95(4), 480–492. https://doi.org/10.1016/j.rse.2004.12.009

Yan, L., & Roy, D. P. (2015). Improved time series land cover classification by missing-observation-adaptive nonlinear dimensionality reduction. Remote Sensing of Environment, 158, 478–491. https://doi.org/10.1016/J.RSE.2014.11.024

Yuan, F., & Bauer, M. E. (2007). Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. Remote Sensing of Environment, 106(3), 375–386. https://doi.org/10.1016/j.rse.2006.09.003

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. Remote Sensing of Environment, 84(3), 471–475. https://doi.org/10.1016/S0034-4257(02)00135-9

Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. Remote Sensing of Environment, 144, 152–171. https://doi.org/10.1016/J.RSE.2014.01.011

Chapter 4. The value of OpenStreetMap Historical Contributions as a Source of Sampling Data for Multi-temporal Land Use/Cover Maps

Cláudia M. Viana, Luis Encalada, Jorge Rocha

Article published at the International Journal of Geo-Information (2019), Volume 8, Part 3, 116. https://doi.org/10.3390/ijgi8030116

The structure and formatting are slightly adapted.

Abstract

OpenStreetMap (OSM) is a free, open-access Volunteered geographic information (VGI) platform that has been widely used over the last decade as a source for Land Use Land Cover (LULC) mapping and visualization. However, it is known that the spatial coverage and accuracy of OSM data are not evenly distributed across all regions, with urban areas being likelier to have promising contributions (in both quantity and quality) than rural areas. The present study used OSM data history to generate LULC datasets with one-year timeframes as a way to support regional and rural multi-temporal LULC mapping. We evaluated the degree to which the different OSM datasets agreed with two existing reference datasets (CORINE Land Cover and the official Portuguese Land Cover Map). We also evaluated whether our OSM dataset was of sufficiently high quality (in terms of both completeness accuracy and thematic accuracy) to be used as a sampling data source for multi-temporal LULC maps. In addition, we used the near boundary tag accuracy criterion to assesses the fitness of the OSM data for producing training samples, with promising results. For each annual dataset, the completeness ratio of the coverage area for the selected study area was low. Nevertheless, we found high thematic accuracy values (ranged from 77.3% to 91.9%). Additionally, the training samples thematic accuracy improved as they moved away from the features' boundaries. Features with larger areas (> 10 ha), e.g. Agriculture and Forest, had a steadily positive correlation between training samples accuracy and distance to feature boundaries.

Keywords: OpenStreetMap (OSM); Volunteered Geographic Information (VGI); land use land cover; mapping; accuracy; sampling data

4.1. Introduction

Since the end of the 20th century, land use and land cover (LULC) maps have been extensively generated at different spatial and temporal scales. The 2000 Global Land Cover map (Fritz et al., 2003) and the 2000 CORINE Land Cover (CLC) (Büttner & Feranec, 2002) map are two examples. Multi-temporal LULC maps can be used to monitor land changes over time, enabling the creation of indicators that can measure changes and support land management. Mapping results have a significant impact on our understandings of LULC patterns and can affect monitoring, characterization, and quantification outcomes. Nevertheless, one of the main challenges regarding LULC map production is the difficulty of distinguishing and accurately mapping land attributes.

Over the last decade, volunteered geographic information (VGI) platforms (Goodchild, 2007), as well as other data contributed by social network communities (Estima & Painho, 2013b; See, Mooney, et al., 2016), have been widely used as sources for LULC mapping and visualizations (Arsanjani, Mooney, & Zipf, 2015; Estima & Painho, 2013a, 2014, 2015; Fonte & Martinho, 2017; Fonte et al., 2017; Arsanjani & Vaz, 2015; See, Estima, et al., 2016). Crowdsourced content from online platforms, accessed, and exchanged by citizens, has emerged as a supplementary data source with significant implications for LULC database production (Sui et al., 2013). This nontraditional data source is not necessarily a substitute for official data, but is considered to be complementary (Goodchild & Li, 2012).

OpenStreetMap (OSM) is a free, open-access VGI platform to which volunteers from all over the world collaboratively contribute data, and it is a particularly promising source of information for LULC analysis (Arsanjani, Mooney, & Zipf, 2015; Haklay, 2010; Mooney & Corcoran, 2014). Several factors have been essential to its success, including its availability of up-to-date data with global coverage, improvements in data quality (mainly driven by an increase in the number of contributors over time) (Fonte et al., 2017; Zook, Graham, Shelton, & Gorman, 2010), and its extensive volume and variety of thematic attribute data (Fonte et al., 2017). OSM therefore has potential for use in the long-term mapping and monitoring of LULC changes (Estima & Painho, 2013a) and could plausibly be used to improve the production, verification, and validation of LULC maps (Arsanjani et al., 2013; Estima & Painho, 2013a, 2015; Fonte & Martinho, 2017; Fonte et al., 2017; Johnson & Iizuka, 2016; Yang et al., 2017). A multi-temporal trajectory can be achieved using OSM data (Neis et al., 2012), not only for LULC mapping but also for ground-validated data creation (Yang et al., 2017).

However, it is known that both the spatial coverage and the accuracy of OSM data are not homogeneous across all regions (Haklay, 2010; Hecht et al., 2013), with urban areas being likelier to have promising contributions (in both quantity and quality) than rural areas (Helbich et al., 2012; Peter Mooney et al., 2012). Several authors have recently highlighted quality issues on VGI platforms (Almendros-Jiménez & Becerra-Terón, 2018; Antoniou & Skopeliti, 2015; Comber et al., 2013; Estima & Painho, 2013b; Fonte et al., 2015; See et al., 2013; Senaratne et al., 2017; Tracewski et al., 2017). OSM data still lack formal standards, such as those established by the International Organization for Standardization (ISO) 19157: 2013 Geographic information-Data quality (Dorn et al., 2015; Fonte et al., 2017). Five data quality criteria are commonly mentioned in the literature: (1) completeness, (2) temporal accuracy, (3) logical consistency, (4) positional accuracy, and (5) thematic accuracy (Arsanjani, Mooney, Zipf, et al., 2015; Arsanjani & Fonte, 2016; Guptill et al., 1995).

The thematic accuracy (Foody et al., 2013) of the OSM platform's LULC mapping is currently a hot research topic (Arsanjani, Mooney, & Zipf, 2015; Arsanjani & Fonte, 2016; Arsanjani & Vaz, 2015; Estima & Painho, 2015; Fonte et al., 2017), and a number of studies have compared OSM data with authoritative reference data (Arsanjani, Mooney, Zipf, et al., 2015; Dorn et al., 2015; Estima & Painho, 2013a, 2015; Haklay, 2010; Yagoub, 2017). For example, two studies of mainland Portugal (Estima & Painho, 2013a, 2015) found a total of 76.7% agreement between the data from OSM and CLC maps, with artificial surfaces, forests, and water bodies presenting promising results. Similarly, a study of Vienna (Arsanjani et al., 2013) compared OSM data to data from the Global Monitoring for Environment and Security Urban Atlas (GMESUA) and found agreement of 76% to 91%. More recently, another comparison of OSM and GMESUA data (Fonte & Martinho, 2017) found an agreement of about 90%.

Completeness accuracy (Hecht et al., 2013) is also commonly used to assess OSM data since it allows for the evaluation of territorial coverage. It has been used mainly to evaluate the completeness of OSM's data on road networks (Barron et al., 2014; Nasiri et al., 2018; Neis et al., 2012), buildings (Brovelli & Zamboni, 2018; Hecht et al., 2013; Rehrl et al., 2016), and LULC features (Arsanjani & Fonte, 2016; Arsanjani & Vaz, 2015; Estima & Painho, 2013a).

4.1.1. Related Works

Recent studies (Antoniou et al., 2016; Jonietz & Zipf, 2016; Sehra et al., 2017) have also included OSM contributors' update history in data quality assessments. OSM full history file stores extra information associated with data contributions, such as timestamps that provide the exact date and time of contributions (Barron et al., 2014; Nasiri et al., 2018; Rehrl et al., 2016). For example, one study (Nasiri et al., 2018) that assessed OSM's road network completeness and positional accuracy using OSM data history concluded that the use of historical information improved the quality of OSM data by up to 14%. OSM data history can moreover be used for a number of different purposes (Nasiri et al., 2018) and allow for new perspectives on the reliability of OSM data at different scales and timeframes (Nasiri et al., 2018; Neis et al., 2012). OSM data history is available since 2005.

In the present study, OSM data history has been accessed to generate different LULC datasets based on the contribution year (using the timestamps) in order to create regional and rural multi-temporal LULC maps. This had two primary purposes. First, we sought to evaluate the degree to which OSM datasets bounded by one-year timeframes agreed with extant authoritative datasets (i.e., CLC and the official Portuguese Land Cover Map [COS]), using both completeness and thematic accuracy as quality parameters. Second, we sought to evaluate whether the OSM datasets were of sufficient quality to be used as sampling data sources for multi-temporal LULC maps. For this second assessment, we used near boundary tag accuracy (NBTA) to evaluate the fitness of the OSM data for producing training samples, by looking at the extent to which a feature's proximity to a boundary influenced its attribute (tag) accuracy.

This research was conducted in the largest Portuguese district, Beja. It is a predominantly rural region with high natural and economic value, and is characterized by a mixed agro-silvo-pastoral ecosystem (Viana & Rocha, 2018). In addition, the region has recently undergone rapid LULC changes (Allen et al., 2018; Viana et al., 2019; Viana & Rocha, 2018), and the limited number of reference LULC data for this region emphasizes the importance of finding supplementary LULC data to support the identification and monitoring of these changes.

Our research is discussed in the rest of this paper. Chapter 4.2 describes the study area and our data. Chapter 4.3 details our methodology. Chapter 4.4 presents our OSM quality results. Finally, in Chapter 4.5 we discuss the implications and main conclusions of our results.

4.2. Study Area and Data

4.2.1. Study Area

Beja is a district located in the southeast of Portugal, in the Alentejo region (Figure 4.1). It is the largest Portuguese district, with an area of 10,229.05 km², and as of 2011 it had a population of 152,758 residents (INE, 2012). Urban density is low, and agricultural and forest areas dominate. The southeastern part of the district is flat, while in the northern and western parts the extensive plains are intersected by tiny hills. The valley of the River Guadiana, which traverses the eastern part of the district in a north-south direction, is the district's main geographical feature.

Figure 4.1. Location of Beja District and its Municipalities

4.2.2. Datasets

4.2.2.1. The OSM dataset and history file

OSM data can be freely downloaded, for example from Geofabrik or Planet OSM, in the form of raw datasets covering different regions (countries, continents, or any other administrative level) around the globe. OSM data represent physical features (objects) on the ground, and their tags (i.e., labels) are used to describe the objects (i.e., class description). The OSM data include a variety of physical feature types, with land use, natural features, waterways, amenities, and highways being the most commonly represented (Mooney & Corcoran, 2012). Feature descriptions can be found on the OSM wiki page.

OSM history file include records of all historical contributions, including recent ones, and are accessible as either XML or PBF formatted files. They provide a history of every modification made to a geographical feature's shape or tag. This means that if, for example, a feature's shape has been changed once, there will then be two entries for the same feature—the original feature and the modified one. A sample illustration and more information can be found in Nasiri et al. (2018). In the present study, we accessed the history file to see the timestamps (day/month/year and time) of each feature's creation and modifications.

4.2.2.2. Official reference datasets

The two reference LULC datasets used in this study were the 2012 CLC and the 2015 official Portuguese COS. The CLC is produced by the Portuguese General Directorate for Territorial Development (DGT) in coordination with the European Environment Agency, while the COS is produced exclusively by the DGT. Both datasets are freely available for download from the DGT website, and both use hierarchical and *a priori* nomenclature systems. The COS nomenclature was produced to match the CLC one. Thus, despite the fact that COS has five disaggregation levels compared to the CLC's three, the COS's first three levels are similar to the three CLC levels, thereby enabling comparisons between them. The dataset characteristics and metadata are shown in Table 4.1, and their descriptive statistics are shown in Table 4.2. More details about their nomenclature characteristics can be seen in Estima and Painho (Estima & Painho, 2015).

Table 4.1. Characteristics of the CORINE Land Cover (CLC) and the official Portuguese Land Cover Map (COS) datasets

Characteristics	CLC	COS
Data model	Vector	
Spatial representation	Polygons	
Nomenclature	Hierarchical (3 levels - 44 classes)	Hierarchical (5 levels - 225 classes)
Scale	1:100,000	1:25,000
Spatial resolution	20 m	0.5 m
Minimum Mapping Unit (MMU)	25 ha	1 ha
Minimum distance between lines	100 m	20 m
Base data	Satellite images	Air-photo maps
Production method	Semi-automated production and visual interpretation	Visual interpretation

Table 4.2. Descriptive statistics of reference datasets (area in ha)

	Beja District	
	CLC (2012)	COS (2015)
Total polygons	11,306	34,793
Minimum area of polygons	0.01	0.01
Maximum area of polygons	107,667.31	26,542.24
Mean area of polygons	90,78	29.50
Standard Deviation area	1,065.51	280.20

4.3. Methods

Figure 4.2 shows our main steps. Briefly, our first step was to download the OSM data history file and filter the contributions according to their timestamps. We obtained datasets for seven different years and resolved all logical inconsistencies, such as overlapping features. Second, we established a relationship between the OSM and CLC/COS first level of nomenclature. Third, we intersected the datasets to determine the area corresponding to the OSM dataset that matched the reference dataset. Fourth, we calculated the datasets' completeness and thematic accuracies for 2012 and 2015. Fifth, random points were generated as training samples. Finally, in step six we calculated the NBTA for 2015.

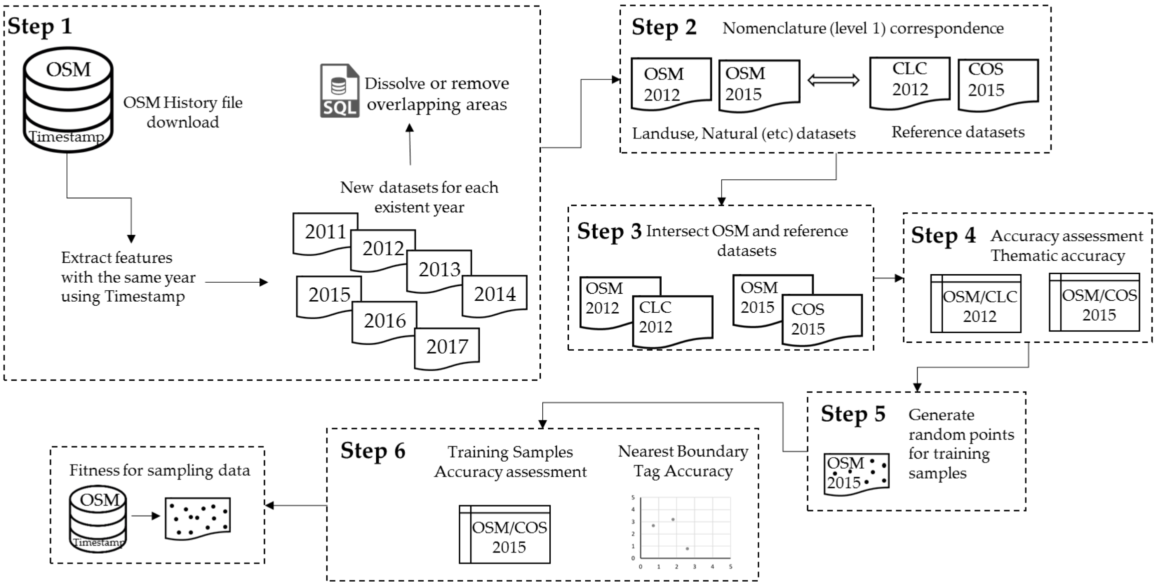


Figure 4.2. Workflow representing the proposed methodology

4.3.1. Processing the OSM Datasets

We began by downloading from Geofabrik the latest OSM history file available at the time (May 7, 2018) that covered Portugal. All OSM objects with the feature types of land use, natural features, airways, amenities, buildings, highways, historic features, leisure features, man-made features, power structures, public transportation, railroads, shops, sports, tourism, forests, and waterways were retrieved. We limited our results to features (polygons) found in the Beja district. By filtering the features by contribution year (using the timestamp data), we generated multi-temporal datasets that included only the features which were created or modified during any given year. We found contributions for seven different years (2011– 2017). Since reference data for the study area only exists for 2012 (CLC) and 2015 (COS), we identified and selected all the OSM data contributions that had been created/modified in these two years (2012 and 2015) and used these two OSM datasets in our study for comparison with the reference datasets.

A common problem when processing OSM data is the presence of logical inconsistencies, such as overlapping features (Arsanjani, Mooney, Zipf, et al., 2015; Arsanjani & Fonte, 2016; Arsanjani & Vaz, 2015; Estima & Painho, 2013a). This problem is even more common when history files are used, since they contain records of every single modification. It is therefore essential to dissolve or remove the overlapping features in order to not overestimate areas and to ensure that only one attribute (tag) is kept for analysis (Arsanjani, Mooney, Zipf, et al., 2015; Arsanjani & Vaz, 2015; Estima & Painho, 2013a). To ensure logical consistency between the annual datasets, when overlapped features had the same attribute we dissolved the overlapping areas, while when the overlapped features' attributes did not match the conflict areas were removed.

In addition, we examined the provenance of OSM data for our study area to confirm that there were no bulk imports from known sources (e.g. CLC; COS). In this study the polygons´ geometry for each reference dataset and both OSM datasets was compared by using the Feature Compare tool of ArcGIS 10.5 software.

4.3.2. Relationship Between Dataset Nomenclatures

Several studies have stressed the difficulties of using datasets from different agencies due to the lack of direct relationships between their classes (Arsanjani, Mooney, Zipf, et al., 2015; Arsanjani & Vaz, 2015; Estima & Painho, 2013a; Haack et al., 2015). In Estima and Painho (Estima & Painho, 2013a), the authors attempted to reconcile the three nomenclature levels of the CLC, the OSM land use feature, and the OSM natural areas feature, based on the official descriptions of the CLC and OSM classes. Given their remarkable results, we decided to partially follow their nomenclature correspondence for the first level of the CLC and COS nomenclature, namely, (1) artificial surfaces, (2) agricultural areas, (3) forests, (4) wetlands, and (5) water (see Table 4.3). For the purposes of the study, the OSM features type airways, amenities, buildings, highways, historic features, leisure features, man-made features, power structures, public transportation, railroads, shops, sports, and tourism were considered to be (1) artificial surfaces. The OSM feature type wood was classed as (3) forest, while the OSM feature type waterway was classed as (5) water.

Table 4.3. Nomenclature correspondence for the first level of the CLC and COS nomenclature

Landuse feature type			
OSM tag	CLC/COS (Level 1)	OSM tag	CLC/COS (Level 1)
Abutters	1	Harbour	1
Allotments	2	Industrial	1
Basin	5	Landfill	1
Beach	3	Leisure	1
Brownfield	1	Meadow	2
Cemetery	1	Military	?
Commercial	1	Museum	1
Conservation	3	Not_known	?
Construction	1	Orchard	2
Farm	2	Park	1
Farmland	2	Public	1
Farmyard	2	Quarry	1
Garages	1	Railway	1
Garden	1	Recreation_groun	1
Grass	2	Reservoir	5
Greenfield	3	Retail	1
Greenhouse	2	Salt_pond	4
Greenhouse_horti	2	Scrub	3
Residential	1	Scrubs	3
University	1	Vineyard	2
Village_green	1	Waste_water_plan	1
Wood	3	Water	5
Natural feature type			
OSM tag	CLC/COS (Level 1)	OSM tag	CLC/COS (Level 1)
Grassland	2	Fell	3
Scrub	3	Bare_rock	3
Wood	3	Park	3
Scree	3	Forest	3
Beach	3	Wetland	4
Sand	3	Water	5
Rock	3	Riverbank	5

4.3.3. Accuracy Assessment Criteria

4.3.3.1. OSM completeness accuracy

It can be difficult to draw definitive conclusions from OSM data due to the heterogeneity of contributions across different regions (Hecht et al., 2013). This challenge poses particular limitations when analyzing rural areas (Haklay, 2010; Zielstra & Zipf, 2010). Accordingly, the completeness accuracy criterion is commonly used to assess the quality of OSM data (Arsanjani, Mooney, Zipf, et al., 2015; Arsanjani & Fonte, 2016; Neis et al., 2012). Completeness accuracy is defined as the completeness of a dataset and measures the presence or absence of features. Typically, the total number of features is computed for point features, while for line features the total length is computed. These calculations are then compared to the reference dataset (Hecht et al., 2013). However, for polygon features, it is assumed that the overall area of the study area is the maximum area possible, and it is therefore not necessary to compare it to reference data (Arsanjani, Mooney, Zipf, et al., 2015).

In this study, completeness accuracy was calculated as a ratio of the OSM features' overall area (A_{OSM}) and the total area of the study area (A_{Ref}). It is presented here as a percentage, as described in Equation 4.1. A ratio of 100 means that the OSM dataset provides full coverage of the study area. This criterion was measured for all OSM dataset features for 2012 and 2015 for each LULC class.

$$Completeness = \frac{A_{OSM}}{A_{Ref}} X\ 100 \tag{4.1}$$

4.3.3.2. OSM thematic accuracy

Thematic accuracy is another common criterion used to evaluate the quality of OSM dataset features in LULC mapping (Arsanjani et al., 2013; Estima & Painho, 2013a, 2015; Fonte & Martinho, 2017). Thematic accuracy describes the accuracy of the features' attributes (tags) by computing the differences between OSM dataset features and those of the chosen reference datasets (Arsanjani, Mooney, Zipf, et al., 2015).

In this study, we used an overlap function to determine which features overlapped between the OSM datasets (2012 and 2015) and the reference datasets (CLC and COS, respectively). Following common statistical approaches (Arsanjani, Mooney, Zipf, et al., 2015; Herold et al., 2008), for each annual dataset the overlapping areas were computed in a confusion matrix, which presented the correct and incorrect mapped areas for each LULC class. The rows denote the occurrences of an actual class (OSM) and the columns denote the occurrences of reference data classes (CLC or COS). Several measures were obtained from this, including overall thematic accuracy, individual user accuracy, producer accuracy, and the Kappa index of agreement (Arsanjani, Mooney, Zipf, et al., 2015; Arsanjani & Fonte, 2016; Landis & Koch, 1977). The overall thematic accuracy measure provides the overall percentage of the correctly mapped OSM features by dividing the correct mapped area by the total mapped area. User accuracy measures the probability that any given LULC class from an OSM dataset will actually match the reference dataset, while producer accuracy indicates the probability that a particular LULC class from the reference dataset is classified as such in OSM (Arsanjani, Mooney, Zipf, et al., 2015). The Kappa index measures the degree of agreement between the OSM dataset and the reference dataset (Landis & Koch, 1977).

4.3.3.3. Near boundary tag accuracy (NBTA)

We also used NBTA to measure the fitness of the OSM data as a source of sampling data to support regional and rural LULC mapping. NBTA measures the extent to which the proximity of an OSM feature's boundary influences the accuracy of the attribute (tag) in the training sample. We computed the NBTA for the most recent OSM dataset (2015). First, it was necessary to create training samples by generating random points inside the OSM features. Since OSM feature areas are very heterogeneous, it would be inappropriate specify an exact number of random points to be generated inside each OSM feature; instead, the total number of random points generated was proportional to the area of each OSM feature. In addition, the shortest distance allowed between any two randomly placed points was 30 m, because this is the most common spatial resolution of satellite images, such as Landsat (Figure 4.3). We used point features instead of polygon features since this minimized the effects of comparing data with different mapping scales. Each random point generated took the attribute of the corresponding OSM polygon feature.

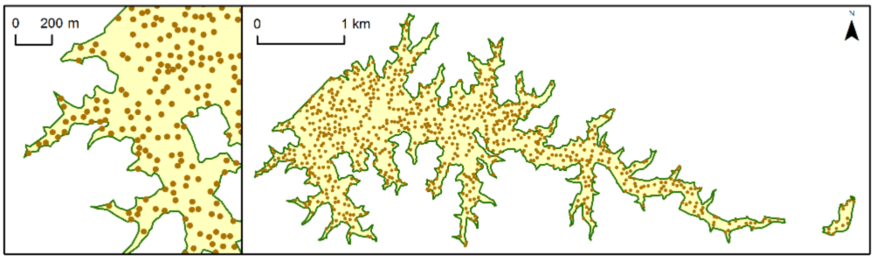


Figure 4.3. Example of random points distribution

Second, thematic accuracy was assessed following the same procedure described in Chapter 4.3.3.2. Also, the Euclidean distance from each random point to the nearest segment of the OSM feature's boundary was computed. The distance values (in ascending order) and the accumulated thematic accuracy were then plotted. The purpose of this process was to ascertain whether the training samples generated for each LULC class near the border of an OSM feature might be inherently less accurate.

However, the relationship between the training samples accuracy and their proximity to the corresponding feature boundaries did not consider the influence of each feature area in explaining the degree of accuracy, since the distance to a feature boundary varies across feature areas. Are features with larger or smaller areas likely to have more or less accuracy? Therefore, the distance values from all training samples were standardized using the maximum and minimum distance values of the corresponding OSM feature. Similarly, the accumulated accuracy of all training samples was computed by considering each OSM feature. The relationship between a feature's boundary proximity and its thematic accuracy is denoted by the Pearson correlation coefficient (R), which was calculated separately for each OSM feature. OSM features with fewer than three training samples were excluded from the analysis (20% of all polygons, for 2015) because when $n = 2$ the R coefficient would always be –1 or +1, except in the improbable circumstance that both y-values perfectly matched.

4.4. Results

4.4.1. OSM Dataset Analysis

The identification and selection of OSM data contributed in 2012 and 2015 allowed us to create two different datasets for analysis. However, these datasets presented some logical inconsistencies that had to be resolved. There were 994 and 2,100 non-overlapping features in 2012 and 2015, respectively. After dissolving or removing the overlapping features following the procedure described in Chapter 4.4.3.1, the final datasets were comprised of 1,215 and 2,863 distinct features in 2012 and 2015, respectively.

The descriptive statistics for both years are presented in Table 4.4. It is notable that the total area covered increased significantly between 2012 and 2015 (+155%). In addition, between 2012 and 2015 the features' minimum area doubled, while their maximum area increased by 2.5 times. As expected, the standard deviation measure also confirmed high variations between the features' areas for each annual dataset, as most the natural/anthropological features fit power law (like) distributions.

Table 4.4. Descriptive statistics for 2012 and 2015 OpenStreetMap (OSM) datasets

	OSM	
	2012	2015
Total polygons	1215	2863
Minimum area of features (ha)	0.0003	0.0006
Maximum area of features (ha)	396.45	1000.75
Mean area of features (ha)	3.55	3.84
Standard Deviation area (ha)	17.27	32.97
Total area (ha)	4314.88	10986.02

In addition, the statistical analysis shows that the 98% of the features in 2012 OSM dataset and the 97% of the features in 2015 OSM dataset have less than 25 ha and the 76.13% of the features in 2012 OSM dataset as well as the 79.9% of the features in 2015 OSM dataset have less than 1 ha. In this way we are able to verify that the OSM used data are not imported from the reference datasets since CLC minimum area unit is 25 ha and COS minimum area unit is 1 ha. When we compared the geometry of the 2012 and 2015 OSM features with the CLC and COS datasets, less than 0.1% matching was obtained, proving that no bulk imports were made from these reference data.

4.4.2. OSM Completeness Accuracy

The completeness ratio of the coverage area for the selected study area has been increasing year after year (Table 4.5). However, in 2012, contributions covered less than 1% of the total district area, while in 2015 contributions covered about 1%. Following Table 4.6, Water was the largest LULC class in both annual OSM datasets (45% in 2012 and 46% in 2015). In 2012, agricultural areas represented about 40% of the total dataset, followed by artificial surfaces (16%), while in 2015 both classes were around 20%. In contrast, in 2012 forests were not significantly represented, but in 2015 they accounted for 14% of the total dataset. It is also interesting to note that in 2012 there were only four LULC classes represented (artificial surfaces, agricultural areas, forests, and water), while in 2015 there were five, due to the additional presence of wetlands. However, wetlands accounted for less than 0.001% of total coverage. In addition, the relationship between the total features' contributions and the corresponding area for each class demonstrated that both agricultural areas and water areas had consistently huge polygons drawn by contributors.

Table 4.5. OSM datasets completeness

	Completeness (%)						
	2011	2012	2013	2014	2015	2016	2017
Total	0.07	0.43	0.81	0.90	1.07	2.8	5.9

Table 4.6. Descriptive statistics per class of 2012 and 2015 OSM datasets

LULC Class	Total features		Area (ha)		OSM Class Coverage (%)		Completeness (%)	
	2012	2015	2012	2015	2012	2015	2012	2015
Artificial Surfaces	800	1,612	672.42	2,187.47	15.58	19.91	0.07	0.21
Agricultural areas	152	392	1,705.71	2,145.09	39.53	19.53	0.17	0.21
Forest	5	95	5.07	1,543.57	0.12	14.05	0	0.15
Wetlands	0	19	0	17.84	0	0.16	0	0
Water	258	745	1,931.68	5,092.05	44.77	46.35	0.19	0.50
Total	1,215	2,863	4,314.88	10,986.02	100	100	0.43	1.07

4.4.3. OSM Thematic Accuracy

A confusion matrix was computed for each annual dataset (Tables 4.7 and 4.8) in order to evaluate its thematic accuracy. The results showed that thematic accuracy was 77.3% in 2012 and 91.9% in 2015, indicating a very high agreement (i.e., accuracy) between the 2015 OSM dataset and the reference dataset (COS). However, there was a wide variation among the accuracy values for each class. In particular, agricultural areas were highly accurate, with a 99.8% user accuracy rate in 2012 and a 94.5% user accuracy rate in 2015, suggesting that the areas classified by contributors as agricultural areas closely matched those in the reference datasets (CLC and COS, respectively). While the producer's accuracy rates were not quite as high (71.9% in 2012 and

79.1% in 2015), they were still high enough to suggest that this class was correctly shown on the OSM. Artificial surfaces also had consistently high accuracy values; in 2012, the user and producer accuracy rates were 70.9% and 75.5%, respectively, and in 2015 they increased to 80.9% and 95.3%, respectively.

Other classes showed a strong increase in user accuracy between 2012 and 2015. User accuracy for forests jumped from 0% in 2012 to more than 97% in 2015, and user accuracy for water increased from about 60% in 2012 to over 94% in 2015. In contrast, producer accuracy for water was very high in both years (about 99.7%). Producer accuracy for forests improved from 0% in 2012 to over 88% in 2015. However, wetlands had null values for both user and producer accuracy in both years (0%). Overall, following the standards set by (J. R. Landis & Koch, 1977), the Kappa index indicated substantial agreement between the OSM and reference maps in 2012 (0.65) and an almost perfect agreement in 2015 (0.88).

Table 4.7. Thematic accuracy for 2012 (in ha)

		Reference data (CLC 2012)						
	LULC Class	Artificial Surfaces	Agricultural areas	Forest	Wetlands	Water	Total	User Accuracy
2012 OSM dataset	Artificial Surfaces	477	151	43	0	1	672	70.98
	Agricultural areas	0	1,702	4	0	0	1,706	99.77
	Forest	0	3	0	0	2	5	0
	Wetlands	0	0	0	0	0	0	0
	Water	155	512	108	0	1,157	1,932	59.88
	Total	632	2,368	155	0	1,160	4,315	
	Producer Accuracy	75.47	71.87	0	0	99.74		77.31
	Kappa índex							0.645

Table 4.8. Thematic accuracy for 2015 (in ha)

		Reference data (COS 2015)						
	LULC Class	Artificial Surfaces	Agricultural areas	Forest	Wetlands	Water	Total	User Accuracy
2015 OSM dataset	Artificial Surfaces	1,771	368	48	0	1	2,188	80.94
	Agricultural areas	64	2,028	52	0	2	2,146	94.50
	Forest	2	29	1,504	0	8	1,543	97.47
	Wetlands	0	10	5	0	3	18	0
	Water	21	130	96	50	4,794	5,091	94.16
	Total	1,858	2,565	1,705	50	4,808	10,986	
	Producer Accuracy	95.31	79.06	88.21	0	99.71		91.91
	Kappa índex							0.884

4.4.4. Assessing the Fitness of OSM Data

Using the method described in Chapter 4.3.3.3., about 24,175 random points were generated from the 2015 OSM dataset features. We evaluated their thematic accuracy and found a very high overall accuracy (88.5%). NBTA was used to assess the extent to which an OSM feature's proximity to a boundary influenced the attribute (tag) accuracy of the training sample. We computed the Euclidean distance from each training sample to the nearest segment of the OSM feature's boundary and plotted the distance values (in ascending order) and the accumulated thematic accuracy (Figure 4.4). The cross-reading between the resulting accuracy values and training sample distances formed the basis for our understanding of whether a feature's proximity to a boundary had any influence on the training sample's accuracy. As shown in Figure 4.4, it appeared that training samples near the feature boundary had a lower accuracy, with a trend toward increased accuracy as the feature's distance from the boundary increased.

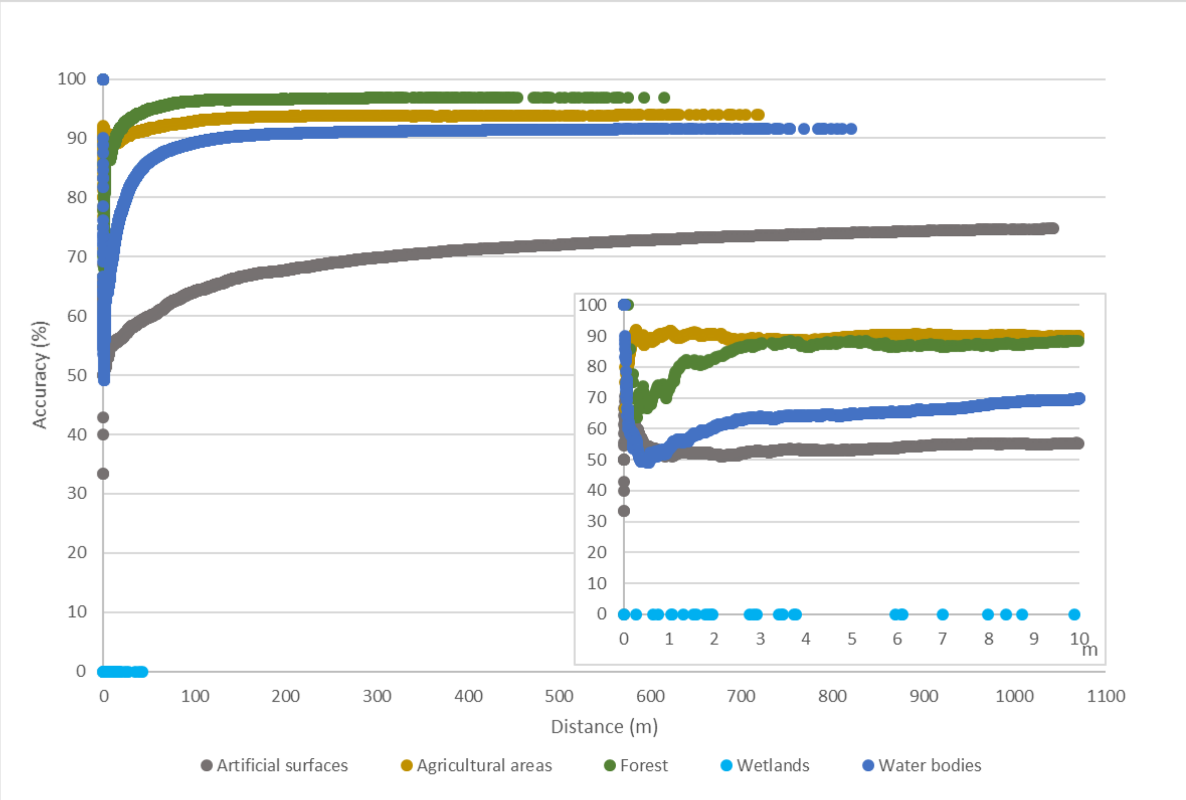


Figure 4.4. Overall relationship between training samples accuracy and feature boundary proximity

However, there was significant variability between the different LULC classes. Wetlands always presented 0% accuracy, as expected given the null thematic accuracy in 2015 described in Chapter 4.4.2., mainly due to the low total coverage (less than 0.001%). Forests, agricultural areas, and water had the highest overall accuracy. For the first two, this was expected since these areas are typically large and homogeneous. For both these classes, low accuracy values predominantly appeared close to the feature boundaries, but at ≥ 1 m from the boundary, high accuracy values were recorded linearly and were always above 85%. Water features behaved somewhat similarly, but accuracy only consistently improved to 70% at ≥ 10 m from the boundary, increasing linearly thereafter but reaching 85% accuracy only at ≥ 45 m from the feature boundary. Overall, artificial surfaces behaved somewhat differently from the other major classes. Training samples for artificial surfaces reached around 60% accuracy within 0.5 m of the feature boundary, but this trend quickly reversed as distance increased, with accuracy decreasing to 50–55% at 0.5–40 m from the boundary. At > 40 m from the boundary, accuracy values reached 60% and increased linearly, reaching 70% accuracy at 250 m and continuing to increase linearly with distance, eventually reaching a maximum of 75% accuracy.

Figure 4.5 illustrates how the maximum distance from a training sample to the nearest feature boundary varies moderately between each LULC class. On average (excluding wetlands), errors were very close to feature boundaries, but a more nuanced reading of the accuracy values and training sample distances can be seen. For features in the water class, at ≥ 406 m from the boundary (49% of the maximum distance) there were no misclassified (incorrect) training samples. Agricultural areas had similar results, with 100% accuracy at ≥ 427 m from the boundary (59% of the maximum distance). Both forests and artificial surfaces had high accuracy values beginning at > 32% of the maximum distance (corresponding to ≥ 199 m for forests and ≥ 337 m for artificial surfaces).

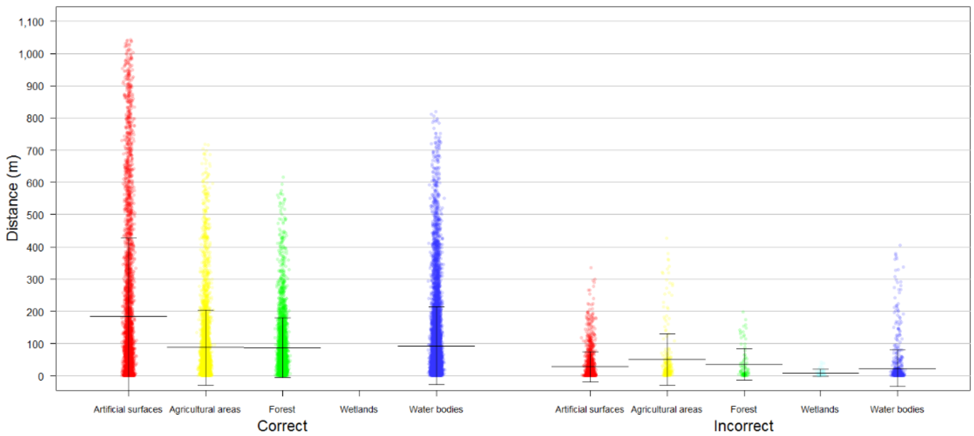


Figure 4.5. Correct and incorrect training samples of each LULC class according to their distance from feature boundary

The correlation between training sample accuracy and their proximity to feature boundary considering the influence of each feature area was tested with the R coefficient. High *R* values, indicated that accuracy was strongly - positively or negatively correlated - to the proximity of the boundary. In most cases, we found a positive correlation between high attribute accuracy and training samples' distance to feature boundary (Figure 4.6). In general, training samples were more likely to show improved accuracy as they moved away from the feature boundary. In addition, features with larger areas (> 10 ha) had a steadily positive correlation between training samples accuracy and distance to feature boundaries. On the other hand, this relationship is not clear for features with small areas (<5 ha), since correlation coefficient varies proportionally directly and inversely to feature boundary proximity.

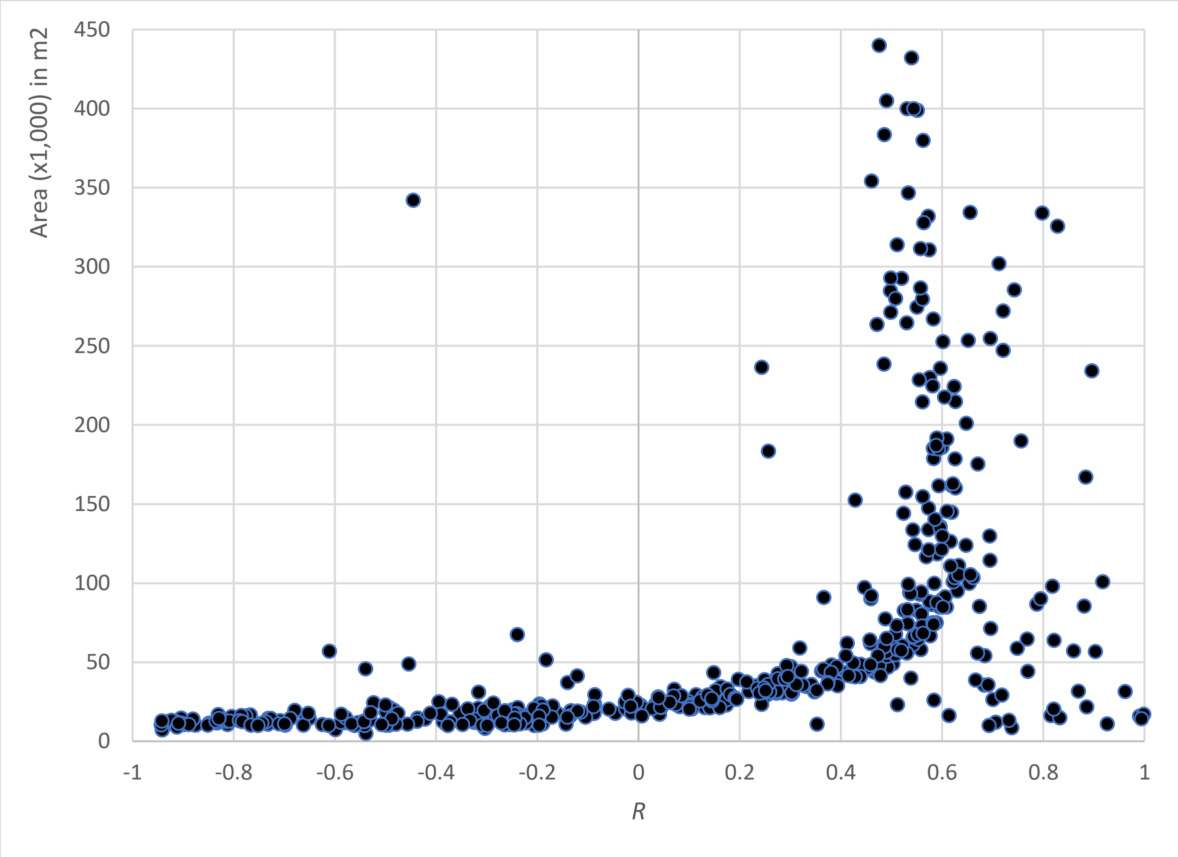


Figure 4.6. Pearson correlation Coefficient (R) between tag accuracy and features boundary proximity

4.5. Discussion

Over the last decade, OSM data have emerged as a supplementary source of data for LULC mapping, mainly due to improvements in the quality of OSM data (Neis et al., 2012). LULC map accuracy assessments are extremely important since they measure the quality of the LULC map and allow for improvements in analysis reliability (Arsanjani, Mooney, & Zipf, 2015; Arsanjani & Fonte, 2016; Arsanjani & Vaz, 2015; Estima & Painho, 2015; Fonte et al., 2017). However, the main focus of most LULC mapping applications that use OSM data has been on urban areas and at local scales, and some studies have emphasized the OSM data's lack of homogeneity, both in terms of its spatial coverage and in terms of its accuracy from region to region (Haklay, 2010; Hecht et al., 2013). The present study therefore proposed a methodology for creating different OSM datasets based on each OSM data contribution year, and used completeness and thematic accuracy assessments to evaluate the degree to which the OSM datasets agreed with the reference ones. In addition, we proposed NBTA as a criterion with which to evaluate the quality of OSM data as a sampling data source for multi-temporal LULC maps.

4.5.1. Completeness and Thematic Accuracy Assessment

Completeness accuracy and thematic accuracy are the two criteria that are most frequently mentioned to assess the quality of OSM data (Arsanjani, Mooney, & Zipf, 2015; Arsanjani & Fonte, 2016; Arsanjani & Vaz, 2015; Brovelli & Zamboni, 2018; Estima & Painho, 2013a, 2015; Fonte et al., 2017; Hecht et al., 2013; Rehrl et al., 2016). In particular, some authors have emphasized the importance of using completeness accuracy as a quality measurement since OSM data's territorial coverage can be extremely variable between locations (Arsanjani, Mooney, Zipf, et al., 2015; Arsanjani & Fonte, 2016; Neis et al., 2012), which can in turn limit the ability to draw any widely applicable conclusions (Hecht et al., 2013). Other studies have also cited a lack of OSM data contributions in rural areas (Helbich et al., 2012; Mooney et al., 2012) which we also confirmed in this study: at the time we downloaded the OSM data for the Beja district, the contribution area over all seven years (2011-2017) was less than 12% of the total Beja area. In spite of the increase year after year, the area covered in 2012 alone was less than 1% of the total Beja area, and in 2015 was still only about 1%. Thus, it is worth noting the increase in total coverage area between 2015-2016 (+1.7%) and 2016-2017 (+3.1%), which indicates an increase in volunteer participation in this region and suggests a potential for continued increases in subsequent years.

In our methodology, we have only included features which were created or modified during any given year and this has influenced the low value of data for each annual dataset. Nevertheless, we decided to follow this approach because OSM data can be updated daily, either by image interpretation, or by importing data (e.g., CLC, GPS devices). In the case of image interpretation, the satellite layer of Bing Maps is used as the background image in OSM edits. Thus, the contribution is influenced by the user's personal knowledge as well as by the background image used at the time of OSM data capture. Our purpose in doing was to attempt to reduce errors, as we could not know if features that were not updated by users were still valid, no longer present, or if it was merely that no user had elected to update that feature during that time period. Furthermore, the assessment of the provenance of OSM data for our study area reveals low probability of bulk imports since less than 0.1% of matching was obtained.

High agreement between OSM data and reference data have been found in a number of studies (Arsanjani et al., 2013; Estima & Painho, 2013a, 2015; Fonte & Martinho, 2017). In this study, we also have found high thematic accuracy values for both the 2012 and 2015 datasets (77.3% and 91.9%, respectively), signaling a significant improvement in the quality of the data between 2012 and 2015. Nevertheless, the literature has mainly attributed quality improvements to increases in contributors over time (Fonte et al., 2017; Zook et al., 2010) and we believe that differences between the two reference datasets we used (CLC for 2012 and COS for 2015) may also help to explain the substantial improvements in our thematic accuracy findings for 2012 and 2015. The different cartographic properties of the CLC and COS, such as their scale and spatial resolution (Meneses et al., 2018), may help to explain their different findings. The COS dataset has a minimum mapping unit (MMU) of 1 ha and a spatial resolution of 0.5 m, compared to the 25 ha and 20 m of the CLC; as such, the CLC dataset has greater polygon generalization than the COS. These comparisons provide a brief glimpse of some differences in reported results quality when LULC datasets with different cartographic properties are compared (García-Álvarez, 2018; Meneses et al., 2018), and could explain the different accuracy values for each LULC class here, as represented by the form and area of each feature of OSM data.

4.5.2. NBTA Accuracy Assessment

We introduced NBTA as a method to evaluate whether the quality of OSM data is suitable for it to be used as a sampling data source for LULC mapping. As expected, training samples closer to feature boundaries had higher levels of uncertainty. However, the degree of uncertainty varied significantly for each LULC class.

Beja is a region defined by strong processes of desertification and is dominated by large croplands with well-defined limits (Allen et al., 2018; Viana et al., 2019). These characteristics help to explain why the forest and agricultural classes had the highest accuracy values, as these classes are typically represented as large homogeneous areas that are easier to map. Nevertheless, there was a slight difference between the two classes, with agricultural areas having slightly lower accuracy values than forests—perhaps due to the fact that agricultural areas, while typically defined by large crop areas, in some cases also have small spaces with trees, ponds, and houses, potentially resulting in some confusion. Contributors may have preferred to draw large features to represent the crops, ignoring the existence of small areas with different LULC types. Water features, which behave similarly, had slightly larger wrongly-classified areas, likely due to the fact that water boundaries are perennially associated with changing weather conditions.

Artificial surfaces presented two different behaviors. For consolidated urban areas, such as the region’s main town (Beja) (Viana et al., 2019), the features behaved identically to the agricultural and forest areas. However, in more dispersed settlements (which were predominant in the Beja district), differences in interpretation of the MMU between the contributors (who were more likely to map everything in detail) and the technicians responsible for COS mapping (who were bound to the rules of cartography at a scale 1:25000) could explain the comparatively low accuracy. In addition, dispersed settlement areas demarcated by the contributors as artificial surfaces had relatively small dimensions, since contributors start considering houses far apart from each other's as isolated elements.

Finally, the wetlands class showed a complete disagreement between the two datasets duo to semantic incoherencies. The OSM feature descriptions of wetlands suggests that this class is mainly comprised of flood zones, mostly along water lines, whereas COS defined wetlands as areas of swamps and marshes. In sum, there were three classes where discrepancies between the OSM and reference datasets were essentially geometric (forest, agricultural, and water areas), one where they were semantic (wetlands), and one where they were both semantic and geometric (artificial surfaces). This finding is shown in more detail in Figure 4.7.

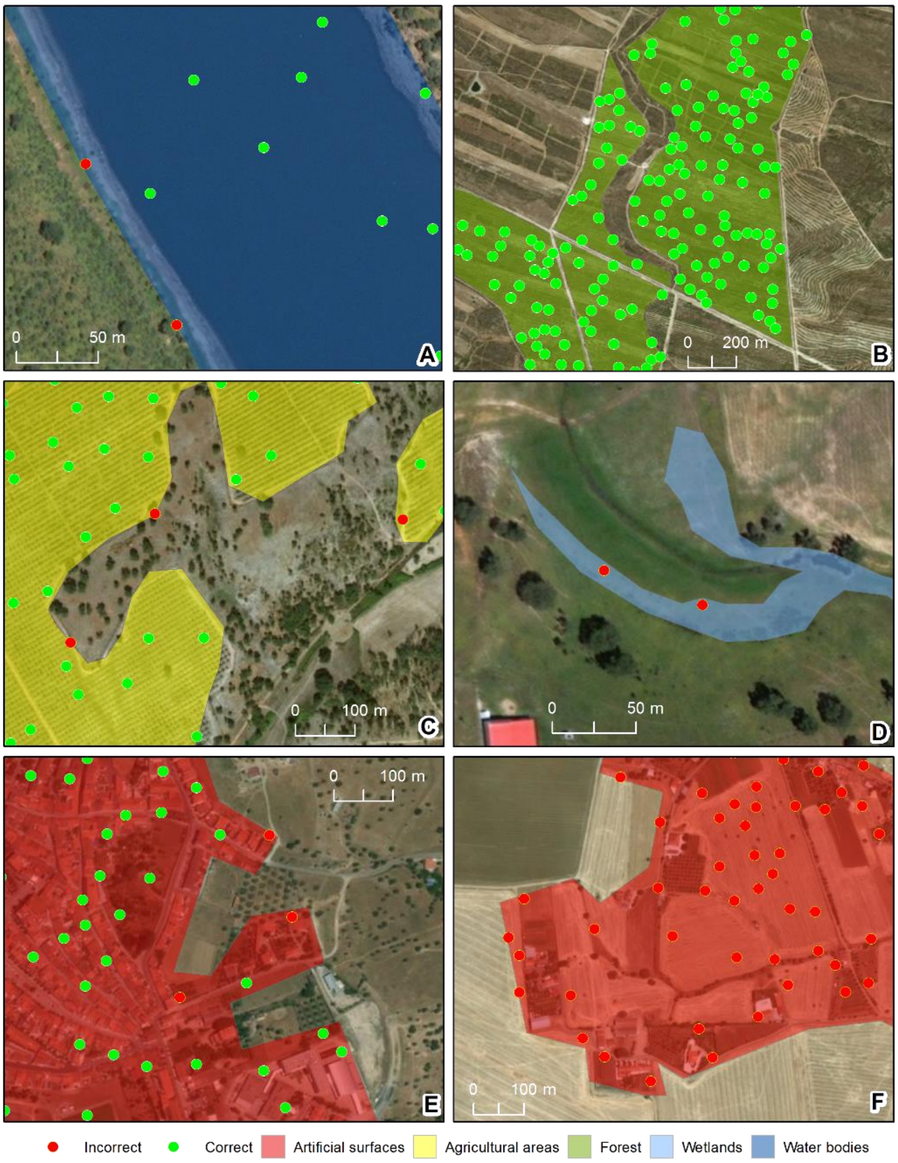


Figure 4.7. Example of correct and incorrect training samples for each LULC class

Some of the semantic incoherencies found in this study may be related to the used nomenclature correspondence. As some authors already have mentioned, the nomenclature harmonization between different datasets (Arsanjani, Mooney, Zipf, et al., 2015; Arsanjani & Vaz, 2015; Estima & Painho, 2013a; Haack et al., 2015) can be a difficult process and can influence the results (Fonte & Martinho, 2017).

Comparing the influence of each feature area on the NBTA yielded some interesting results. We found that the training samples' accuracy was proportional to their proximity to the polygon's boundary, but this proportionality was somewhat dependent on the area of the polygon. Features with areas > 10 ha had a steady positive correlation, presenting higher level of thematic accuracy as training samples moved away from the features' boundaries. Looking at the descriptive statistics for each LULC class in the 2015 OSM dataset, the classes represented by these larger areas were mainly agricultural areas and forests. In features of < 5 ha, this relationship between accuracy and the proximity to the polygon's boundary was not clear, since training samples demonstrated both positive and negative correlations to boundary proximity. These areas were mainly comprised of artificial surfaces with very disparate geometries, including both large and small areas and consolidated and dispersed settlements.

4.6. Conclusions

OSM data history was accessed to generate LULC datasets based on the contribution year (timestamp). Two different datasets comprising all contributions made in 2012 and in 2015 were created. Although downloading the data history was straightforward, the data exploration required some Structured Query Language (SQL) knowledge in order to obtain data elements, such as timestamps, that are stored on raw packages. This may prevent or simply restrain the use of OSM data history by common users. In addition, there are several logical inconsistencies in the OSM data that need to be analyzed and resolved, which are time-consuming.

Our research was conducted at the district level in a predominantly rural region that has undergone rapid LULC changes. In rural areas, where reference data are scarce, LULC data withhigh accuracy, even in small quantities, will always be of significant value. Thus, OSM platforms should be seen as a valid source of data, both in the production and updating of LULC maps and as a sample source for training purposes in supervised multi-temporal remote sensing classifications. If these data are used as auxiliary data to classify satellite images, the use of timestamps to create, for example, multi-temporal year-based or month-based datasets could improve the quality of future classifications. Additional research should investigate whether the use of OSM data history and the division of features by their year of contribution influence the accuracy of OSM data, and whether they can be used as ground-truth auxiliary data. Furthermore, the present study demonstrated that OSM LULC classes (artificial surfaces, agricultural areas, forests, and water) were as accurate as the official reference dataset to which they were compared (COS 1:25 0000 map), and thus have great potential as auxiliary data for use in mapping applications. More analyses should be carried out in other regions. Ultimately, OSM data are freely available and their use is not highly time-consuming. The approach used here could therefore also be usefully applied at a larger scale (e.g., country level).

References

Allen, H., Simonson, W., Parham, E., Santos, E. de B. e, & Hotham, P. (2018). Satellite remote sensing of land cover change in a mixed agro-silvo-pastoral landscape in the Alentejo, Portugal. International Journal of Remote Sensing, 00(00), 1–21. https://doi.org/10.1080/01431161.2018.1440095

Almendros-Jiménez, J., & Becerra-Terón, A. (2018). Analyzing the Tagging Quality of the Spanish OpenStreetMap. ISPRS International Journal of Geo-Information, 7(8), 323. https://doi.org/10.3390/ijgi7080323

Antoniou, V, & Skopeliti, A. (2015). Measures and indicators of VGI quality: an overview. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 345–351. https://doi.org/10.5194/isprsannals-II-3-W5-345-2015

Antoniou, Vyron, Touya, G., & Raimond, A.-M. (2016). Quality analysis of the Parisian OSM toponyms evolution. European Handbook of Crowdsourced Geographic Information, 97–112.

Arsanjani, J. J., & Fonte, C. C. (2016). On the Contribution of Volunteered Geographic Information to Land Monitoring Efforts. European Handbook of Crowdsourced Geographic Information, August, 269–284. https://doi.org/http://dx.doi.org/10.5334/bax.t. License: CC-BY 4.0

Arsanjani, J. J., Helbich, M., Bakillah, M., Hagenauer, J., & Zipf, A. (2013). Toward mapping land-use patterns from volunteered geographic information. International Journal of Geographical Information Science, 27(12), 2264–2278. https://doi.org/10.1080/13658816.2013.800871

Arsanjani, J. J., Mooney, P., & Zipf, A. (2015). OpenStreetMap in GIScience. https://doi.org/10.1007/978-3-319-14280-7

Arsanjani, J. J., Mooney, P., Zipf, A., & Schauss, A. (2015). Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets (pp. 37–58). Springer, Cham. https://doi.org/10.1007/978-3-319-14280-7_3

Arsanjani, J. J., & Vaz, E. (2015). An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises. International Journal of Applied Earth Observation and Geoinformation, 35, 329–337. https://doi.org/10.1016/J.JAG.2014.09.009

Barron, C., Neis, P., & Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. Transactions in GIS, 18(6), 877–895. https://doi.org/10.1111/tgis.12073

Brovelli, M., & Zamboni, G. (2018). A New Method for the Assessment of Spatial Accuracy and Completeness of OpenStreetMap Building Footprints. ISPRS International Journal of Geo-Information, 7(8), 289. https://doi.org/10.3390/ijgi7080289

Büttner, G., & Feranec, J. (2002). The CORINE Land Cover upddate 2000. Techinical guidelines. EEA Technical Report, 89.

Comber, A., Brunsdon, V., See, L., & Fritz, S. (2013). Evaluating Global Land Cover Datasets: Comparing VGI on Cropland with Formal Data. GI_Forum 2013. Creating the GISociety, 2013, 91–95. https://doi.org/10.1553/giscience2013s91.This

Dorn, H., Törnros, T., Zipf, A., Dorn, H., Törnros, T., & Zipf, A. (2015). Quality Evaluation of VGI Using Authoritative Data—A Comparison with Land Use Data in Southern Germany. ISPRS International Journal of Geo-Information, 4(3), 1657–1671. https://doi.org/10.3390/ijgi4031657

Estima, J., & Painho, M. (2013a). Exploratory analysis of OpenStreetMap for land use classification. Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information - GEOCROWD '13, 39–46. https://doi.org/10.1145/2534732.2534734

Estima, J., & Painho, M. (2013b). Flickr geotagged and publicly available photos: Preliminary study of its adequacy for helping quality control of Corine Land Cover. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7974 LNCS(PART 4), 205–220. https://doi.org/10.1007/978-3-642-39649-6-15

Estima, J., & Painho, M. (2014). Photo Based Volunteered Geographic Information Initiatives. International Journal of Agricultural and Environmental Information Systems, 5(3), 73–89. https://doi.org/10.4018/ijaeis.2014070105

Estima, J., & Painho, M. (2015). Investigating the Potential of OpenStreetMap for Land Use/Land Cover Production: A Case Study for Continental Portugal. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience (pp. 273–293). Springer, Cham. https://doi.org/10.1007/978-3-319-14280-7

Fonte, C. C., Bastin, L., Foody, G., Kellenberger, T., Kerle, N., Mooney, P., Olteanu-Raimond, A.-M., & See, L. (2015). Vgi Quality Control. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W5, 317–324. https://doi.org/10.5194/isprsannals-II-3-W5-317-2015

Fonte, Cidália C., & Martinho, N. (2017). Assessing the applicability of OpenStreetMap data to assist the validation of land use/land cover maps. International Journal of Geographical Information Science, 31(12), 2382–2400. https://doi.org/10.1080/13658816.2017.1358814

Fonte, Cidália Costa, Patriarca, J. A., Minghini, M., Antoniou, V., See, L., & Brovelli, M. A. (2017). Using OpenStreetMap to Create Land Use and Land Cover Maps: Development of an Application. In C. E. C. Campelo, M. Bertolotto, & P. Corcoran (Eds.), Volunteered Geographic Information and the Future of Geospatial Data: Vol. i (pp. 113–137). IGI Global. https://doi.org/10.4018/978-1-5225-2446-5

Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., & Boyd, D. S. (2013). Assessing the Accuracy of Volunteered Geographic Information arising from Multiple Contributors to an Internet Based Collaborative Project. Transactions in GIS, 17(6), 847–860. https://doi.org/10.1111/tgis.12033

Fritz, S., Bartholomé, E., Belward, A., Hartley, A., Stibig, H.-J., Eva, H., Mayaux, P., Bartalev, S., Latifovic, R., Kolmert, S., Roy, P. S., Agrawal, S., Bingfang, W., Wenting, X., Ledwith, M., Pekel, J.-F., Giri, C., Mücher, S., De Badts, E., … Defourny, P. (2003). Harmonisation, mosaicing and production of the Global Land Cover 2000 database (Beta Version).

García-Álvarez, D. (2018). The Influence of Scale in LULC Modeling. A Comparison Between Two Different LULC Maps (SIOSE and CORINE). In M. T. Camacho Olmedo, M. Paegelow, J. F. Mas, & F. Escobar (Eds.), Geomatic Simulations and Scenarios for Modelling LUCC. A Review and Comparison of Modelling Techniques (pp. 187–213). Springer, Cham. https://doi.org/10.1007/978-3-319-60801-3_10

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. Spatial Statistics, 1, 110–120. https://doi.org/https://doi.org/10.1016/j.spasta.2012.03.002

Guptill, S. C., Morrison, J. L., & International Cartographic Association. (1995). Elements of spatial data quality. Elsevier Science.

Haack, B., Mahabir, R., & Kerkering, J. (2015). Remote sensing-derived national land cover land use maps: a comparison for Malawi. Geocarto International, 30(3), 270–292. https://doi.org/10.1080/10106049.2014.952355

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. Environment and Planning B: Planning and Design, 37(4), 682–703. https://doi.org/10.1068/b35097

Hecht, R., Kunze, C., Hahmann, S., Hecht, R., Kunze, C., & Hahmann, S. (2013). Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. ISPRS International Journal of Geo-Information, 2(4), 1066–1091. https://doi.org/10.3390/ijgi2041066

Helbich, M., Amelunxen, C., Neis, P., & Zipf, A. (2012). Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata. Proceedings of GI_Forum, January 2012, 24–33. https://doi.org/10.1080/13658816

Herold, M., Mayaux, P., Woodcock, C. E., Baccini, A., & Schmullius, C. (2008). Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. Remote Sensing of Environment, 112(5), 2538–2556. https://doi.org/10.1016/J.RSE.2007.11.013

INE. (2012). Censos 2011 Resultados Definitivos – Região Alentejo. Instituto Nacional de Estatística.

Johnson, B. A., & Iizuka, K. (2016). Integrating OpenStreetMap crowdsourced data and Landsat time-series imagery for rapid land use/land cover (LULC) mapping: Case study of the Laguna de Bay area of the Philippines. Applied Geography, 67, 140–149. https://doi.org/10.1016/j.apgeog.2015.12.006

Jonietz, D., & Zipf, A. (2016). Defining Fitness-for-Use for Crowdsourced Points of Interest (POI). ISPRS International Journal of Geo-Information, 5(9), 149. https://doi.org/10.3390/ijgi5090149

Keßler, C., Trame, J., & Kauppinen, T. (n.d.). Tracking Editing Processes in Volunteered Geographic Information: The Case of OpenStreetMap.

Landis, J. R., & Koch, G. G. (1977). Measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174.

Meneses, B., Reis, E., Vale, M. J., & Reis, R. (2018). Modelling the Land Use and Land cover changes in Portugal: A multi-scale and multi-temporal approach. Finisterra, 53(107). https://doi.org/10.18055/Finis12258

Meneses, B., Reis, E., Reis, R., Vale, M., Meneses, B. M., Reis, E., Reis, R., & Vale, M. J. (2018). The Effects of Land Use and Land Cover Geoinformation Raster Generalization in the Analysis of LUCC in Portugal. ISPRS International Journal of Geo-Information, 7(10), 390. https://doi.org/10.3390/ijgi7100390

Mooney, P., & Corcoran, P. (2012). Using OSM for LBS–An analysis of changes to attributes of spatial objects. In G. Ortag & Gartner F. (Eds.), Advances in location-based services, Lecturenotes in geoinformation and cartography (pp. 165–179). Springer-Verlag.

Mooney, Peter, & Corcoran, P. (2014). Analysis of interaction and co-editing patterns amongst openstreetmap contributors. Transactions in GIS, 18(5), 633–659. https://doi.org/10.1111/tgis.12051

Mooney, Peter, Corcoran, P., Mooney, P., & Corcoran, P. (2012). Characteristics of Heavily Edited Objects in OpenStreetMap. Future Internet, 4(1), 285–305. https://doi.org/10.3390/fi4010285

Nasiri, A., Ali Abbaspour, R., Chehreghan, A., & Jokar Arsanjani, J. (2018). Improving the Quality of Citizen Contributed Geodata through Their Historical Contributions: The Case of the Road Network in OpenStreetMap. ISPRS International Journal of Geo-Information, 7(7), 253. https://doi.org/10.3390/ijgi7070253

Neis, P., Zielstra, D., & Zipf, A. (2012). The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. Future Internet, 4(4), 1–21. https://doi.org/10.3390/fi4010001

Rehrl, K., Gröchenig, S., Rehrl, K., & Gröchenig, S. (2016). A Framework for Data-Centric Analysis of Mapping Activity in the Context of Volunteered Geographic Information. ISPRS International Journal of Geo-Information, 5(3), 37. https://doi.org/10.3390/ijgi5030037

See, L., Comber, A., Salk, C., Fritz, S., van der Velde, M., Perger, C., Schill, C., McCallum, I., Kraxner, F., & Obersteiner, M. (2013). Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts. PLoS ONE, 8(7), 1–11. https://doi.org/10.1371/journal.pone.0069958

See, L., Estima, J., Pődör, A., Jokar, J., Laso-Bayas, J., & Vatseva, R. (2016). Sources of VGI for Mapping. In G. Foody, L. See, S. Fritz, P. Mooney, A.-M. Olteanu-Raimond, C. C. Fonte, & V. Antoniou (Eds.), Mapping and the Citizen Sensor (pp. 13–35). Ubiquity Press. https://doi.org/https://doi.org/10.5334/bbf.a

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H.-Y., Milčinski, G., Nikšič, M., Painho, M., Pődör, A., Olteanu-Raimond, A.-M., & Rutzinger, M. (2016). Crowdsourcing, Citizen Science or Volunteered Geographic

Information? The Current State of Crowdsourced Geographic Information. ISPRS International Journal of Geo-Information, 5(5), 55. https://doi.org/10.3390/ijgi5050055

Sehra, S., Singh, J., Rai, H., Sehra, S. S., Singh, J., & Rai, H. S. (2017). Assessing OpenStreetMap Data Using Intrinsic Quality Indicators: An Extension to the QGIS Processing Toolbox. Future Internet, 9(2), 15. https://doi.org/10.3390/fi9020015

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (Muki). (2017). A review of volunteered geographic information quality assessment methods. International Journal of Geographical Information Science, 31(1), 139–167. https://doi.org/10.1080/13658816.2016.1189556

Sui, D., Goodchild, M., & Elwood, S. (2013). Volunteered Geographic Information, the Exaflood, and the Growing Digital Divide BT - Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice (D. Sui, S. Elwood, & M. Goodchild (eds.); pp. 1–12). Springer Netherlands. https://doi.org/10.1007/978-94-007-4587-2_1

Tracewski, L., Bastin, L., & Fonte, C. C. (2017). Repurposing a deep learning network to filter and classify volunteered photographs for land cover and land use characterization. Geo-Spatial Information Science, 20(3), 252–268. https://doi.org/10.1080/10095020.2017.1373955

Viana, C. M., Oliveira, S., Oliveira, S. C., & Rocha, J. (2019). Land Use/Land Cover Change Detection and Urban Sprawl Analysis. In H. R. Pourghasemi & C. Gokceoglu (Eds.), Spatial Modeling in GIS and R for Earth and Environmental Sciences (1st ed.). Elsevier.

Viana, C. M., & Rocha, J. (2018). Spatiotemporal analysis and scenario simulation of agricultural land use land cover using GIS and a Markov chain model. In A. Mansourian, P. Pilesjö, L. Harrie, & R. von Lammeren (Eds.), Geospatial Technologies for All : short papers, posters and poster abstracts of the 21th AGILE Conference on Geographic Information Science.

Yagoub, M. M. (2017). Assessment of OpenStreetMap (OSM) Data: The Case of Abu Dhabi City, United Arab Emirates. Journal of Map & Geography Libraries, 13(3), 300–319. https://doi.org/10.1080/15420353.2017.1378150

Yang, D., Fu, C. S., Smith, A. C., & Yu, Q. (2017). Open land-use map: a regional land-use mapping strategy for incorporating OpenStreetMap with earth observations. Geo-Spatial Information Science, 20(3), 269–281. https://doi.org/10.1080/10095020.2017.1371385

Zielstra, D., & Zipf, A. (2010). A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. 13th AGILE International Conference on Geographic Information Science, 1–15.

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. World Medical & Health Policy, 2(2), 7–33. https://doi.org/10.2139/ssrn.2216649

Part IV. Dynamics of agricultural land use: modelling spatiotemporal changes and factors

Chapter 5. A framework to evaluate land change simulation software with an illustration of a Cellular Automata – Markov model

Cláudia M. Viana, Robert Gilmore Pontius Jr, Jorge Rocha

Article submitted at the Computers, environment and urban systems

The structure and formatting are slightly adapted.

Abstract

Numerous models exist to simulate land change to facilitate communication with an audience of stakeholders. Our manuscript poses a framework to help users decide whether to use any particular model by asking four questions: 1) Can the user understand the model? 2) Can the audience understand the model? 3) Can the user control the model? and 4) Does the model address the goals of the specific application? Our manuscript applies this framework to a popular Cellular Automata - Markov model. We compare the software's documentation to the model's behaviour in 120 runs. Results show that the Cellular Automata feature corrupts the Markov feature in counterintuitive ways that the model's documentation fails to describe. The model's counterintuitive behaviour is likely to cause users to misinterpret the validation metrics and to miscommunicate with audiences. We encourage users to apply our framework to any simulation model.

Keywords: Cellular automata; CA-Markov; Markov model; model evaluation; IDRISI software

5.1. Introduction

Over the last decades, models have been the predominant tool to support the analysis of future changes among land categories (Houet et al., 2010; Noszczyk, 2019). Numerous land change models have been developed to depict possible land-use changes and help to support landscape planning and environmental management (see e.g., Fuglsang et al., 2013; Murray-Rust et al., 2014; Roodposhti et al., 2019; Sankarrao et al., 2021; Verweij et al., 2018). Therefore, it can be difficult for a prospective user to decide which model(s) to use because many such models exist. Accordingly, our manuscript proposes a framework to guide users to decide whether to use any particular model to address the goals of a specific application. The framework consists of four questions: 1) Can the user understand the model? 2) Can the audience understand the model? 3) Can the user control the model? and 4) Does the model address the goals of the specific application? The answers to these questions must be Yes for a user to be able to control the model's parameters to create various scenarios of landscape change in a manner that will communicate insights that are helpful to accomplish the goals. A model must be well documented for the user to have any hope to answer Yes to question 1. A model must be sufficiently straightforward to answer Yes to question 2. A model must have parameters that allow the user to control the output to answer Yes to question 3. The answer to question 4 depends on the alignment between the goals and the model's algorithm, but if the answers to questions 1-3 are No, then the likely answer to question 4 is also No. We recommend prospective users apply our framework to evaluate the appropriateness of any model for the user's specific goals.

We illustrate the use of our framework to evaluate the Cellular Automata – Markov model (CA-Markov) in the Selva version of the IDRISI software (Eastman, 2012; Eastman, 2006). We select IDRISI's CA-Markov model for several reasons. CA-Markov is integrated into the GIS IDRISI, which makes CA-Markov easily accessible to IDRISI's 100,000 users worldwide. IDRISI's CA-Markov is better documented than several other models so we have a better chance at answering Yes to question 1 than we would have had with some other models. Cellular Automata (CA) and Markov are general mathematical concepts that exist in other models, so our analysis can offer insights that could be relevant to other models. The creators of IDRISI software introduced the CA-Markov module as experimental, which warrants its testing (Eastman & Toledano, 2018). The creators of IDRISI have considered removing CA-Markov from the software, but users wanted CA-Markov to remain in the software as users have frequently claimed that CA-Markov is helpful. CA-Markov in IDRISI's has been popular in the literature (see e.g., Aksoy and Kaptan, 2021; Aliani et al., 2019; Ghosh et al., 2017; Guan et al., 2011; Halmy et al., 2015; Hamad et al., 2018; Huang et al., 2020; Hyandye and Martz, 2017; Liping et al., 2018; Palmate et al., 2017; Radwan et al., 2019; Rimal et al., 2017; Sang et al., 2011; Santé et al., 2010; Wang and Murayama, 2017; Yirsaw et al., 2017).

Any model that simulates transitions among land use categories through time and space must perform two tasks. First, the model must specify the number of observations that transition from one category to another category during each time interval, which is a concept known as quantity. The Markov part plays the role to specify the quantity in CA-Markov. Second, the model must specify the spatial distribution of the transitions, which is a concept known as allocation. The Cellular Automata part and suitability maps influence the allocation in CA-Markov (Eastman, 2012; Mas et al., 2014). If a model's parameters allow the user to specify the quantity separately from the allocation, then the user can control two important components to create various scenarios of simulated change, thus the user might be able to answer Yes to question 3.

The behavior and output of the model derive from three factors: 1) the user's decisions concerning how to format the input data, 2) the user's selection of the model's parameters, and 3) the software's design (see e.g., Dahal and Chow, 2015; Liao et al., 2016; Lin et al., 2020; Samat, 2006; Varga et al., 2020; Wu et al., 2012). In our testing, we consider two ways to format the input data. Specifically, we examine two possibilities to define the categories and four possibilities to format the spatial resolution of the input maps. We consider two ways to set the model's parameters concerning the spatial filter of the CA. We examine three possibilities to select the filter's shape and five possibilities to select the filter's size. Thus, our number of combinations of possibilities is two times four times three times five, which generates 120 runs. The software dictates the model's design, which we do not control.

Some users are tempted to evaluate a model based on its predictive power, which can be an unfair criterion to evaluate a model's algorithm for several reasons. First, there are many ways to format the input data, and each way can influence the model output. Second, there are many ways to select the model's parameters and each combination of parameters can influence the output. More importantly, the purpose of many models is to extrapolate from the calibration time interval, not necessarily to predict accurately during the extrapolation time interval. If the patterns in the reference data are not stationary from the calibration interval to the extrapolation interval, then an extrapolation from the calibration time interval will not have predictive power because the patterns in the landscape are not consistent through time. Thus, low measures of validation might be due more to the landscape than to the model's algorithm. Nevertheless, validation is an important consideration for many users, while some users apply flawed criteria for validation or do not validate at all (Agarwal et al., 2002; Bennett et al., 2013; Bradley et al., 2016; Olmedo et al., 2015; van Vliet et al., 2016). Therefore, our manuscript illustrates an insightful method of validation for our 120 model runs, which is important if the user's goal is prediction.

We illustrate our framework by applying IDRISI's CA-Markov to a ten thousand square kilometers region in Beja District, Portugal, which is a mixed agro-silvo-pastoral environment with low urban density (Viana, Girão, et al., 2019; Viana & Rocha, 2020). Our purpose is to illustrate our framework to judge a model's appropriateness for a specific application. Our results expose how CA-Markov behaves for this application, which is similar to the behavior that Varga et al., (2019) observed in Hungry but is not necessarily how CA-Markov would behave in all applications. Nevertheless, our results give insights into IDRISI's CA-Markov algorithm that its documentation does not describe clearly.

5.2. Materials and Methods

5.2.1. Study area and data

The study area is the Beja District located in southeastern Portugal, with an area that covers 11% of Portugal’s mainland territories and is home to 153 thousand inhabitants in 2011 (INE, 2011). To the southeast, the Beja landscape is flat; to the north and west, the extensive plains are cut by tiny hills. In general, the landscape is characterised by a mixed agro-silvo-pastoral land-use system. Figure 5.1 shows the data that our manuscript uses.

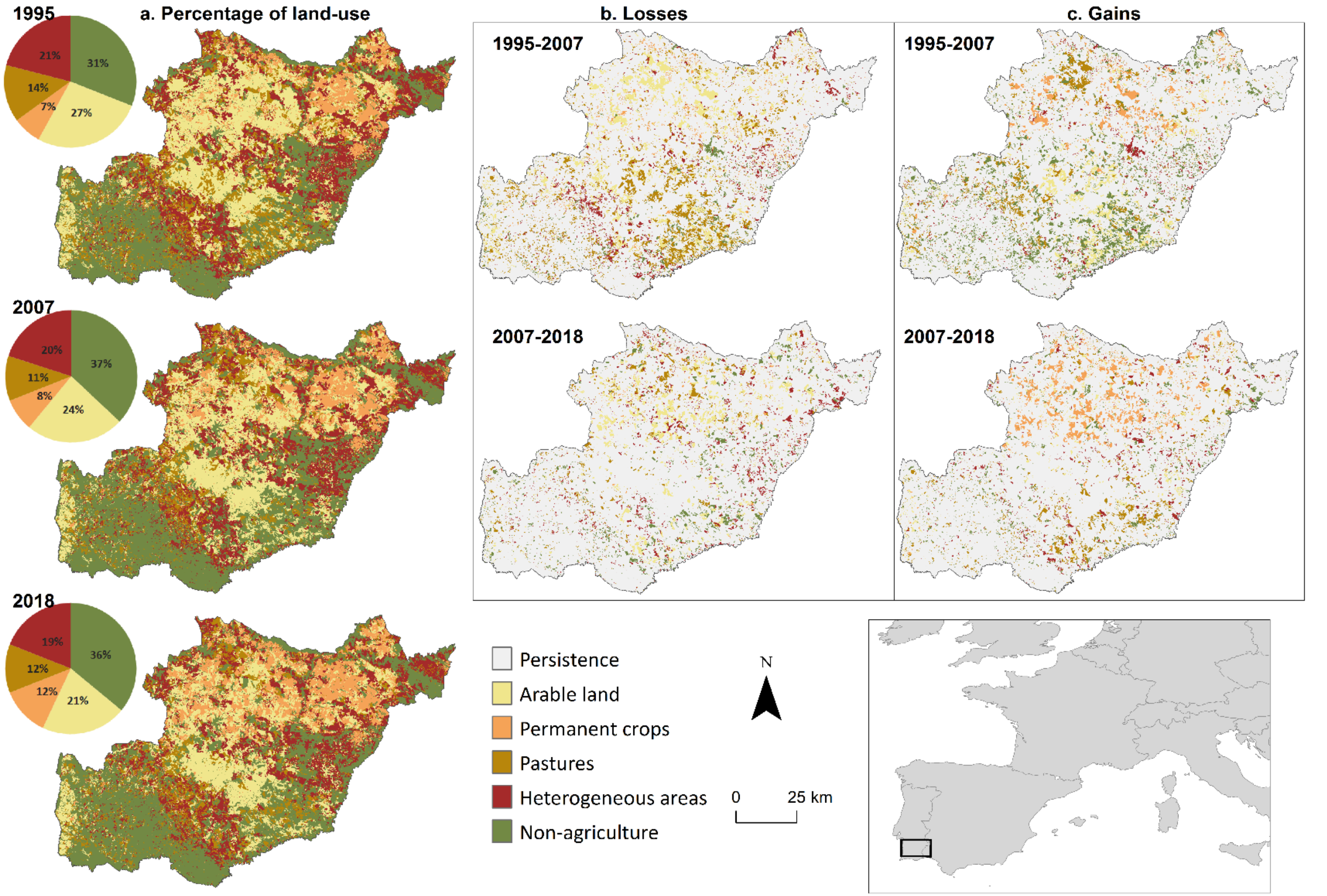


Figure 5.1. Land-use at 1995, 2007, 2018 and changes during 1995-2007 and 2007-2018 for the Beja district with a five-category classification

This study uses 1995, 2007, and 2018 land-use maps (COS) produced by the Portuguese General Directorate for Territorial Development (DGT). These maps are produced by photointerpretation, with vector polygon data, cartographic accuracy of 0.5 m, minimum mapping unit of 1 ha, and a hierarchical nomenclature system (DGT, 2018). COS 1995, 2007, and 2018 maps were converted into raster format following 10 × 10 m, 20 × 20 m, 50 × 50 m, and 100 × 100 m resolution, and reclassified into two and five land-use categories. The two categories are the agricultural category and the aggregation of all the non-agriculture categories, which are both first hierarchical levels of the COS nomenclature. The five categories are the non-agriculture categories and four types of agricultural categories based on the second hierarchical level: arable land, permanent crops, pastures, and heterogeneous areas. Table 5.1 specifies how our two and five land-use categories derive from the COS nomenclature levels.

Table 5.1. Land-use category description and percentage coverage

Category in model	Category in COS (level)	Description
Non-agriculture	Artificial surfaces, forest and semi-natural areas, wetlands, and water bodies (1)	Urban fabric; artificial non-agricultural vegetated areas; Forests; open spaces with little or no vegetation; inland wetlands.
Agricultural areas	Agricultural areas (1)	Areas principally occupied by agriculture, interspersed with significant natural or semi-natural areas.
Arable land	Arable land (2)	Lands that are rain-fed or irrigated under a rotation system used for annually harvested plants and fallow lands.
Permanent crops	Permanent crops (2)	Lands not under a rotation system. Includes fruit orchards, olive groves, shrub orchards such as vineyards.
Pastures	Pastures (2)	Lands that are permanently used (at least 5 years) for fodder production.
Heterogeneous areas	Heterogeneous areas (2)	Landscapes in which permanent crops on the same parcel, meadows, and/or pastures are intimately mixed with natural vegetation or natural areas.

5.2.2. IDRISI's CA-Markov model

We used the CA-Markov model in the Selva version of the IDRISI software (Eastman, 2012; Eastman, 2006). The CA-Markov model has three stages: 1) calculation of the transition area matrix, 2) development of the suitability images, and 3) allocation of change (Eastman & Toledano, 2018; Eastman, 2012; Eastman, 2006). Figure 5.2 presents an overview of the modelling process framework. Stage one consists of calculating the quantity of extrapolated change. Thus, the Markov transition area matrix is computed based on land-use map in an initial time (1995), land-use map in a subsequent time (2007), and extrapolation year (2018). In stage two, the user creates the suitability maps. Stage three is the allocation of the extrapolated change from 2007 to 2018. Our experimental design using the CA-Markov model repeats the simulation for the two thematic scales, four spatial resolutions, three neighbourhood shapes, and five neighbourhood sizes. This generates 120 combinations; thus we ran CA-Markov 120 times. We used validation to compare each run of simulated change to the reference change from 2007 to 2018. The following subsections describe each stage and the modelling procedures.

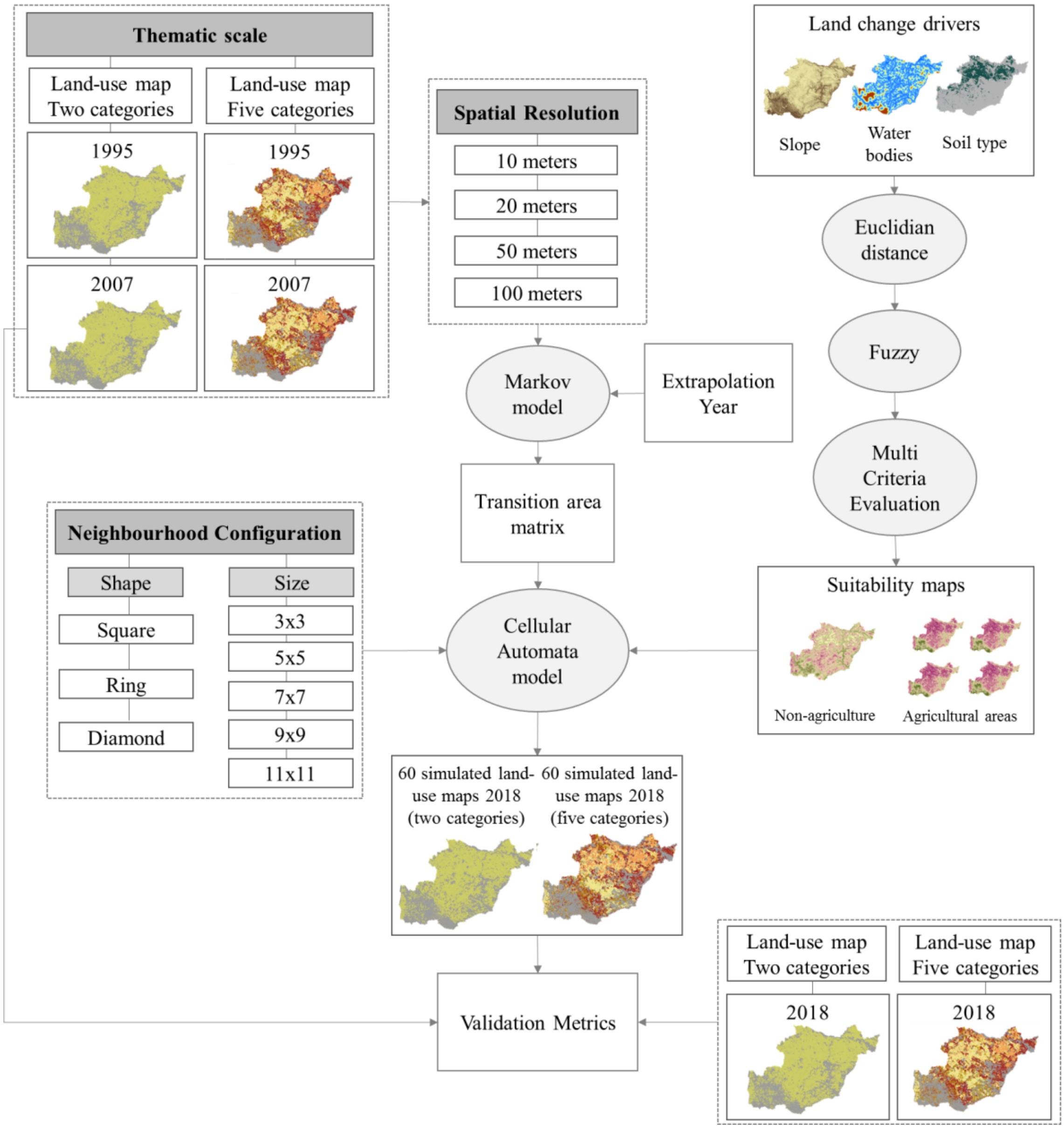


Figure 5.2. Modelling process framework

5.2.2.1. Calculation of the Transition Area Matrix

The first stage computes the transition area matrix, which extrapolates land change in terms of quantity. In this stage, we used the Markov module. The premise of the Markov model is that the land-use at a future time point is an extrapolation of the change from the initial time point to the end time point of the calibration time interval (Muller & Middleton, 1994). The calibration time interval combined with the extrapolation year generates matrix **A**.

$$\mathbf{A} = \begin{bmatrix} a_{ii} & \cdots & a_{iJ} \\ \vdots & \ddots & \vdots \\ a_{Ji} & \cdots & a_{JJ} \end{bmatrix} \tag{5.1}$$

Each entry a_{ij} in matrix **A** is the area that is category i at the start of the extrapolation and category j at the end of the extrapolation. The number of categories is J. The calibration interval is the 12 years between 1995 and 2007. The extrapolation interval is the 11 years between 2007 and 2018. We accounted for the difference between the duration of the calibration and extrapolation intervals by specifying in IDRISI's Markov module that the number of extrapolated years is 11 (Eastman, 2012; Eastman, 2006).

5.2.2.2. Development of the Suitability Maps

The user must construct the suitability maps in the second stage. Each suitability map portrays the appropriateness of each cell for a particular land-use category. Slope, water bodies, and soil type determine our suitability maps. The slope data were collected from Instituto Geográfico Português (IGEO), water bodies were collected from the Agência Portuguesa do Ambiente (APA), and soil type data (Cartas de Solos e de Capacidade de Uso do Solo) were collected from the Direção-Geral de Agricultura e Desenvolvimento Rural (DGADR). A suitability map was generated for the non-agriculture category and the agricultural land category. Each agricultural land category uses the same suitability map. Specifically, flatter slopes have higher suitability for both categories. Closer distance to water bodies indicates high suitability for the agricultural category, while longer distance indicates high suitability for the non-agricultural category. Soil type data were reclassified into the 0–1 range where 1 indicates suitable soil types for agriculture, while 0 indicates suitable soil types for the non-agricultural category. Finally, the weighted linear combination multi-criteria evaluation method was used to set the weights for slope as 0.4, distance to water bodies as 0.4, and soil type as 0.2. Data availability dictated variable selection while expert opinion determined the weights.

5.2.2.3. Change Allocation

In the third stage, a CA process allocates the extrapolated quantities from the transition area matrix **A**. Conceptually, the CA-Markov model consists of cells, states, time, neighbourhoods, and transition rules (Torrens, 2000). The following equation expresses how the simulated change derives from a cell's category, the number of cells that change from one category to another, the suitability maps, and the cell's neighbours (White & Engelen, 2000; Wolfram, 1984).

$$^{t+1}C_m = f(\ ^{t}C_m, a_{ij}, S_{mj},\ ^{t}N_m) \tag{5.2}$$

where $^{t+1}C_{mn}$ denotes the category at time t+1 of cell m; $^{t}C_m$ denotes the category at time t of cell m; a_{ij} is the Markov matrix entry that gives the extrapolated size of change from category i to category j; S_{mj} is the suitability index of cell m for category j, and $^{t}N_m$ denotes the condition at time t of the neighbourhood of cell m.

IDRISI's CA-Markov module requires the basis land-use map 2007 (reference time 2) and the Markov transition area matrix for the extrapolation. The procedure allows the user to define a spatial filter. The spatial filter defines a cell's neighbourhood in terms of shape and size (Pan, Roth, Yu, & Doluschitz, 2010; Verburg, de Nijs, van Eck, Visser, & de Jong, 2004). We examined three filter shapes as diamond, ring, or square; and five filter's size as 3 × 3, 5 × 5, 7 × 7, 9 × 9, and 11 × 11 (Figure 5.3). The diamond shape includes pixels that are within the same Manhattan distance from the central cell. The ring shape is the outermost cells at a distance identical to the square type, where the neighbourhood size is equal to the number of cells on one side (Pan et al., 2010; White & Engelen, 2000). The square shape is a square neighbourhood that fills the ring (Pan et al., 2010; White & Engelen, 2000).

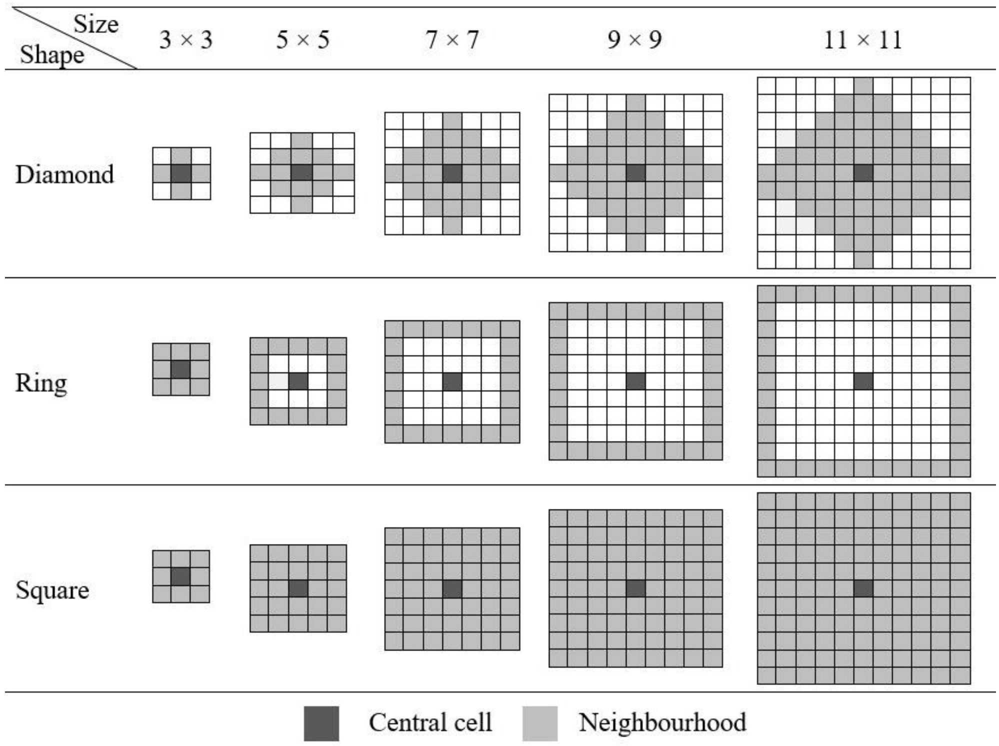


Figure 5.3. Neighbourhood filter's shape and size configurations

5.2.3. Model validation

We used the three-way cross-tabulation technique to compare the simulated change to the reference change. We compared the reference map for 2007, the reference map for 2018, and each of the 120 simulated maps for 2018 (Pontius Jr et al., 2008, 2011). This comparison generates five components: Misses, Hits, Wrong Hits, False Alarms, and Correct Rejections (Pontius Jr et al., 2008, 2011). Misses are reference change areas simulated as persistence. Hits are reference change simulated as changes to the correct land-use category. Wrong Hits are reference change simulated as change to the wrong land-use category, which requires more than two categories. False Alarms are reference persistence simulated as change. Correct Rejections are reference persistence areas simulated as persistence. These components generate the figure of merit (FOM), which provides a measure of the accuracy of change, where 0% means no intersection between simulated and reference change areas, while 100% means perfect intersection between simulated and reference change areas (Perica & Foufoula-Georgiou, 1996; Pontius Jr et al., 2008, 2011; Varga et al., 2019).

$$FOM = \frac{Hits}{Misses + Hits + Wrong\ Hits + False\ Alarms} 100\% \tag{5.3}$$

In addition, we computed the change per year during four intervals: reference change during the calibration interval (Eq. 5.4), the extrapolated change according to the Markov transition area matrix (Eq. 5.5), the simulated change according to CA-Markov (Eq. 5.6), and the reference change during the extrapolation interval (Eq. 5.7). The values were computed as square kilometres per year. If the model works as the user manual led us to believe, then Equation 5.5 would equal Equation 5.6.

$$\text{Annual reference change during calibration interval} = \frac{\text{Change area from 1995 to 2007}}{\text{12 years}} \tag{5.4}$$

$$\text{Annual change via Markov extrapolation} = \frac{\text{Markov change area from 2007 to 2018}}{\text{11 years}} \tag{5.5}$$

$$\text{Annual simulated change} = \frac{\text{Hits+Wrong Hits+ False Alarms from 2007 to 2018}}{\text{11 years}} \tag{5.6}$$

$$\text{Annual reference change during extrapolation interval} = \frac{\text{Change area from 2007 to 2018}}{\text{11 years}} \tag{5.7}$$

5.3. Results

We conducted 120 model runs consisting of one run for the two ways to format the input data and two ways to set the model's parameters concerning the spatial filter of the cellular automata. Figure 5.4 shows for each run a bar stacked with brown, blue, and orange segments. The sum of the brown, blue, and orange segments is the annual reference change during the calibration interval (Eq. 5.4). If the annual reference change were equal to the annual simulation change, then each stack would be entirely brown. However, the mathematical behaviour of Markov chains implies deceleration of change from the calibration time interval to the extrapolation time interval. The orange segment shows the deceleration of change that the Markov procedure implies. Simulations of the combined Markov and cellular automata had less change than the Markov procedure dictated. The blue segments indicate how much less the simulated change was than the Markov change. The brown bars show the simulated changes from the combined Markov and cellular automata (Eq. 5.6). If the simulation had conformed to the Markov procedure, then the simulated change would be the union of the brown and blue segments. The blue segments show that the CA-Markov simulated less change than what the Markov procedure dictated and that the filter shape and size influence the deficit. This deficit is more extreme with five categories and finer spatial resolutions.

Figure 5.4 uses dotted rectangles to show the annual reference change during the extrapolation interval (Eq. 5.7). Reference change during extrapolation interval is less than reference change during calibration interval because change on the landscape decelerated. If the CA-Markov model would have respected the deceleration that Markov implies, then simulated change would have been more than the reference change during the extrapolation interval because the change on the landscape decelerated from the calibration interval to the extrapolation interval. However, the effect of the cellular automata caused the simulated change to be less than the reference change during the extrapolation interval for all runs that used five categories. Figure 4 presents the variation in input data format vertically down the rows of boxes, and the variation in spatial filter horizontally across the columns of boxes. Most of the variation is among the rows of boxes that show the format of the input data. There is some variation horizontally due to the size of the spatial filter and trivial variation due to the shape of the filter. The user's most influential decision is the number of categories, then the data's spatial resolution. The user's decision concerning filter size is more influential than filter shape.

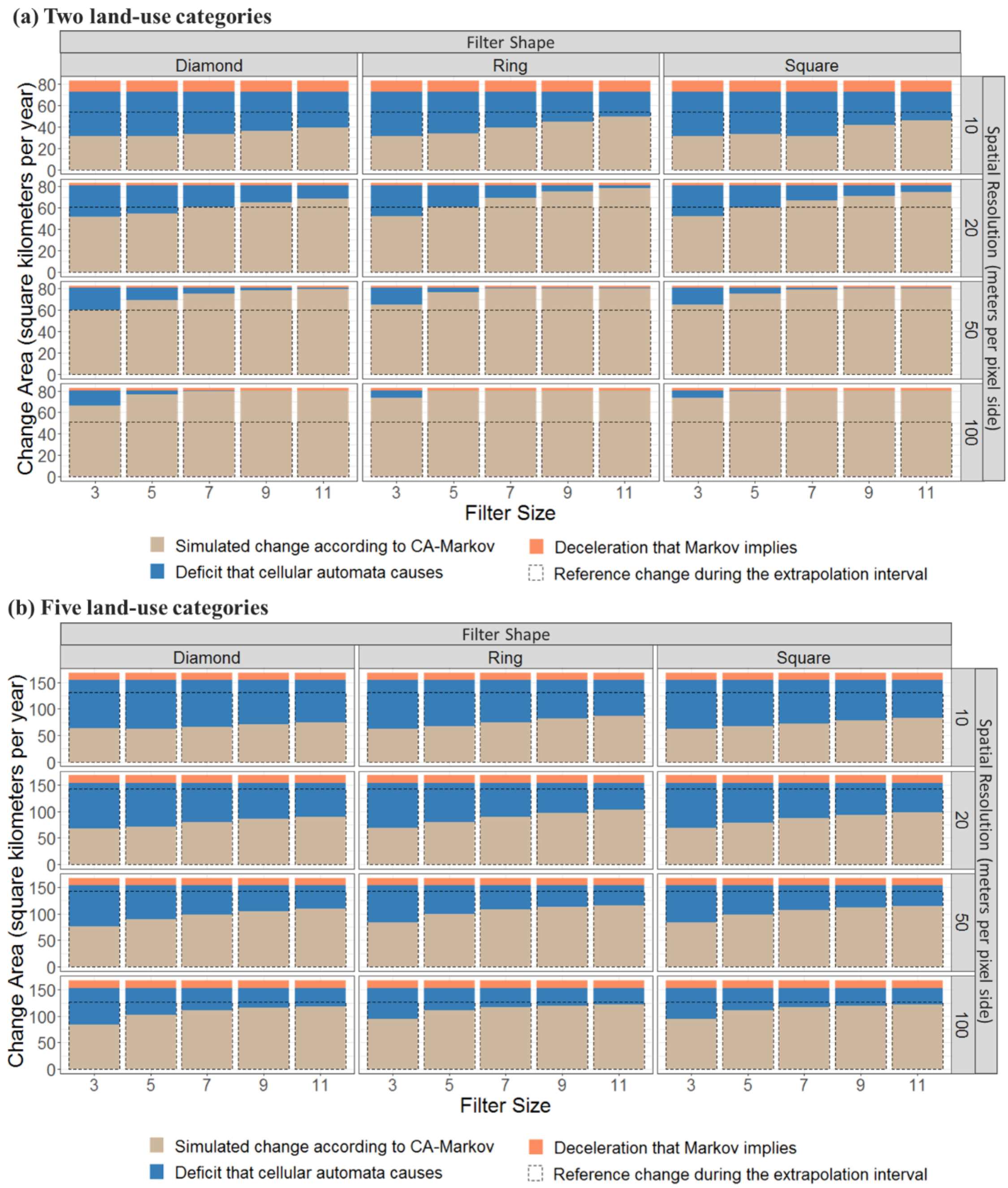
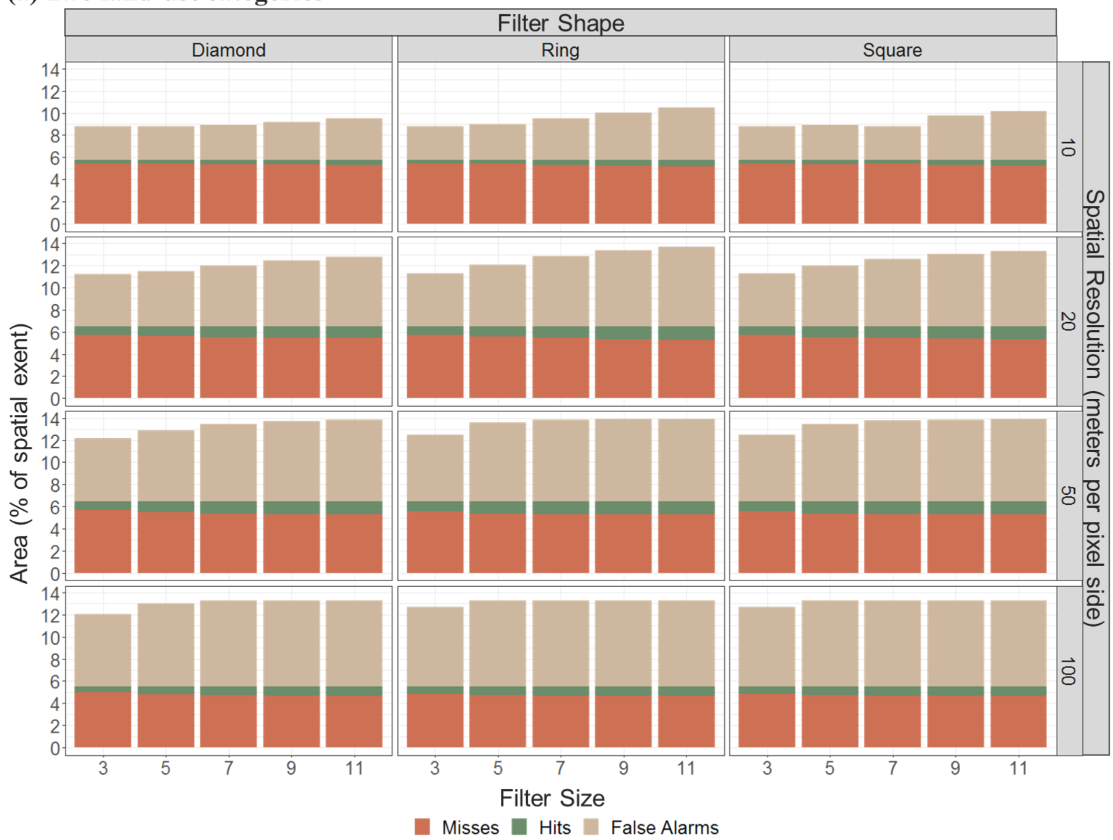


Figure 5.4. Annual change for 120 runs

Figure 5.5 presents the Misses, Hits, Wrong Hits, and False Alarms generated from the comparison of the reference map for 2007, the reference map for 2018, and each of the 120 simulated maps for 2018. Results vary most by the number of categories, then by the spatial resolution. The spatial filter's shape and size generate smaller variations in the four components of validation. The results confirm that the number of categories and the spatial resolution influence the extrapolated quantity. With two categories, the size of the overall error is 8 to 12 times the size of Hits. With five categories, the size of the overall error is 18 to 21 times the size of Hits. Coarser resolutions caused an increase in simulated change, thus an increase in False

Alarms. In addition, five categories for the input maps doubled the Misses and increased the sizes of Hits and False Alarms compared to the two categories for the input maps. False Alarms are smaller than Misses for most of the runs because the simulated changes during the extrapolation interval are smaller than the reference change during the extrapolation interval. The main reason for the deficit is that the cellular automata caused the simulated change to be less than the change that the Markov matrix dictated, which Figure 5.4 shows. If a user were to see only Figure 5.5, then the user would conclude that the error derives from how the model dictated less change than the reference change during the extrapolation interval, but Figure 5.4 shows that the proper conclusion is just the opposite. If simulation would have respected the Markov extrapolation, then the simulation would have simulated more change than what occurred according to the reference data during the extrapolation time interval. Thus the counterintuitive behaviour of the CA-Markov simulation is likely to lead to a misinterpretation of Figure 5.5 unless the user knows the counterintuitive behaviour of the model, which in the software's manual does not describe. All the FOM values are from 4 to 9%.

(a) Two land-use categories



(b) Five land-use categories

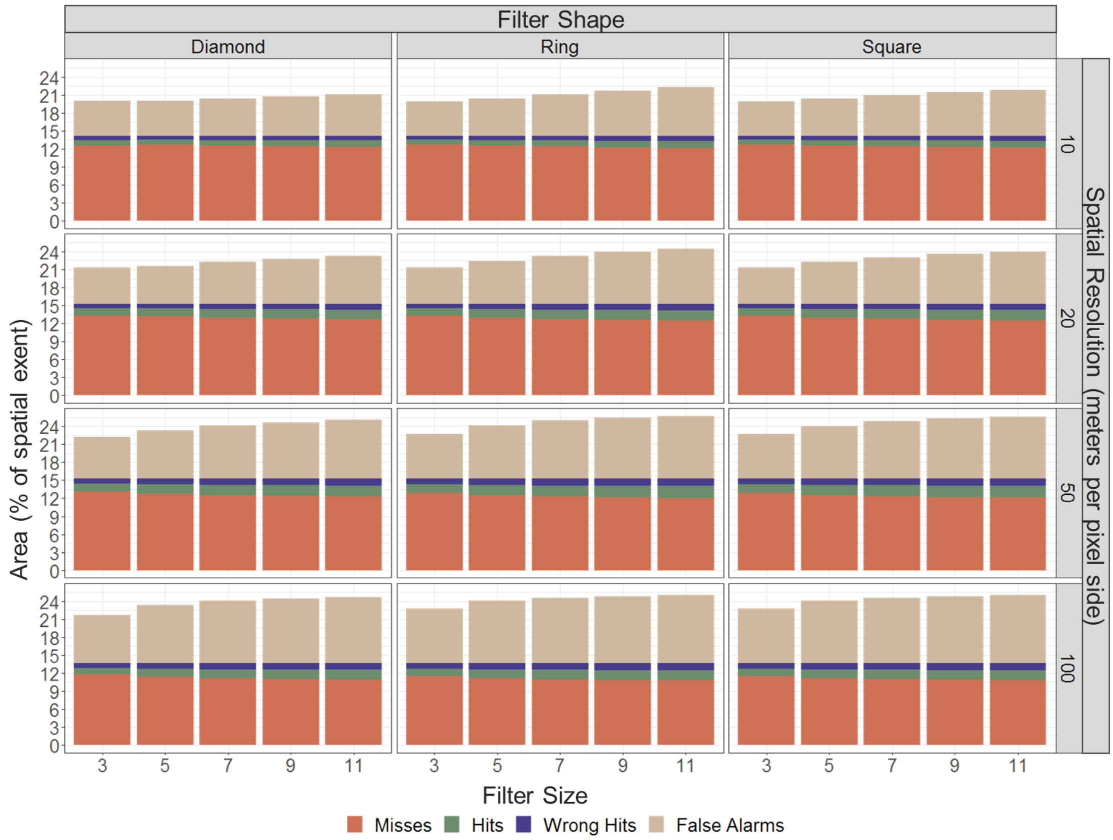


Figure 5.5. Overall simulation hits and errors in the spatial extent

Figure 5.6 shows the spatial allocation of Misses (12%), Hits (2%), Wrong Hits (1%), False Alarms (10%), and Correct rejections (75%) for the run with five land categories at the 50 m resolution with a square spatial filter shape and 11 × 11 size. The spatial filter causes the CA-Markov model to allocate the gain of a category around the patches where the category existed at the start of the simulation. Therefore, CA-Markov failed to simulate a leapfrog pattern, where an isolated category patch gains without spatial connection to nearby category patches. The FOM value is 7.7% for the result in Figure 5.6.

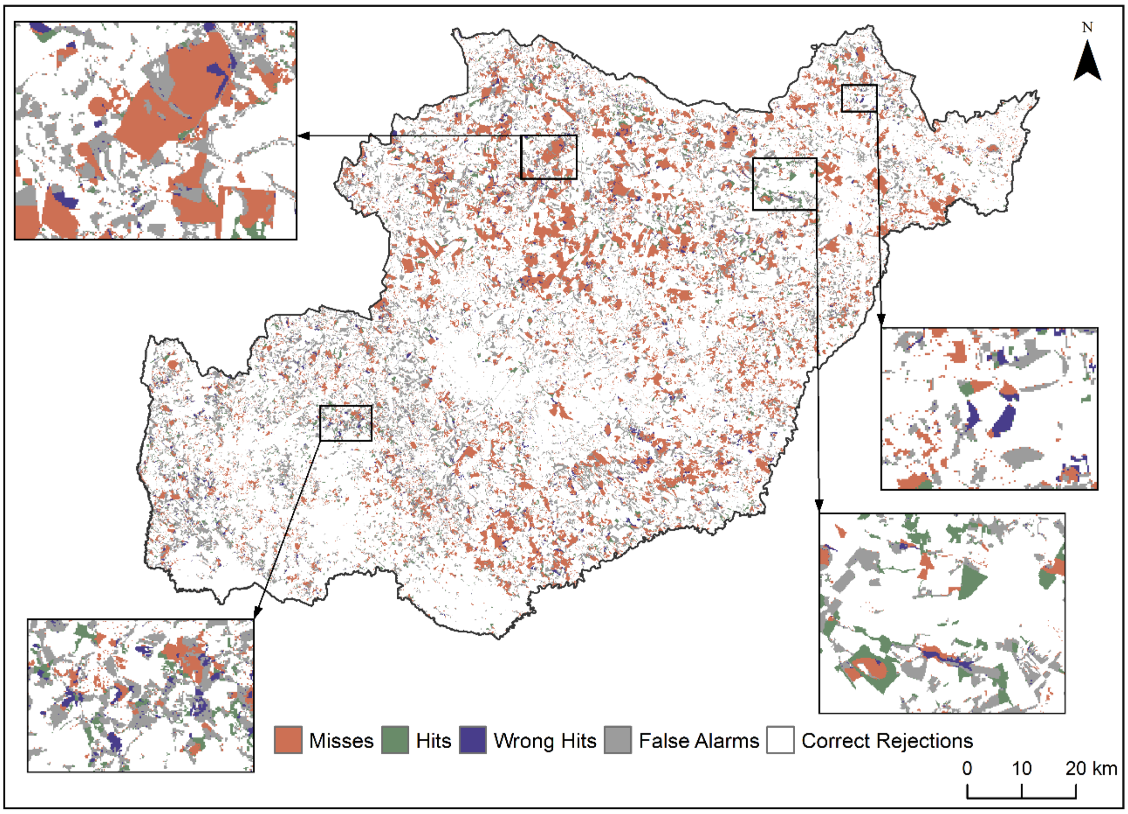


Figure 5.6. Example of the spatial allocation of hits, misses, false alarms, wrong hits, and correct rejections

5.4. Discussion

The four questions of our proposed evaluation framework are answered below to determine whether IDRISI's CA-Markov model is appropriate for our specific application. The first question is "*Can the user understand the model?*" Our experience was that our understanding required insights that are not in the software's documentation. We suspect that some users will not realize that the model simulates a different quantity of change than what the Markov transition area matrix dictates, thus are likely to misinterpret the results. The users who realize will not be able to understand why by reading the software's documentation, despite CA-Markov in IDRISI is better documented than most models and is integrated into a GIS that comes with customer support.

The second question "*Can the audience understand the model?*" The answer is unlikely given the answer to question 1. In the vast literature that has used Idiris's CA-Markov, we have never read a discussion that CA-Markov fails to follow the quantities that the Markov transition area matrix dictates, besides a few papers in which Pontius was an author (Pontius Jr & Malanson, 2005; Varga et al., 2019, 2020). If the user does understand how the model behaves, then the audience will unlikely be able to understand. If the user misinterprets the results, then the audience is also likely to misinterpret the results.

The third question is "*Can the user control the model?*" Our experience was No because we do not know why CA-Markov fails to follow its own Markov transition area matrix. The software dictates the model's design, which we do not control. CA-Markov advertises that Markov controls the quantity and CA controls the allocation. Indeed, our results show the Markov procedure extrapolates slower change than the calibration interval, which is a property of the Markov matrix. However, simulated change is slower than the Markov extrapolation, which the spatial filter influences. The user wants to modify only the spatial allocation of change, then the user could modify the spatial filter, and not realize that the spatial filter also influences the quantity of change. These results show that CA influences the quantity (Pontius Jr & Malanson, 2005; Varga et al., 2019). Thus, the user does not have complete control over the model. The model's parameters do not allow the user to specify the quantity separately from the allocation, therefore the user cannot control both important components to create various scenarios of simulated change. Moreover, the reference change slows from the calibration interval to the extrapolation interval, which is a function of the landscape. Thus, overall, the quantity error derives from a mix of the behavior of the Markov matrix, the CA's corruption of the quantities, and the non-stationarity of the reference data from the calibration interval to the extrapolation interval.

In addition, our results show the decisions concerning data format such as the number of categories and spatial resolution are fundamental for understanding the model output (Pontius Jr et al., 2018). In our experiment, the number of categories was more influential than the spatial resolution. For all our runs, the allocation error is more than twice the size of the quantity error, and the overall error is larger than the hits. The quantity difference in the validation compares the size of CA-Markov simulation to the size of reference change during extrapolation interval. Thus, the difficulty to understand the behavior of the model is likely to lead directly to a misinterpretation of the validation metrics especially when a user does only a few runs. Moreover, when the validation shows a disagreement in quantity, then the user will likely think that the quantity disagreement derives from a combination of the Markov extrapolation and the difference in speed of the reference change from the calibration interval to the extrapolation interval. But the real reason could be the effect of the spatial filter on reducing the simulated quantity. Figure 5.4a shows that variation in the filter size can cause the model to simulate more than the reference change during the extrapolation interval in some cases and less in other cases. This finding supports our answer to the first three questions.

The validation procedure reflects a combination of the extrapolation and the non-stationarity of the reference data from the calibration to the extrapolation interval. However, the purpose of the model is to extrapolate from the calibration interval not to predict the future. The software is responsible for the model extrapolating as advertised, not for the stationarity of the landscape through time. Nevertheless, a large number of users assume that the model is good when a validation metric is high, even when they lack a definition of high. Our experiment shows even when simulated quantity matches the reference quantity for the validation, such as the run in the lower right corner of Figure 4.4b, the reason for the match derives from a glitch in the model. Therefore, the user's misunderstanding will cause the user and the audience to misinterpret the results in the validation. Our findings illustrate why we must not evaluate the performance of a model based on the validation, but rather based on whether the model behaves in a manner that the user can understand and control by reading the software's documentation.

The last question is "*Does the model address the goals of the specific application?*" The answer to this question depends on the user's goals. We began the CA-Markov modeling to integrate knowledge regarding agricultural land system dynamics using a GIS-based approach. Particularly, we choose the CA-Markov model because authors of previous work routinely claimed that the simulation model was helpful to provide insights for conscious and thrifty management of land systems, which could result in effective environmental planning and informed policy decisions for the future. However, after using the CA-Markov model, we found that we could not communicate helpful insights for landscape management, due mainly to the fact that the answers to the first three questions were not Yes. If we could control the quantities, then we might have been able to use the model to express various scenarios but given that the CA influences the quantities in ways that the software's documentation does not describe, the CA part of the model would not allow the model to portray various scenario storylines that dictate the quantities. Therefore, the model did not address the goals of our specific application because we were not able to understand and control all aspects of the behaviour of the IDRISI's CA-Markov model. Our results exposed how CA-Markov behaved for our application, which is similar to the behaviour that Varga et al., (2019) observed in Hungry but is not necessarily how CA-Markov would behave in other applications.

Hence, based on our results and the considerations discussed above we question why so many authors continue to use CA-Markov regularly (e.g., Aksoy and Kaptan, 2021; Nyamekye et al., 2021) and claim that the model is useful and successful for sustainable management of environmental systems? Probably because they did not ask themselves the four questions. Our findings emphasize the importance of our four questions. Therefore, we intend that the four questions of our framework be used to guide users to decide whether to use any particular land change model, helping to identify more clearly the characteristics of a model simulation behavior before getting buried in the details of the analysis and preventing the misinterpretation of the outputs. It is not clear whether the documentation is any better for other land change models. We know of some models where it is difficult to obtain documentation or customer support.

5.5. Conclusion

The four questions proposed in this manuscript constitute a first step that is helpful to evaluate the usefulness of a land change simulation model to address the goals of a specific application. Our framework of evaluation is illustrated based on the CA-Markov model in the Selva version of IDRISI's software, while our framework can offer insights that could be relevant to other models. Our questions were helpful to evaluate the appropriateness of CA-Markov model for our purpose. In particular, we were not able to understand or control the model because the CA-Markov model simulates a different quantity of change than what the Markov transition area matrix dictates, which is a behavior that the software's documentation does not describe. In addition, 120 runs revealed the decisions concerning data format such as the number of categories and spatial resolution as well as the spatial filter are fundamental for understanding the model output. For all runs, the allocation error was more than twice the size of the quantity error, and the size of the overall error was 8 to 21 times the size of hits. Thus, if the purpose of the model is to predict change accurately, then our selection of parameters and suitability maps were not appropriate or the CA-Markov algorithm might not be effective for our landscape. If the purpose of the model is to portray possible scenarios, then we would have had difficulty because we could not control the quantity of simulated change.

Clark Labs introduced the CA-Markov decades ago as an experimental model, and CA-Markov remains popular. Our manuscript reveals difficulty in understanding and controlling the CA-Markov model. Clark Labs has since developed the Land Change Modeler, which does not have the spatial filter that causes CA-Markov to lose control of the simulated quantity. A next step should be to evaluate the Land Change Modeler in TerrSet (formerly IDRISI) and its documentation with respect to our four questions. We encourage users to ask themselves our four questions during all phases of modeling to determine the appropriateness of any model.

References

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P., & Schweik, C. M. (2002). A Review and Assessment of Land-Use Change Models: Dynamics of Space, Time, and Human Choice. Apollo The International Magazine Of Art And Antiques, 62. https://doi.org/10.1289/ehp.6514

Aksoy, H., & Kaptan, S. (2021). Monitoring of land use/land cover changes using GIS and CA-Markov modeling techniques: a study in Northern Turkey. Environmental Monitoring and Assessment 2021 193:8, 193(8), 1–21. https://doi.org/10.1007/S10661-021-09281-X

Aliani, H., Malmir, M., Sourodi, M., & Kafaky, S. B. (2019). Change detection and prediction of urban land use changes by CA–Markov model (case study: Talesh County). Environmental Earth Sciences, 78(17), 546. https://doi.org/10.1007/s12665-019-8557-9

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., & Andreassian, V. (2013). Characterising performance of environmental models. Environmental Modelling & Software, 40, 1–20. https://doi.org/10.1016/J.ENVSOFT.2012.09.011

Bradley, A. V., Rosa, I. M. D., Pontius, R. G., Ahmed, S. E., Araújo, M. B., Brown, D. G., Brandão, A., Câmara, G., Carnerio, T. G. S., Hartley, A. J., Smith, M. J., & Ewers, R. M. (2016). SimiVal, a multi-criteria map comparison tool for land-change model projections. Environmental Modelling & Software, 82, 229–240. https://doi.org/10.1016/J.ENVSOFT.2016.04.016

Dahal, K. R., & Chow, T. E. (2015). Characterization of neighborhood sensitivity of an irregular cellular automata model of urban growth. International Journal of Geographical Information Science, 29(3), 475–497. https://doi.org/10.1080/13658816.2014.987779

DGT. (2018). Especificações Técnicas da Carta de Uso e Ocupação do Solo (COS) de Portugal Continental para 1995, 2007, 2010 e 2015.

Eastman, J. R., & Toledano, J. (2018). A Short Presentation of CA_MARKOV. In M. T. Camacho Olmedo, M. Paegelow, J. F. Mas, & F. Escobar (Eds.), Geomatic Approaches for Modeling Land Change Scenarios (1st ed., pp. 481–484). Springer International Publishing. https://doi.org/10.1007/978-3-319-60801-3_33

Eastman, J.R. (2012). Idrisi Selva Manual. Clark University, Worcester MA.

Eastman, J Ronald. (2006). Guide to GIS and Image Processing. Clark University: Worcester, MA, USA, 1(May), 87–131.

Fuglsang, M., Münier, B., & Hansen, H. S. (2013). Modelling land-use effects of future urbanization using cellular automata: An Eastern Danish case. Environmental Modelling & Software, 50, 1–11. https://doi.org/10.1016/J.ENVSOFT.2013.08.003

Ghosh, P., Mukhopadhyay, A., Chanda, A., Mondal, P., Akhand, A., Mukherjee, S., Nayak, S. K., Ghosh, S., Mitra, D., Ghosh, T., & Hazra, S. (2017). Application of Cellular automata and Markov-chain model in geospatial environmental modeling- A review. Remote Sensing Applications: Society and Environment, 5(January), 64–77. https://doi.org/10.1016/j.rsase.2017.01.005

Guan, D., Li, H., Inohae, T., Su, W., Nagaie, T., & Hokao, K. (2011). Modeling urban land use change by the integration of cellular automaton and Markov model. Ecological Modelling, 222(20–22), 3761–3772. https://doi.org/10.1016/j.ecolmodel.2011.09.009

Halmy, M. W. A., Gessler, P. E., Hicke, J. A., & Salem, B. B. (2015). Land use/land cover change detection and prediction in the north-western coastal desert of Egypt using Markov-CA. Applied Geography, 63, 101–112. https://doi.org/10.1016/j.apgeog.2015.06.015

Hamad, R., Balzter, H., & Kolo, K. (2018). Predicting land use/land cover changes using a CA-Markov model under two different scenarios. Sustainability (Switzerland), 10(10), 3421. https://doi.org/10.3390/su10103421

Houet, T., Verburg, P. H., & Loveland, T. R. (2010). Monitoring and modelling landscape dynamics. In Landscape Ecology (Vol. 25, Issue 2, pp. 163–167). Springer. https://doi.org/10.1007/s10980-009-9417-x

Huang, Y., Yang, B., Wang, M., Liu, B., & Yang, X. (2020). Analysis of the future land cover change in Beijing using CA–Markov chain model. Environmental Earth Sciences, 79(2), 60. https://doi.org/10.1007/s12665-019-8785-z

Hyandye, C., & Martz, L. W. (2017). A Markovian and cellular automata land-use change predictive model of the Usangu Catchment. International Journal of Remote Sensing, 38(1), 64–81. https://doi.org/10.1080/01431161.2016.1259675

INE. (2011). XV Recenseamento Geral da População. Instituto Nacional de Estatística.

Liao, J., Tang, L., Shao, G., Su, X., Chen, D., & Xu, T. (2016). Incorporation of extended neighborhood mechanisms and its impact on urban land-use cellular automata simulations. Environmental Modelling and Software, 75, 163–175. https://doi.org/10.1016/j.envsoft.2015.10.014

Lin, J., Li, X., Li, S., & Wen, Y. (2020). What is the influence of landscape metric selection on the calibration of land-use/cover simulation models? Environmental Modelling & Software, 129, 104719. https://doi.org/10.1016/J.ENVSOFT.2020.104719

Liping, C., Yujun, S., & Saeed, S. (2018). Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques—A case study of a hilly area, Jiangle, China. PLoS ONE, 13(7), e0200493. https://doi.org/10.1371/journal.pone.0200493

Mas, J.-F., Kolb, M., Paegelow, M., Olmedo, M. T. C., & Houet, T. (2014). Inductive pattern-based land use/cover change models: A comparison of four software packages. Environmental Modelling & Software, 94–111. https://doi.org/http://dx.doi.org/10.1016/j.envsoft.2013.09.010

Muller, M. R., & Middleton, J. (1994). A Markov model of land-use change dynamics in the Niagara Region, Ontario, Canada. Landscape Ecology, 9(2), 151–157. https://doi.org/10.1007/BF00124382

Murray-Rust, D., Robinson, D. T., Guillem, E., Karali, E., & Rounsevell, M. (2014). An open framework for agent based modelling of agricultural land use change. Environmental Modelling & Software, 61, 19–38. https://doi.org/10.1016/J.ENVSOFT.2014.06.027

Noszczyk, T. (2019). A review of approaches to land use changes modeling. Human and Ecological Risk Assessment: An International Journal, 25(6), 1377–1405. https://doi.org/10.1080/10807039.2018.1468994

Nyamekye, C., Kwofie, S., Agyapong, E., Ofosu, S. A., Arthur, R., & Appiah, L. B. (2021). Integrating support vector machine and cellular automata for modelling land cover change in the tropical rainforest under equatorial climate in Ghana. Current Research in Environmental Sustainability, 3, 100052. https://doi.org/10.1016/j.crsust.2021.100052

Olmedo, M. T. C., Pontius Jr, R. G., Paegelow, M., & Mas, J. F. (2015). Comparison of simulation models in terms of quantity and allocation of land change. Environmental Modelling and Software, 69, 214–221. https://doi.org/10.1016/j.envsoft.2015.03.003

Palmate, S. S., Pandey, A., & Mishra, S. K. (2017). Modelling spatiotemporal land dynamics for a trans-boundary river basin using integrated Cellular Automata and Markov Chain approach. Applied Geography, 82, 11–23. https://doi.org/10.1016/j.apgeog.2017.03.001

Pan, Y., Roth, A., Yu, Z., & Doluschitz, R. (2010). The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. Computers, Environment and Urban Systems, 34, 400-408. https://doi.org/10.1016/j.compenvurbsys.2010.03.003

Perica, S., & Foufoula-Georgiou, E. (1996). Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions. Journal of Geophysical Research Atmospheres, 101(21), 26347–26361. https://doi.org/10.1029/96jd01870

Pontius Jr, G. R., & Malanson, J. (2005). Comparison of the structure and accuracy of two land change models. International Journal of Geographical Information Science, 19(2), 243–265. https://doi.org/10.1080/13658810410001713434

Pontius Jr, R. G., Boersma, W., Castella, J. C., Clarke, K., Nijs, T., Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C. D., McConnell, W., Mohd Sood, A., Pijanowski, B., Pithadia, S., Sweeney, S., Trung, T. N., Veldkamp, A. T., & Verburg, P. H. (2008). Comparing the input, output, and validation maps for several models of land change. Annals of Regional Science, 42(1), 11–37. https://doi.org/10.1007/s00168-007-0138-2

Pontius Jr, R. G., Castella, J.-C., de Nijs, T., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C. D., McConnell, W., Mohd Sood, A., Pijanowski, B., Verburg, P., & Veldkamp, A. T. (2018). Lessons and Challenges in Land Change Modeling Derived from Synthesis of Cross-Case Comparisons. In Geotechnologies and the Environment (Vol. 19, pp. 143–164). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-52522-8_8

Pontius Jr, R. G., Peethambaram, S., & Castella, J.-C. (2011). Comparison of Three Maps at Multiple Resolutions: A Case Study of Land Change Simulation in Cho Don District, Vietnam. Annals of the Association of American Geographers, 101(1), 45–62. https://doi.org/10.1080/00045608.2010.517742

Radwan, T. M., Blackburn, G. A., Whyatt, J. D., & Atkinson, P. M. (2019). Dramatic Loss of Agricultural Land Due to Urban Expansion Threatens Food Security in the Nile Delta, Egypt. Remote Sensing, 11(3), 332. https://doi.org/10.3390/rs11030332

Rimal, B., Zhang, L., Keshtkar, H., Wang, N., & Lin, Y. (2017). Monitoring and Modeling of Spatiotemporal Urban Expansion and Land-Use/Land-Cover Change Using Integrated Markov Chain Cellular Automata Model. ISPRS International Journal of Geo-Information, 6(9), 288. https://doi.org/10.3390/ijgi6090288

Roodposhti, M. S., Aryal, J., & Bryan, B. A. (2019). A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. Environmental Modelling & Software, 112, 70–81. https://doi.org/10.1016/J.ENVSOFT.2018.10.006

Samat, N. (2006). Characterizing the scale sensitivity of the cellular automata simulated urban growth: A case study of the Seberang Perai Region, Penang State, Malaysia. Computers, Environment and Urban Systems, 30, 905–920. https://doi.org/10.1016/j.compenvurbsys.2005.11.002

Sang, L., Zhang, C., Yang, J., Zhu, D., & Yun, W. (2011). Simulation of land use spatial pattern of towns and villages based on CA-Markov model. Mathematical and Computer Modelling, 54(3–4), 938–943. https://doi.org/10.1016/j.mcm.2010.11.019

Sankarrao, L., Ghose, D. K., & Rathinsamy, M. (2021). Predicting land-use change: Intercomparison of different hybrid machine learning models. Environmental Modelling & Software, 145, 105207. https://doi.org/10.1016/J.ENVSOFT.2021.105207

Santé, I., García, A. M., Miranda, D., & Crecente, R. (2010). Cellular automata models for the simulation of real-world urban processes: A review and analysis. Landscape and Urban Planning, 96(2), 108–122. https://doi.org/10.1016/j.landurbplan.2010.03.001

Torrens, P. M. (2000). How cellular models of urban systems work (1. theory). Centre for Advances Spatial Analysis Working Paper 28. University College London, UK.

van Vliet, J., Bregt, A. K., Brown, D. G., van Delden, H., Heckbert, S., & Verburg, P. H. (2016). A review of current calibration and validation practices in land-change modeling. In Environmental Modelling and Software (Vol. 82, pp. 174–182). Elsevier Ltd. https://doi.org/10.1016/j.envsoft.2016.04.017

Varga, O. G., Pontius Jr, R. G., Singh, S. K., & Szabó, S. (2019). Intensity Analysis and the Figure of Merit's components for assessment of a Cellular Automata – Markov simulation model. Ecological Indicators, 101, 933–942. https://doi.org/10.1016/j.ecolind.2019.01.057

Varga, O. G., Pontius Jr, R. G., Szabó, Z., & Szabó, S. (2020). Effects of Category Aggregation on Land Change Simulation Based on Corine Land Cover Data. Remote Sensing, 12(8), 1314. https://doi.org/10.3390/rs12081314

Verburg, P. H., de Nijs, T. C. M., van Eck, J. R., Visser, H., & de Jong, K. (2004). A method to analyse neighbourhood characteristics of land use patterns. Computers, Environment and Urban Systems, 28: 667–690. https://doi.org/10.1016/j.compenvurbsys.2003.07.001

Verweij, P., Cormont, A., Kok, K., van Eupen, M., Janssen, S., te Roller, J., de Winter, W., Pérez-Soba, M., & Staritsky, I. G. (2018). Improving the applicability and transparency of land use change modelling: The iCLUE model. Environmental Modelling & Software, 108, 81–90. https://doi.org/10.1016/J.ENVSOFT.2018.07.010

Viana, C. M., Girão, I., & Rocha, J. (2019). Long-Term Satellite Image Time-Series for Land Use/Land Cover Change Detection Using Refined Open Source Data in a Rural Region. Remote Sensing, 11(9), 1104. https://doi.org/10.3390/rs11091104

Viana, C. M., & Rocha, J. (2020). Evaluating Dominant Land Use/Land Cover Changes and Predicting Future Scenario in a Rural Region Using a Memoryless Stochastic Method. Sustainability, 12(10), 4332. https://doi.org/10.3390/su12104332

Wang, R., & Murayama, Y. (2017). Change of Land Use/Cover in Tianjin City Based on the Markov and Cellular Automata Models. ISPRS International Journal of Geo-Information, 6(5), 150. https://doi.org/10.3390/ijgi6050150

White, R., & Engelen, G. (2000). High resolution integrated modelling of the spatial dynamics of urban and regional systems. Computers, Environment and Urban Systems, 24: 383-440.

Wolfram, S. (1984). Cellular Automata as models of complexity. Nature, 311, 5985: 419–424.

Wu, H., Zhou, L., Chi, X., Li, Y., & Sun, Y. (2012). Quantifying and analyzing neighborhood configuration characteristics to cellular automata for land use simulation considering data source error. Earth Science Informatics, 5(2), 77–86. https://doi.org/10.1007/s12145-012-0097-8

Yirsaw, E., Wu, W., Shi, X., Temesgen, H., & Bekele, B. (2017). Land Use/Land Cover Change Modeling and the Prediction of Subsequent Changes in Ecosystem Service Values in a Coastal Area of China, the Su-Xi-Chang Region. Sustainability, 9(7), 1204. https://doi.org/10.3390/su9071204

Chapter 6. Evaluating Dominant Land Use/Land Cover Changes and Predicting Future Scenario in a Rural Region Using a Memoryless Stochastic Method

Cláudia M. Viana, Jorge Rocha

Article published at the Sustainability (2020), Volume 12, Part 10, 4332.
https://doi.org/10.3390/su12104332
The structure and formatting are slightly adapted.

Abstract

The present study used the official Portuguese land use/land cover (LULC) maps (*Carta de Uso e Ocupação do Solo*, COS) from 1995, 2007, 2010, 2015, and 2018 to quantify, visualize, and predict the spatiotemporal LULC transitions in the Beja district, a rural region in the southeast of Portugal, which is experiencing marked landscape changes. Here, we computed the conventional transition matrices for in-depth statistical analysis of the LULC changes that have occurred from 1995 to 2018, providing supplementary statistics regarding the vulnerability of inter-class transitions by focusing on the dominant signals of change. We also investigated how the LULC is going to move in the future (2040) based on matrices of current states using the Discrete-Time Markov Chain (DTMC) model. The results revealed that, between 1995 and 2018, about 28% of the Beja district landscape changed. Particularly, croplands remain the predominant LULC class in more than half of the Beja district (in 2018 about 64%). However, the behavior of the inter-class transitions was significantly different between periods, and explicitly revealed that arable land, pastures, and forest were the most dynamic LULC classes. Few dominant (systematic) signals of change during the 1995–2018 period were observed, highlighting the transition of arable land to permanent crops (5%) and to pastures (2.9%), and the transition of pastures to forest (3.5%) and to arable land (2.7%). Simulation results showed that about 25% of the territory is predicted to experience major LULC changes from arable land (−3.81%), permanent crops (+2.93%), and forests (+2.60%) by 2040.

Keywords: landscape pattern; transition matrix; systematic processes; change detection; Discrete Time Markov Chains

6.1. Introduction

In many areas of the globe, the scale and extent of land use/land cover (LULC) changes have been affected by socioeconomic and biophysical factors (e.g., (Feranec et al., 2017; Fuchs et al., 2015). More specifically, environmental conditions, internal and external policies, the functioning of local and national markets, and the demographic situation have been pointed out as the main causes directly responsible for the LULC changes (Deng et al., 2008; Qingshui Lu et al., 2011; Turner et al., 1993). Therefore, over the last decade, for dynamic visualization and quantification of the spatial patterns of the LULC changes, Geographic Information Systems (GIS) have been widely used, since they allow for the use of vector or remote sensing data, expressed spatially and temporally (Hansen & Loveland, 2012; Rousta et al., 2018; Uddin et al., 2018). Particularly, LULC change monitoring and assessment is a practice that assists a more conscious and thrifty management of this natural resource, which could result in effective environmental planning and informed policy decisions.

Commonly, in LULC change analysis, the LULC transitions are quantified and visualized based on a transition matrix by computing the inter-class percentage and by measuring, e.g., the persistence, gain/loss, net, and rate of the changes, in order to achieve comprehensive monitoring of physical changes over time (Mertens et al., 2000; Petit et al., 2001). Although these are the most commonly used measures, since they are very instructive for understanding the nuance of LULC changes, there are other measures that are equally informative and can contribute to more insights regarding LULC change processes (Pontius Jr, Shusas, & Mceachern, 2004) and to the design of anticipatory measure responses to changes. For instance, the quantification of both net change and swap (which comprises both the gain and loss of an LULC class in different locations of the study area), as well the gross gains/losses, allows for the detection of the inter-class transitions that have occurred as a systematic or as a random process of change (Braimoh, 2006; Pontius Jr et al., 2004). Thus, the dominant signals of LULC changes can be distinguished.

Hence, whereas a systematic transition occurs by a continuous or general process of change, having a gradual tendency for progressive increment influenced by different factors (e.g., population growth or internal and external policies) (Lambin, Geist, & Lepers, 2003), a random transition occurs as a single abrupt episode process of change influenced by factors that occur suddenly (e.g., floods or wildfires) (Braimoh, 2006; Lambin et al., 2003). Accordingly, distinguishing systematic (dominant) processes within a pattern of LULC changes, which tends to evolve in a progressive way driven by established and continued processes (see, e.g., Braimoh, 2006; Teferi et al., 2013), allows one to focus on the mechanisms of change and further in-depth assessment of the potential change's main drivers. Therefore, a better understanding of the LULC dynamics and better policies and measures in line can reduce the adverse impacts of dominant LULC changes (Braimoh, 2006; Teferi et al., 2013).

In addition, apart from LULC change dynamics analysis, since the end of the 20th century, the prediction of future scenarios at different spatial and temporal scales has been extensively carried out (Han et al., 2015; Hyandye & Martz, 2017; Yirsaw et al., 2017). In fact, the study of past and future LULC changes has been very promising for understanding past trends (Fuchs et al., 2015) and potential future developments, being a decision-making basis for environmental management and the sustainability of the ecosystems. Therefore, as a complex process, LULC dynamics is regularly modeled through several methods that are efficient in the simulation and prediction of future LULC (Agarwal et al., 2002), including the discrete-time Markov chain (DTMC) model (more commonly known as the Markov chain (Gagniuc, 2017)) (Hathout, 2002; Weng, 2002), the cellular automata (CA) model (Hamad et al., 2018; Liu & Feng, 2016; Yirsaw et al., 2017), and neural networks (Li & Yeh, 2002), among others. Of these approaches, the DTMC model is of countless importance for modeling LULC changes (Ahmed et al., 2013; Basharin et al., 2004; Iacono et al., 2015), as it has the advantages of a time dimension and is able to represent the LULC change data, since it has the ability to quantify not only the states of conversion between LULC classes but also the rate of conversion among them. Thus, it is a stochastic process, where space is discrete (i.e., mainly based on probabilities, not certainties).

Context and Background

The last decades of the 20th century and the first decades of the 21st century have shown that European and Mediterranean rural areas have been undergoing significant changes (Feranec et al., 2010; Pinto-Correia, 2000). Indeed, one of the most remarkable changes in LULC in Europe, identified in Fuchs et al.'s (Fuchs et al., 2015) research, is the increase in forest to the detriment of the agricultural land. Likewise, the same trend has been observed in Portugal (Jan Feranec et al., 2010). Indeed, the LULC changes in Portugal have been the most dynamic among the European Union (EU) countries (Allen et al., 2018; Jan Feranec et al., 2010).

For instance, despite the existence of numerous studies regarding the LULC changes in Portugal (Meneses et al., 2018; Tavares et al., 2012; Viana, Girão, et al., 2019), the primary research focus has been on urban areas rather than rural areas. However, rural areas and agriculture have an important role in the regional and national economy of individual countries, as they constitute an important source of income for part of the population. Therefore, it is urgent to monitor the dynamics of rural LULC changes to promote and ensure a weighted and parsimonious use of the available resources.

Accordingly, given the stark changes of the Portuguese landscape, this study seeks to unravel the LULC changes situation in the Beja district, the largest Portuguese district in terms of area and the second in terms of percentage of agricultural land (about 47% in 2010); a rural region where we can find the *Montado* (named *Montado* in Portugal and *Dehesa* in Spain), an agrosilvopastoral agricultural heritage system (Correia, 1993; Muñoz-Rojas et al., 2019), that has been indicated as a globally important agricultural system according to the Globally Important Agricultural Heritage System (GIAHS) program promoted by the Food and Agriculture Organization of the United Nations (FAO) (Koohafkan & Altieri, 2016).

Moreover, this research is an effort to evaluate the spatiotemporal LULC changes with a particular focus on the evaluation of the dominant LULC changes and to simulate the future changes, in order to determine the main LULC processes' transformations and suggest new strategies for formulating better land use policies. The analysis was performed in a 10,229.05 km² area in the southeast of Portugal, which is experiencing marked landscape changes (Viana, Girão, et al., 2019). Thus, the main objectives of the present study were threefold: (i) identify and analyze the spatiotemporal dynamics of LULC over the past 23 years (1995–2018); (ii) quantify and analyze in depth the inter-class dominant (systematic) LULC transitions; and (iii) quantify the states of conversion between LULC classes and simulate and analyze the future LULC changes by 2040 using the DTMC model.

Our research is organized as follows: Chapter 6.2 details the study area, the used data, and the applied methodology. The results of the LULC changes and future LULC developments are described in Chapter 6.3. Finally, in Chapter 6.4, we discussed the main findings of our results and the implications of the applied methodology, presenting also the main conclusions of our research.

6.2. Materials and Methods

6.2.1. Study area

The study area, the Beja district, is located between 07°37'58" W and 07°41'17" W longitude and between 38°08'35" N and 38°10'08" N latitude in the southeast of Portugal in the Nomenclature of Territorial Units for Statistics (NUTS) II Alentejo litoral and Baixo Alentejo (Figure 6.1). The district of Beja is bordered to the north by the Évora district, to the south by the Algarve region, to the east by Spain, and to the west by the district of Setúbal and the Atlantic Ocean. Beja is the largest Portuguese district, with an area of 10,229.05 km² and a resident population of 152,758 inhabitants in 2011 (INE, 2012).

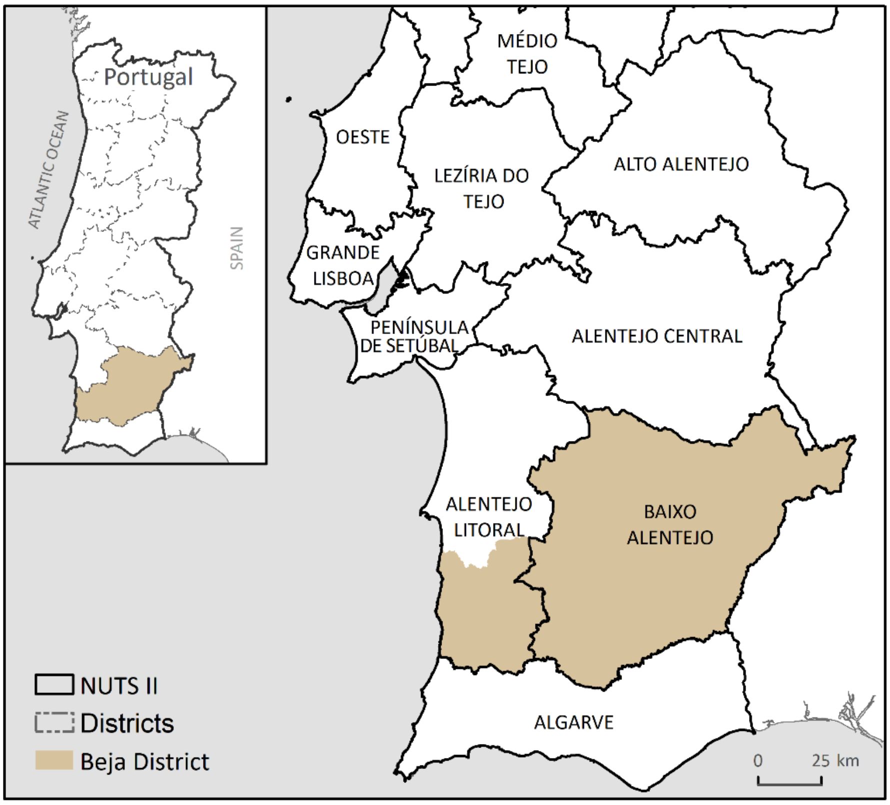


Figure 6.1. Location of study area

This region is characterized by a vast landscape of wheat, cork oaks, and olive trees, where the dominant land use is mixed agrosilvopastoral. The landscape of this region is distinctive, with both fragmented parcels and more compact parcels of land due to the different cropping calendars and field geometries. Dispersed settlements are the predominant urban form in this region. The climate is influenced by its distance from the coast, with a Mediterranean climate characterized by hot and dry summers and wet and cold winters. To the southeast, the territory is flat; to the north and west, the extensive plains are cut by tiny hills, the valley of the Guadiana River is the district's main geographical feature, which crosses from north to south in the eastern part of the district. The Alqueva dam, the largest in the country, is located in this district.

6.2.2. Portuguese Land Cover (COS) Data

The official spatial land cover data (COS) produced by the Portuguese General Directorate for Territorial Development (DGT) for 1995, 2007, 2010, 2015, and 2018 were used. These maps have the highest cartographic detail for the study area and are available for free on the DGT website (http://mapas.dgterritorio.pt/geoportal/catalogo.html). The COS maps are produced by photointerpretation (i.e., over orthophoto maps), with vector (polygons) structure data, a cartographic accuracy of 0.5 m, a minimum mapping unit (MMU) of 1 ha, and a scale of 1:25,000 (DGT, 2018). The five datasets have a hierarchical nomenclature system (a DGT report describes in detail the COS nomenclature characteristics (DGT, 2018)).

Spatiotemporal LULC change analysis was developed considering three major classes of the first hierarchical level of the COS nomenclature, namely, artificial surfaces, forest, and water bodies. In order to specify the major changes in croplands and characterize the type of agricultural activity in this region, we used the second hierarchical level of the COS nomenclature of the agriculture mega class, namely arable land, permanent crops, pastures, and heterogeneous areas. Therefore, a total of seven classes were used (Table 6.1). All LULC maps were clipped to the extent of the case study area, converted to a raster format of 100 × 100 m, and reclassified into seven classes. The cartographic information treatment and the spatiotemporal analysis were developed in QGIS software.

Table 6.1. Land use/Land cover (LULC) class description

LULC Classes	**Description**
Artificial surfaces	Urban fabric; industrial, commercial, and transport units; mine, dump, and construction sites; artificial non-agricultural vegetated areas.
Agricultural - Arable land	Lands under a rotation system used for annually harvested plants and fallow lands, which are rain-fed or irrigated. Includes flooded crops such as rice fields and other inundated croplands.
Agricultural - Permanent crops	Lands not under a rotation system. Includes ligneous crops of standard cultures for fruit production such as extensive fruit orchards, olive groves, chestnut groves, walnut groves shrub orchards such as vineyards and some specific low-system orchard plantation, espaliers, and climbers.
Agricultural - Pastures	Lands that are permanently used (at least 5 years) for fodder production. Includes natural or sown herbaceous species, unimproved or lightly improved meadows and grazed or mechanically harvested meadows. Regular agriculture impact influences the natural development of natural herbaceous species composition.
Agricultural - Heterogeneous areas	Areas of annual crops associated with permanent crops on the same parcel, annual crops cultivated under forest trees, areas of annual crops, meadows and/or permanent crops which are juxtaposed, landscapes in which crops and pastures are intimately mixed with natural vegetation or natural areas.
Forest	Forests; shrub and/or herbaceous vegetation associations; open spaces with little or no vegetation.
Water bodies	Inland wetlands; coastal wetlands; marine waters.

6.2.3. Spatiotemporal LULC Change Analysis

To identify and analyze the spatiotemporal dynamics of LULC in the Beja district, we first presented a brief evolution and distribution analysis of each class between 1995 and 2018. We quantified the changes for each year and reported them spatially and temporally. Secondly, we computed four change matrices in order to understand the main transition dynamics within each LULC class during the periods 1995–2007, 2007–2010, 2010–2015, and 2015–2018. The matrices were presented in graphical form. Each graph for each period shows the changes within each class from the initial time to the final time.

6.2.4. LULC Change Detection

For the LULC change detection in-depth analysis, we compared two maps (1995 and 2018) and produced a cross-tabulation matrix for the overall period (1995–2018). We decided to only analyze the overall period since each period comprises different years. While the first period has 12 years (1995–2007), the second (2007–2010) and the last (2015–2018) have 3 years, and the third has 5 years (2010–2015), so the change detection analysis could be influenced by the year's discrepancy between periods. Thus, the transition matrix represents the percentage of LULC changes within each class (j), where the diagonal values indicate the percentage of LULC that persists (P_{jj}) for each LULC class from initial time (T_1) (1995) to the final time (T_2) (2018), rows indicating the transitions from initial time $\left(P_{j+}\right)$, and columns indicating the transitions in the final time (P_{+j}).

By following Pontius et al.'s (Pontius Jr et al., 2004) study, we obtained several measures from the transition matrix result and grouped them into two categories. The first category includes the (i) persistence, (ii) gain, (iii) loss, (iv) total change, (v) absolute value of net change, and (vi) swapping tendency. Since net change cannot detect the swap between each class, and therefore underestimates the LULC total change, the swapping tendency is calculated, since it comprises both the gain and loss of an LULC class in different locations of the study area (Pontius Jr et al., 2004). Additionally, we presented the LULC vulnerability to transition by calculating the gain-to-loss, lost-to-persistence, and the gain-to-persistence ratios. A description of each measure is in Table 6.2.

Table 6.2. Description of each measure for LULC change detection analysis

Persistence (P_{jj})	The amount of LULC for each class which remains from initial time and the later time.
Gain (G_j)	The amount of gains is the difference between the total value of each LULC class for the later time (P_{+j}) and the value of persistency (P_{jj}).
Loss (L_j)	The amount of losses is the difference between the total for the initial time (P_{j+}) and the value of persistency (P_{jj}).
Total change (C_j)	The total change (overall change) is the sum of gain (G_j) and loss (L_j).
Net change (D_j)	The amount of net change is the maximum value of the gains and losses (P_{+j}) minus the minimum value of gains and losses (P_{j+}).
Swap tendency (S_j)	The amount of swap is two times the minimum value of the gains and losses (G_j, L_j) for each LULC class.
Gain-to-Loss (G/L)	The tendency to experience more gain or loss is the division of gain (G_j) and loss (L_j). Values above 1 suggest larger tendency to gain than loss.
Loss-to-persistence (L_p)	The tendency to experience more loss or persistence is the division of loss (L_j) and persistency (P_{jj}). Values above 1 suggest larger tendency to loss than to persistence.
Gain-to-persistence (G_p)	The tendency to experience more gain or persistence is the division of gain (G_j) and persistency (P_{jj}). Values above 1 suggest larger tendency to gain than to persistence

The second category includes the measurement of systematic and random LULC changes by using the off-diagonal entries of the transition matrix (which indicates the transitions from one LULC class to the other) for the significant inter-class transitions detection. To detect if the inter-class transition is systematic or random, the expected LULC value was compared with the observed LULC value in the matrix. Following the statistical definitions outlined by Pontius et al. (2008), a random transition indicates that the gain from other LULC classes was equivalent to the size of those other losing classes, or the loss to other LULC classes was equivalent to the size of those other gaining classes. Therefore, values close to zero or at zero indicate a partly systematic or random transition, respectively, while numbers far from zero indicate a systematic inter-class transition (Pontius Jr et al., 2004).

As such, the measurement of the systematic and random LULC transitions will result in three values in terms of gain (G_{ij}) and in three values in terms of loss (L_{ij}). Firstly, we compute the expected LULC gains percentage if the gain in each class were to occur randomly:

$$G_{ij} = (P_{+j} - P_{jj}) \left(\frac{P_{i+}}{\sum_{i=1, i \neq j}^{j} P_{i+}} \right) \tag{6.1}$$

where $(P_{+j} - P_{jj})$ is the total gross gain of LULC class j, and P_{i+} is the size of LULC class i in the initial time (1995). $\sum_{i=1,i\neq j}^{j} P_{i+}$ is the sum of the sizes of all LULC classes, excluding LULC classes of j, in the initial time. Secondly, we compute the expected LULC losses percentage if the loss in each class were to occur randomly:

$$L_{ij} = (P_{i+} - P_{ii})\left(\frac{P_{+j}}{\sum_{j=1,j\neq j}^{j} P_{+j}}\right) \tag{6.2}$$

where $(P_{i+} - P_{ii})$ is the observed total gross loss of LULC class i, and P_{+j} is the size of LULC class j in the final time (2018). $\sum_{j=1,j\neq j}^{j} P_{+j}$ is the sum of the sizes of all LULC classes, excluding LULC classes of j, in the final time.

Third, we compute the differences between the observed LULC percentage and the expected LULC percentage under a random process of gain:

$$P_{ij} - G_{ij} \tag{6.3}$$

where P_{ij} is the observed LULC, and G_{ij} is the expected LULC if the gain in each class were to occur randomly. Fourth, we compute the differences between the observed LULC percentage and the expected LULC percentage under a random process of loss:

$$P_{ij} - L_{ij} \tag{6.4}$$

where P_{ij} is the observed LULC, and L_{ij} is the expected LULC if the loss in each class were to occur randomly.

6.2.5. The Discrete-Time Markov Chain (DTMC) Model

6.2.5.1 States Transition Probability of the Markov Chain

To quantify the states of conversion between the LULC classes and simulate and analyze the future LULC changes by 2040, we used the Discrete-Time Markov Chain (DTMC) model. Specifically, the DTMC model use matrices to model the LULC transition probabilities among a set of discrete states. Therefore, a DTMC is characterized by an array of random variables $X_1, X_2, \dots, X_n$ showing a memoryless property (i.e., a Markov property). The memoryless property postulates that the transition probabilities are independent of the chain state before its current state. Thus, the X_{n+1} state distribution relies only upon the current state X_n regardless of the $X_{n-1}, X_{n-2}, \dots, X_{n1}$ states.

$$P_r(X_{n+1} = x_{n+1} | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P_r(X_{n+1} = x_{n+1} | X_n = x_n) \tag{6.5}$$

Strictly, a Markov chain is a probabilistic automaton, where X_n has a set of possible states $S = \{S_1, S_2, \dots, S_r\}$ entitled as the "state space of the chain." The procedure begins in one of S states and changes consecutively from one to another. Each change is named either step or transition, and the transition probability p_{ij} is set as the probability of change from state s_i to state s_j in just one step.

$$p_{ij} = P_r(X_1 = s_j | X_0 = s_i) \tag{6.6}$$

whereas the probability to change from state s_i to state s_j in n steps is given by $p_{ij}^{(n)} = P_r(X_n = s_j | X_0 = s_i)$.

If Equation (7) proves to be true, the DTMC is time-homogeneous, which indicates that the original transition probabilities remain unchanged through time.

$$P_r\left(X_{n+1} = s_j \middle| X_n = s_i\right) = P_r\left(X_n = s_j \middle| X_{n-1} = s_i\right) \tag{6.7}$$

For $k > 0$, if $p_{ij} = P_r\left(X_{k+1} = s_j \middle| X_k = s_i\right) \wedge p_{ij}^{(n)} = P_r\left(X_{n+k} = s_j \middle| X_k = s_i\right)$, then the Markov chain is said to be time-homogeneous. The transition probability can be represented as a transition probabilities matrix (or simply transition matrix) $P = \left(p_{ij}\right)_{i,j}$, where each element (i, j) denotes the position of p_{ij}.

$$P = \begin{bmatrix} p_{ii} & \cdots & p_{ij} \\ \vdots & \ddots & \vdots \\ p_{ji} & \cdots & p_{jj} \end{bmatrix} \tag{6.8}$$

The dispersion through the states can be transcribed as a stochastic $(\sum_i x_i = 1, x_i \geq 0)$ row vector $\boldsymbol{x}$. Hence, $x^{(1)}$ relates to$x^{(0)}$ as $x^{(1)} = x^{(0)}P$ and, recursively, we have $x^{(2)} = x^{(0)}P^2 \wedge x^{(n)} = x^{(0)}P^n$.

6.2.5.2. Classification States and Simulating Future States

The states can be classified by means of their periodicity. A state that is not absorbing is a transient one. A state s_i is said to be transient if there is a state s_k that is attainable from s_j, but j not being attainable from s_k. In other words, if s_j is transient, there is no guarantee that, after leaving s_j, the system will return there. Suppose that, for a chain with initial state x, the number of periods required to return to that state is given by $T^{x \to x}$. If $P(T^{x \to x} < +\infty) = 1 \equiv P(T^{x \to x} = +\infty) > 0$, the x state is transient.

A state s_j is said to be recurrent if it is not transient, i.e., if it is always possible to return to s_j. As such,

$$\begin{matrix} state\ x\ is\ recurrent\ if \\ P(T^{x \to x} < +\infty) = 1\ or\ \ in\ addition\ is \\ P(T^{x \to x} = +\infty) = 0 \end{matrix} \begin{cases} null\ recurrent\ if\ P(T^{x \to x}) = +\infty \\ positive\ recurrent\ if\ P(T^{x \to x}) < \infty \\ absorbing\ if\ P(T^{x \to x} = 1) = 1 \end{cases} \tag{6.9}$$

The time necessary to reach a specific state can also be computed. The so-called hitting time, i.e., the first passage $\left(s_i \to s_j\right)$, is given by the number of steps $\left(T_{ij}\right)$ that the chain requires until it hits the state s_j for the first time, considering that $X_0 = s_i$.

$$h_{ij}^{(n)} = P_r\left(T_{ij} = n\right) = P_r\left(X_n = s_j, X_{n-1} \neq s_j, \ldots, X_1 \neq s_j \middle| X_0 = s_i\right) \tag{6.10}$$

where $h_{ij}^{(n)} = \sum_{k \in S - \{s_j\}} p_{ik} h_{kj}^{(n-1)}$ and $h_{ij}^{(n)} = p_{ij}$.

A measure interconnected with (h) h that is frequently used is the mean first passage time, i.e., the expected hitting time. In an ergodic Markov chain, the mean first passage time from s_i to s_j is the estimated number of steps necessary to go from the initial state s_i to s_j for the first time.

This measure is the (h) average value, explicitly $\bar{h}_{ij} = \sum_{n=1\ldots\infty} nh_{ij}^{(n)}$. When delineating the first passage time, one can assume that $s_i = s_{ij}$ and thus obtain the first recurrence time $T_i = inf\{n \geq 1\colon X_n = s_i | X_0 = s_i\}$.

Therefore, having the recurrent states, we calculated the mean recurrence time, i.e., the expected number of steps to return to the initial state:

$$r_i = \sum_{k=1}^{\infty} k \cdot P(T_i = k) \tag{6.11}$$

Finally, as the period under analysis presents different cycles, that is more dynamic cycles (1995–2007 and 2010–2015) interspersed with less dynamic ones (2007–2010 and 2015–2018), we chose to perform future simulations based on trends of the whole set (1995–2018). Thus, based on the year 2018 and using the trends recorded over 23 years, we created a scenario of LULC for 2040 (instead of 2041 for rounding purposes).

6.2.5.3 Uncertainty of Future States

Uncertainty can then be characterized by the probability vectors that are produced as a sub-product of simulation, made available by most stochastic modeling procedures (Shi et al., 2002). These probabilities provide useful information about the quality of the resulting classification in terms of the uncertainties involved. To fully explore the information of the probability vector, additional measures of uncertainty are needed (der Wel et al., 1998). The greater the uncertainty, the lower the probability that a pixel will be associated with a given class, so uncertainty is the complement of the probability. Formally, if we considerer P_{ij}, the probability that a given pixel j is associated with class i, the uncertainty can be described by $U_{ij} = 1 - P_{ij}$.

However, normally the true class of the pixel is not known and, consequently, the amount of information required to indicate the pixel class is also unknown. The pixel entropy is therefore defined as the expected information index of a piece of information that reveals your true class. For this purpose, the entropy measure combines the uncertainties of the various pixel classes, weighing them by their probabilities. In this way, the global entropy can be estimated from the normalized values of probabilities by class, using the expression $E_{ij} = -\sum P_{ij} * \log\ _2 P_{ij}$.

6.3. Results

6.3.1. LULC Spatiotemporal Evolution and Distribution

The spatiotemporal analysis results showed that the Beja district has undergone major changes between 1995 and 2018 (Table 6.3). At first sight, the predominance of agricultural and forestry areas over the entire study period was observed. However, the occupancy tendency of both classes showed large fluctuations. Examining the results, between 1995 and 2010, there was a loss of almost 6.6% of the agricultural area, but this mega class continued to occupy the largest portion of the territory (in 2010 about 62.88%). Subsequent years (2015–2018) saw an increase in agricultural area (+1.08%).

Table 6.3. LULC from 1995 to 2018 (in %)

LULC Class	1995	2007	2010	2015	2018
Artificial surfaces	0.79	1.04	1.09	1.21	1.24
Agricultural areas	69.43	63.30	62.88	63.12	63.96
Arable land	26.64	23.69	22.83	22.77	20.70
Permanent crops	7.43	8.80	9.87	10.25	12.56
Pastures	14.13	10.63	10.11	10.68	11.50
Heterogeneous areas	21.23	20.18	20.07	19.42	19.20
Forest	28.35	33.68	33.94	33.53	32.66
Water bodies	1.44	1.99	2.09	2.14	2.14

Indeed, except for the permanent crops, there has been a significant decrease in remaining croplands. Particularly, arable land, the cropland with the most significant predominance in the Beja district (in 2018 about 20.7%), presented a decrease of almost 6% between 1995 and 2018, while heterogeneous areas and pastures decreased by about 2% and 2.6%, respectively. Heterogeneous areas are mainly concentrated near forestland, and the pastures class almost disappears in the central region of the Beja district. Permanent crops increased by about 5%, particularly in the northern part of the district, where the patch correspondent to this class is noticeable (Figure 6.2).

By contrast, the forestland presented an area increase of almost 5.6% between 1995 and 2010 (in 2010 about 34%), but in subsequent years (2010–2018) there was a loss of almost 1.3% (in 2018 about 32.7%). It is noticeable that the green area corresponds to the forestland class in the SSW of the study area, probably due to its being closer to the Atlantic Ocean and to its orography (plateaus). However, the increase in forest was especially near the wetlands, untangling the southeast part of the district. As for the artificial surfaces, although substantially reduced in this region, they present a linear growth increment between the period under analysis (close to 0.5%), especially in the north of the district near the Beja airport. The same occurred with the wetlands class with an overall increase of 0.7%, with some visible expression in the northeast part of the district (Figure 6.2).

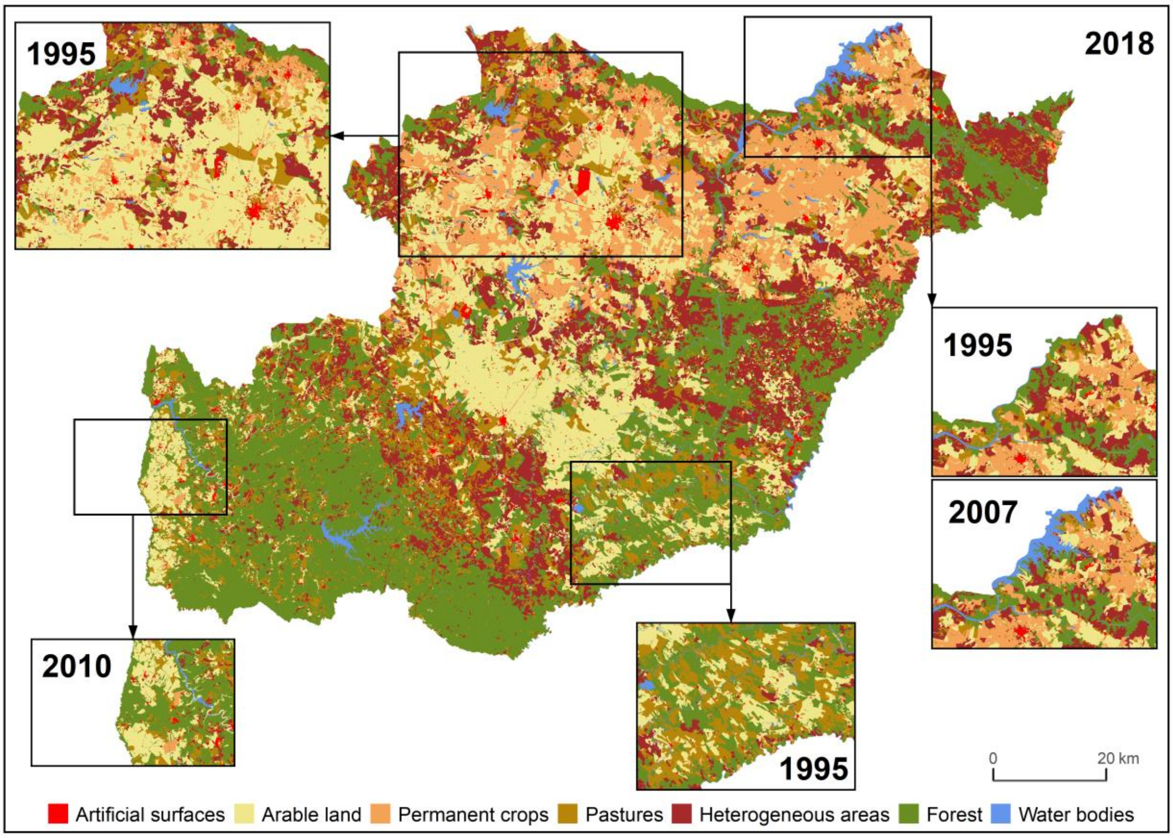


Figure 6.2. LULC map of the Beja district for 2018, and the detailed changes detected between 1995 and 2018

6.3.2. Main LULC Transitions by Period

The analysis of the inter-class transitions by period is present using the cross-tabulation matrices in graphical form for the four periods under analysis (Figure 6.3). Specifically, during the first period (1995–2007), almost 20% of the LULC changed. The results revealed a clear tendency for each type of agricultural use to transition to another agricultural use or to transition to forest. Additionally, the forest class and the pastures class were the two that presented the highest LULC transition values. Indeed, the first period experienced a significant decrease in croplands (−6.3%), mostly due to forestland, with transitions from the pastures (3.8%), heterogeneous areas (1.8%), and arable land (1.7%) classes. However, about 1.8% of forestland also transitioned to agricultural areas. By contrast, arable land and pastures exchange between themselves.

As for the second period (2007–2010), about 3% of the LULC changed. This period experienced the same trend: a decrease in croplands, albeit much less significant (−0.42%). The arable land, pastures, and permanent crops classes transitioned between each other, the highest area transitioning from arable land to permanent crops (0.8%).

.

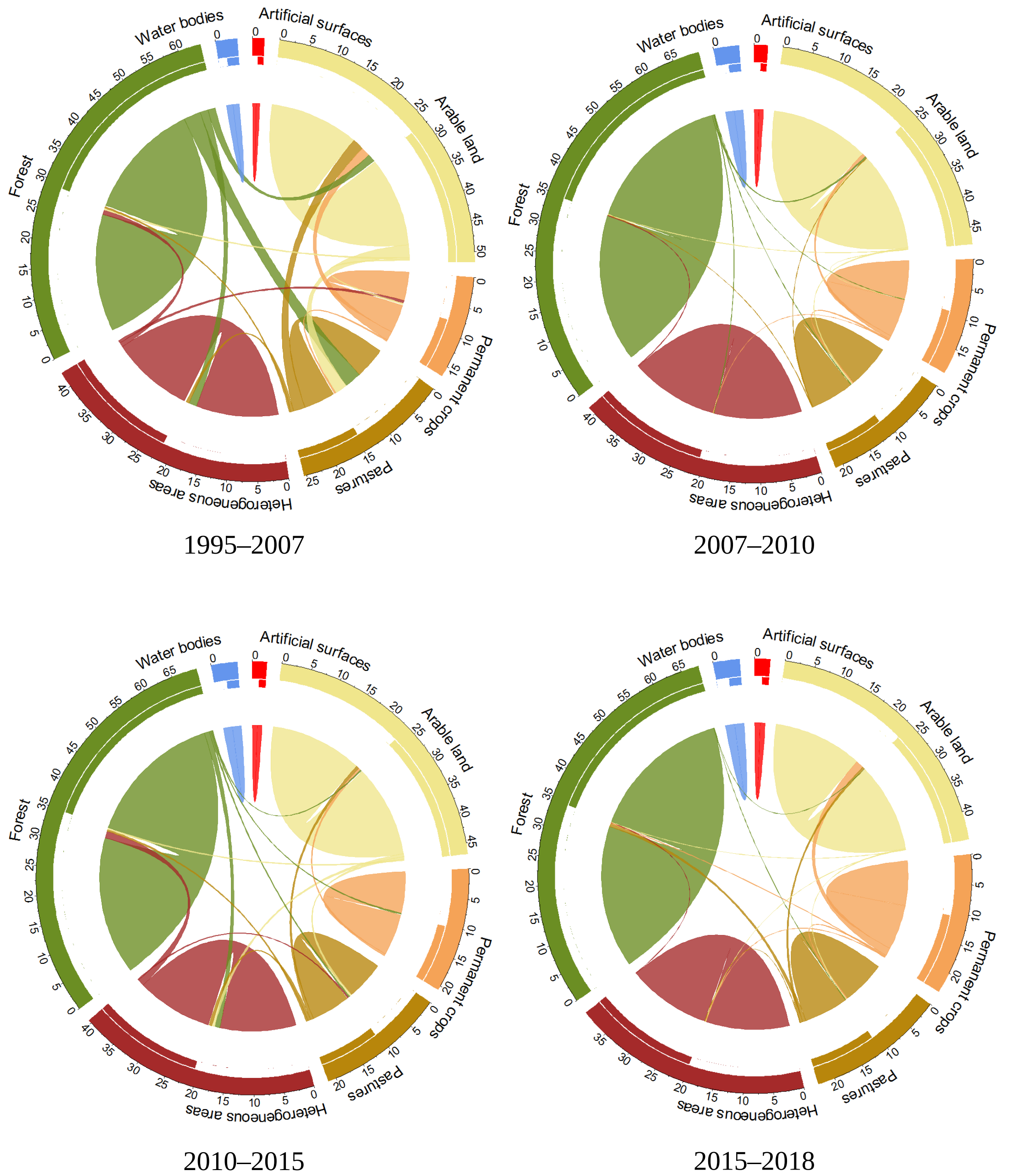


Figure 6.3. LULC transition for each period under analysis. The outer circle is the amount scale (%); the inner circle represents the quantity of the LULC class that remains unchanged between the two dates; the direction of the flow is encoded both by the origin color and by the gap between link and circle segment

The subsequent period (2010–2015) experienced about 8.5% of LULC changes. However, this period counters the decreasing trend in cropland, albeit modestly, as croplands increased but only by about 0.24%. Thus, a transition tendency similar to the first period can be observed, since the major recorded transitions were between agricultural areas and forestland. Accordingly, the heterogeneous areas and forestland classes transitioned between each other, with heterogeneous areas transitioning to forest (1.18%) and forest transitioning to heterogeneous areas (1.55%).

During the last period (2015–2018), almost 4.7% of the LULC changed. This period experienced the highest gain in agricultural areas (+0.84%). Once again, arable land transitioned mainly to pastures (0.7%), and pastures transitioned mainly to arable land (0.5%), while heterogeneous areas transitioned to arable land (0.7%) and to pastures (0.7%). Indeed, the most substantial transition recorded was from arable land to permanent crops (1.8%), followed by the transition from forest to pastures (0.7%), and from arable land to pastures (0.6%). Unlike the past two periods, we verified for the first time more losses than gains for the forest class.

6.3.3. LULC Change Detection

The in-depth analysis of the overall LULC change detection shows that, for the overall period (1995–2018), about 72.1% of the LULC persisted, while 27.9% exhibited transitions. Briefly, the highest LULC transitions were experienced from arable land to permanent crops (4.96%), followed by the transition from pastures to forest (3.48%) (Table 6.4). In addition, the overall period transitions highlighted that forest experienced high gains from agricultural classes, while both artificial surfaces and water bodies showed a minor transition to other LULC classes.

Table 6.4. LULC transition matrix 1995–2018 (in %) *

		2018						
	LULC Class	Artificial Surfaces	Arable Land	Permanent Crops	Pastures	Heterog. Areas	Forest	Water Bodies
1995	Artificial surfaces	0.76	0.00	0.01	0.01	0.00	0.00	0.00
	Arable land	0.16	16.47	4.96	2.92	0.17	1.78	0.18
	Permanent crops	0.03	0.53	6.14	0.12	0.42	0.17	0.03
	Pastures	0.13	2.70	0.92	6.52	0.26	3.48	0.12
	Heterog. areas	0.10	0.64	0.40	1.18	16.23	2.53	0.16
	Forest	0.05	0.37	0.11	0.76	2.13	24.66	0.27
	Water bodies	0.01	0.00	0.01	0.01	0.00	0.04	1.36

* The highest LULC transitions are highlighted in bold.

From Table 6.5, it can be verified that, during the overall period, both the artificial surfaces and water bodies classes experienced a lower percentage of persistence (0.76% and 1.36%, respectively). These results are mostly due to their lower percentage of territory occupancy within the period analysis. Thus, artificial surfaces occupied about 1.2% of the district area, meaning almost 0.5% of gains and only 0.03% of losses. Therefore, the changes in the artificial surfaces class is mostly net change (about 0.45%). Similarly, the water bodies class experienced almost 0.8% of gains against 0.1% of losses between 1995 and 2018. Thus, most of the change in the water bodies class is a net change (0.7%). Hence, artificial surfaces and water bodies experienced a gain–loss ratio of 18.3 and 10.5, respectively; however, according to the loss-to-persistence and gain-to-persistence ratios, both classes presented a greater tendency to persist than to gain or lose. Figure 6.4 presents the net changes for both the artificial surfaces and water bodies classes. While artificial surfaces gains are highly dispersed across the district and usually founded near spots with the same LULC, water bodies gains are not clustered predominantly near to each other. Due to the minimal losses during 1995–2018, it is difficult to identify the dispersion of losses across the district.

Table 6.5. Summary of LULC changes (1995–2018) (in %)

LULC Class	P_j	G_j	L_j	C_j	S_j	D_j	G/L	L_p	G_p
Artificial surfaces	0.76	0.48	0.03	0.50	0.05	0.45	18.32	0.03	0.63
Arable land	16.47	4.23	10.16	14.39	8.46	5.93	0.42	0.62	0.26
Permanent crops	6.14	6.42	1.30	7.71	2.59	5.12	4.96	0.21	1.05
Pastures	6.52	4.99	7.61	12.60	9.97	2.62	0.66	1.17	0.77
Heterogeneous areas	16.23	2.97	5.00	7.97	5.94	2.03	0.59	0.31	0.18
Forest	24.66	8.00	3.69	11.69	7.38	4.31	2.17	0.15	0.32
Water bodies	1.36	0.78	0.07	0.85	0.15	0.70	10.47	0.05	0.57
Total	72.14	27.86	27.86	27.86	17.27	10.59			

Note: P_j = persistence, G_j = gain, L_j = loss, C_j = total change, S_j = Swap, D_j = net change, G/L = gain-to-loss ratio, L_p = loss-to-persistence ratio, G_p = gain-to-persistence ratio

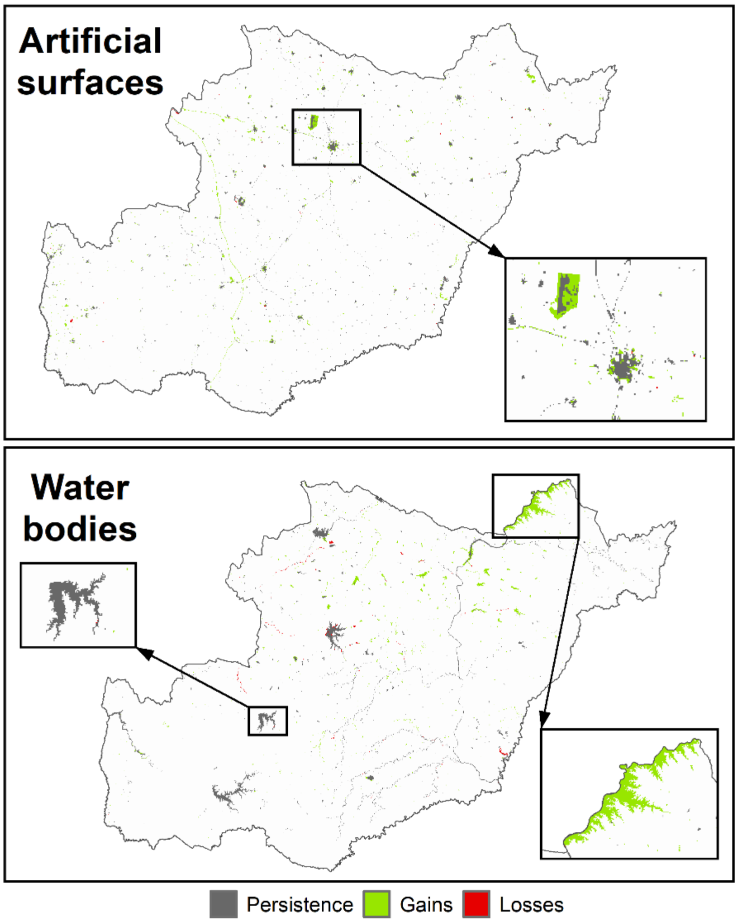


Figure 6.4. Changes in the artificial surfaces and water bodies classes (1995–2018)

In contrast, forest, arable land, and heterogeneous areas presented a high percentage of persistence (24.66%, 16.47%, and 16.23%, respectively), while pastures and permanent crops presented a relatively low percentage of persistence (6.52% and 6.14%, respectively). These results are mostly due to each LULC class percentage of territory occupancy during all the years under study. Nevertheless, permanent crops and forest classes presented more gains than losses during the overall period. The gain–loss ratio indicates that permanent crops presented 4.9 times more gain than loss, while forest presented a ratio of 2.2 times more. However, the influence of net changes was found to be the main driver in permanent crops (66% of the total change for permanent crops), while swapping was the main driver in the forest class (63% of the total change for forest).

However, according to the loss-to-persistence and gain-to-persistence ratios, the forest class experienced a greater tendency to persist than to gain or lose, while the permanent crops class presented a greater tendency to persist than to lose and a greater tendency to gain than to persist. Figure 6.5 presents the net changes for both the permanent crops and forest classes. The gains of forestland are dispersedly distributed across the Beja district and are usually found near spots with the same LULC. Comparatively, permanent crops gains are primarily distributed in the north and east. Gain spots are grouped mainly near to each other, while losses are minimal and show no distribution pattern.

The remaining agricultural classes presented more losses than gains, with a loss–gain ratio of 2.4, 1.5, and 1.7 for arable land, pastures, and heterogeneous areas, respectively. These three agricultural classes indicate both swap and net changes between 1995 and 2018. However, the influence of swapping was found to be the main driver of the total change, which was a change by 59%, 79%, and 75% for arable land, pastures, and heterogeneous areas, respectively. While both arable land and heterogeneous areas presented a greater tendency to persist than to gain or lose, the pastures class presented a greater tendency to lose than to persist and a greater tendency to persist than to gain.

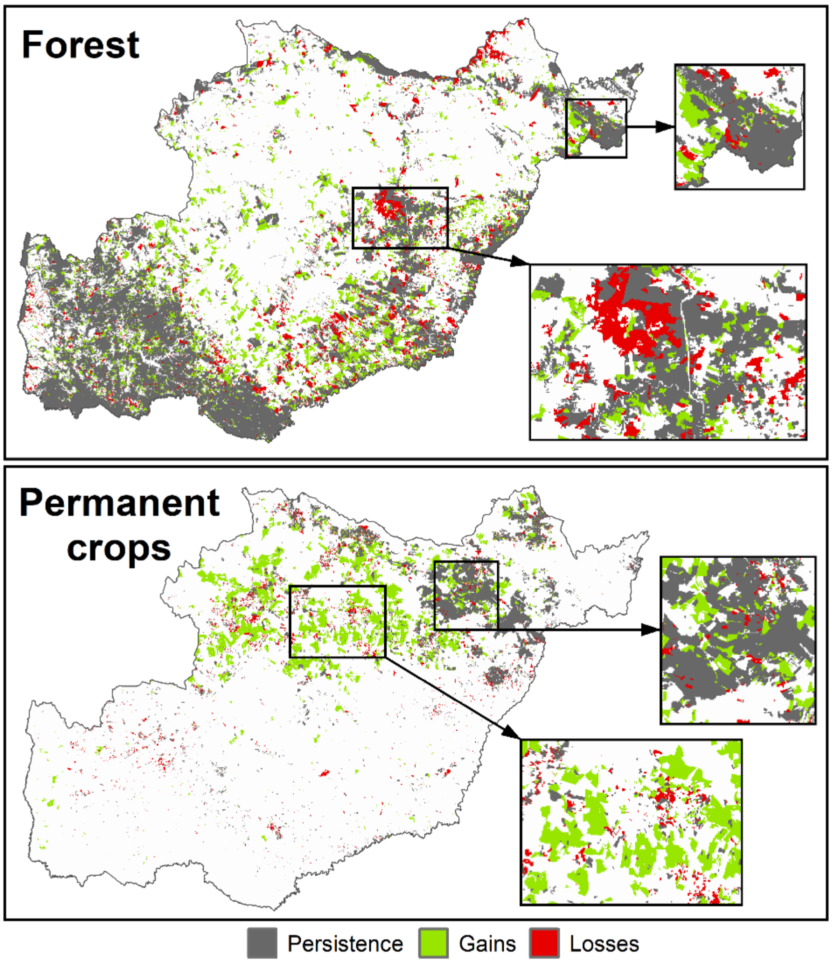


Figure 6.5. Changes in the forest and permanent crops classes (1995–2018)

Figure 6.6 presents the net changes for arable land, pastures, and heterogeneous areas. The three classes present a very distinct change pattern. Arable land losses are mainly consolidated in the north and east of Beja, while the gains are dispersed in the south of the district, predominantly near spots with the same LULC. As for pastures, the gains are spread over the entire territory, while the losses can be found in the south and close to the east. Lastly, the heterogeneous areas show large gains in the center of the territory, while losses are distributed mainly close to where there is forestland.

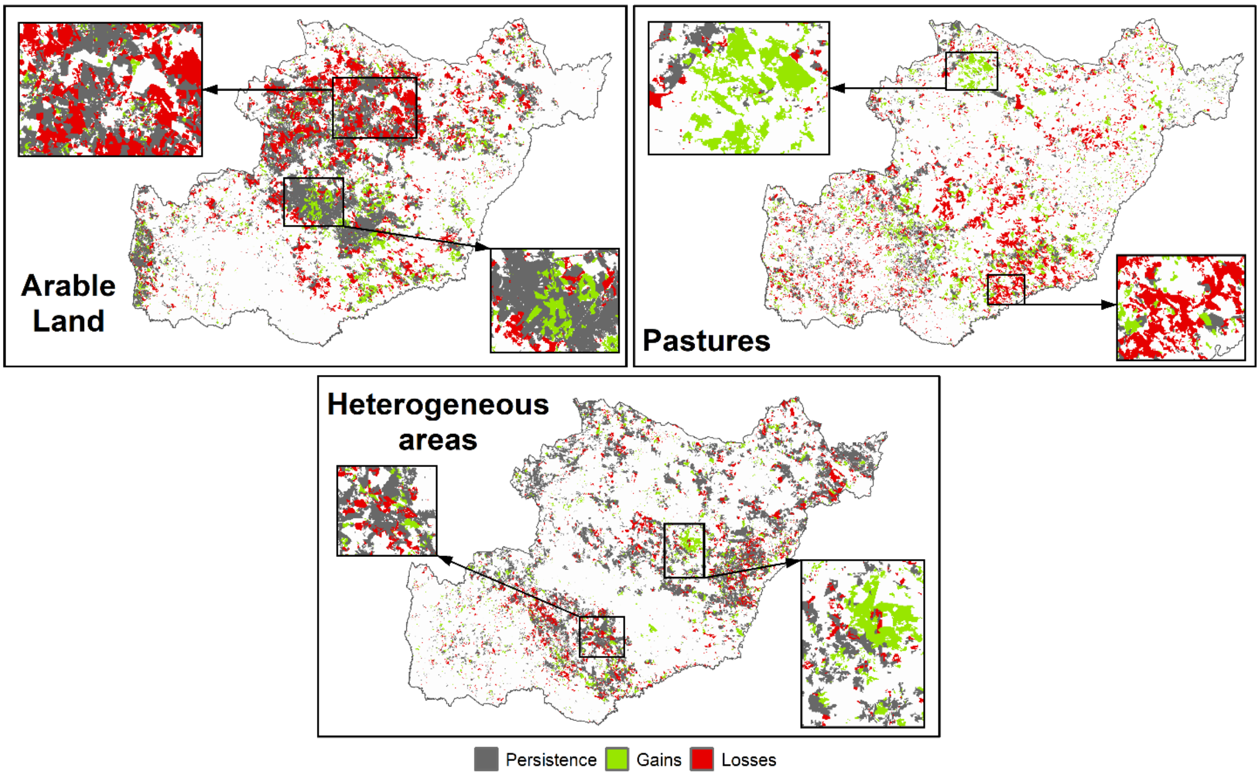


Figure 6.6. Changes in arable land, pastures, and heterogeneous areas classes (1995–2018)

6.3.4. Detection of Random and Systematic LULC Transitions

The random and systematic LULC transitions analysis showed that not all the inter-class transitions during the period under analysis were systematic (dominant) processes within a pattern of LULC changes. Based on Tables 6.6 and 6.7, we are able to determine what the dominant transitions in each LULC class would be if the gain or loss in each class were to occur randomly. Therefore, the arable land and permanent crops classes experienced the strongest dominant signal of change. While arable land systematically transitioned to permanent crops, permanent crops were systematically gaining from arable land simultaneously (3.12%), and arable land was also systematically losing to permanent crops (3.35%). However, permanent crops systematically avoided losses to arable land.

Likewise, while arable land systematically transitioned to pastures, the pastures class was systematically gaining from arable land simultaneously (1.37%), and arable land was also systematically losing to pastures (1.44%). By contrast, while pastures were systematically transitioning to arable land, arable land was systematically gaining from pastures simultaneously (1.88%), and pastures were also systematically losing to arable land (0.92%). Thus, the arable land and pastures classes systematically transitioned between each other. Analyzing the situation more closely, based on Table 6, it should also be noted that permanent crops and arable land systematically avoided gaining from forest (−1.85% and −1.27%, respectively), and permanent crops or arable land avoided gaining from heterogeneous areas (−1.07% and −0.59%, respectively). In Table 7, it can be seen that arable land systematically avoided losing to heterogeneous areas (−2.29%) or to forest (−2.41%), and pastures avoided losing to heterogeneous areas (−1.39%).

Even though forest systematically gained from pastures (1.91%), there was a weak signal of pastures losing to forest (0.68). The analysis of the results in Table 6 also recognized that pastures systematically avoided gaining from forest (−0.89%), while Table 7 indicated that pastures avoided losing to heterogeneous areas (−1.39%). On the one hand, heterogeneous areas systematically gained from forests (1.06%), while heterogeneous areas systematically avoid losing to forest (0.51%). However, forest was systematically losing to heterogeneous areas (1.08%). From Table 6, it also appeared that forest and heterogeneous area classes systematically avoided gaining from arable land (−1.20% and −0.84%, respectively), and forest avoided gaining from permanent crops (−0.66%).

Table 6.6. Inter-class LULC transitions gains

	2018						
1995	Artificial Surfaces	Arable Land	Permanent Crops	Pastures	Heterog. Areas	Forest	Water Bodies
	(i) Expected LULC if the process of change in each class were to occur randomly (%)						
Artificial surfaces	0.76	0.05	0.05	0.05	0.03	0.09	0.01
Arable land	0.13	16.47	1.85	1.55	1.00	2.98	0.21
Permanent crops	0.04	0.43	6.14	0.43	0.28	0.83	0.06
Pastures	0.07	0.81	0.98	6.52	0.53	1.58	0.11
Heterog. areas	0.10	1.22	1.47	1.23	16.23	2.37	0.17
Forest	0.14	1.63	1.97	1.65	1.07	24.66	0.22
Water bodies	0.01	0.08	0.10	0.08	0.05	0.16	1.36
	(ii) Observed LULC minus the proportion expected under the random process (%) *						
Artificial surfaces	0.00	−0.04	−0.05	−0.04	−0.03	−0.09	0.00
Arable land	0.03	0.00	3.12	1.37	−0.84	−1.20	−0.03
Permanent crops	−0.01	0.10	0.00	−0.31	0.14	−0.66	−0.03
Pastures	0.06	1.88	−0.06	0.00	−0.28	1.91	0.01
Heterog. areas	−0.01	−0.59	−1.07	−0.06	0.00	0.16	−0.01
Forest	−0.08	−1.27	−1.85	−0.89	1.06	0.00	0.05
Water bodies	0.00	−0.08	−0.09	−0.08	−0.05	−0.12	0.00

* The magnitude of the value indicates the percent of LULC. The most systematic transitions are highlighted in bold.

Table 6.7. Inter-class LULC transitions losses

	2018						
1995	Artificial Surfaces	Arable Land	Permanent Crops	Pastures	Heterog. Areas	Forest	Water Bodies
	(i) Expected LULC if the process of change in each class were to occur randomly (%)						
Artificial surfaces	0.76	0.01	0.00	0.00	0.01	0.01	0.00
Arable land	0.16	16.47	1.61	1.47	2.46	4.19	0.27
Permanent crops	0.02	0.31	6.14	0.17	0.28	0.48	0.03
Pastures	0.11	1.78	1.08	6.52	1.65	2.81	0.18
Heterog. areas	0.08	1.28	0.78	0.71	16.23	2.02	0.13
Forest	0.07	1.13	0.69	0.63	1.05	24.66	0.12
Water bodies	0.00	0.02	0.01	0.01	0.01	0.02	1.36
	(ii) Observed LULC minus the proportion expected under the random process (%) *						
Artificial surfaces	0.00	0.00	0.00	0.01	0.00	−0.01	0.00
Arable land	0.00	0.00	3.35	1.44	−2.29	−2.41	−0.09
Permanent crops	0.01	0.22	0.00	−0.05	0.13	−0.31	0.00
Pastures	0.02	0.92	−0.16	0.00	−1.39	0.68	−0.06
Heterog. areas	0.02	−0.64	−0.37	0.46	0.00	0.51	0.03
Forest	−0.02	−0.77	−0.57	0.13	1.08	0.00	0.16
Water bodies	0.01	−0.01	0.00	0.00	−0.01	0.02	0.00

* The magnitude of the value indicates the percent of LULC. The most systematic transitions are highlighted in bold.

6.3.5. States Transition Probability of the Markov Chain

Table 6.8 presents the temporal transition probability matrix for the overall period (1995–2018), showing the probability that each LULC type transitions to another (the off-diagonal entries indicate the transitions probability from one LULC class to another). As expected, there is a small probability that artificial surfaces will transition to another use (3%), and the same is observed with the water bodies (5%). For forest areas, there is a probability of an approximately 13% change in use, specifically, approximately 8% towards heterogeneous areas.

With regard to croplands, there is a clear tendency for each type of agricultural use to transition to another agricultural use or to forest. For example, pastures and arable land are the classes that are most likely to transition. Pastures have a 25% probability of transition to forest and a 19% probability of transition to arable land. Arable land presents a 19% probability of transition to permanent crops and an 11% probability of transition to pastures. Heterogeneous areas have a 12% probability of transition to forest. From all croplands, the low probability of permanent crops to transition (about 17%) and the low probability of heterogeneous areas to transition (about 24%) to other LULC classes are noteworthy.

Table 6.8. Markov chain probability transition matrix (1995–2018) (in %)

		2018						
	LULC Class	Artificial Surfaces	Arable Land	Permanent Crops	Pastur es	Heterog. Areas	Fore st	Water Bodies
1995	Artificial surfaces	0.97	0.00	0.01	0.01	0.00	0.00	0.00
	Arable land	0.01	0.62	0.19	0.11	0.01	0.07	0.01
	Permane nt crops	0.00	0.07	0.83	0.02	0.06	0.02	0.00
	Pastures	0.01	0.19	0.07	0.46	0.02	0.25	0.01
	Heterog. areas	0.00	0.03	0.02	0.06	0.76	0.12	0.01
	Forest	0.00	0.01	0.00	0.03	0.08	0.87	0.01
	Water bodies	0.01	0.00	0.01	0.00	0.00	0.03	0.95

6.3.6. Classification States

The analysis in Chapter 6.3.5 showed that each state has a positive probability to achieve any other state. Therefore, all states of the DTMC communicate, and the matrix can be cataloged as irreducible. As there are no transition or absorbing states, the matrix is also recurrent; i.e., all states are recurrent. Finally, as we always obtain $k = 1$, the recurrent states are all aperiodic. Together, since all states are recurrent and aperiodic, communicating the matrix can be cataloged as ergodic. Thus, having a transition probability matrix, the mean first passage time was evaluated. Table 6.9 gives us the mean first passage time between 1995 and 2018. It is clear that a high number of steps are expected to be necessary for transition to artificial surfaces and water bodies, suggesting that both of these LULC classes are almost stable. Faster transitions are, in all cases, towards a forest state.

Having the recurrent states, we calculated the mean recurrence time, i.e., the expected number of steps to return to the initial state. This measure is narrowly related to the mean first passage and states. Even in our case, a returning probability of 1 does not signify that the predictable return time is fixed. The results showed that all the LULC classes represent positive recurrent states (Table 6.10). Thus, if the system initiates in state s_i, in determining the number of steps necessary to return to s_i for the first time, it is obvious that it will eventually return to the initial sate. This is because, in the first step, the system either stays at s_i or changes to other state s_j, and from that state s_j, one will ultimately reach s_i since the chain is ergodic.

Table 6.9. Mean first passage time

	LULC Class	2018 Artificial Surfaces	Arable Land	Permanent Crops	Pastures	Heterog. Areas	Forest	Water Bodies
1995	Artificial surfaces	0.0	61.3	62.0	53.2	55.4	46.1	145.1
	Arable land	232.3	0.0	23.5	26.8	27.0	17.2	134.2
	Permanent crops	233.3	27.9	0.0	33.1	24.2	19.8	135.7
	Pastures	232.1	26.7	34.9	0.0	25.5	12.1	132.6
	Heterog. areas	233.8	36.0	42.1	29.8	0.0	12.7	132.4
	Forest	236.6	40.0	46.6	33.2	20.5	0.0	130.7
	Water bodies	218.5	58.4	58.3	52.7	46.5	31.6	0.0

From Table 6.10, it is noticeable that, in the first period (1995–2005), the LULC classes with the lowest mean recurrence time are forest (7.5) and water bodies (1.4), which denotes the resilience of these two LULC classes. The next period (2005–2010) is similar to the previous one, but we start to see the emergence of permanent crops (9.2). The last two periods emphasize the importance of forest and permanent crops in the study area. However, the roles invert themselves, since in 2015–2018, permanent crops achieved a major role in the landscape of Beja, and this may soon become an absorbing state. As for the overall period from 1995 to 2018, we can confirm the forest resilience with a lower value (2.97), whereas the different changing cycles slightly reduce the importance of permanent crops, especially due to the dynamics between 1995 and 2007.

Table 6.10. Mean recurrence time

LULC Class	1995–2007	2007–2010	2010–2015	2015–2018	1995–2018
Artificial surfaces	22.8	122.0	11.2	76.8	8.62
Arable land	37.7	33.8	5.1	25.1	12.80
Permanent crops	56.4	9.2	6.2	1.5	7.16
Pastures	83.8	99.1	8.7	12.7	17.98
Heterog. areas	21.4	16.5	8.0	12.8	6.74
Forest	7.5	7.8	4.0	9.8	2.97
Water bodies	1.4	1.5	16.1	46.9	7.90

Table 6.11 shows that, on average, and after some time, following the 2015–2018 trend, 67% of the area will be permanent crops. Likewise, considering the 2010–2015 trend, about 20% of the territory will be arable land and about 25% forestland. In the former periods of 1995–2005 and 2005–2010, it could be expected that 72% and 65% of the Beja district area will be water bodies, respectively. Lastly, the analysis of the overall period (1995–2018) allows us to embrace the different dynamics observed, avoiding trends too ephemeral to be considered in the long term and reconfirming forestland as the most likely class to occur.

Table 6.11. Steady sates (%)

LULC Class	1995–2007	2007–2010	2010–2015	2015–2018	1995–2018
Artificial surfaces	0.04	0.01	0.09	0.01	0.11
Arable land	0.03	0.03	0.20	0.04	0.07
Permanent crops	0.02	0.11	0.16	0.67	0.13
Pastures	0.01	0.01	0.11	0.08	0.05
Heterog. areas	0.05	0.06	0.12	0.08	0.14
Forest	0.13	0.13	0.25	0.10	0.33
Water bodies	0.72	0.65	0.06	0.02	0.12

6.3.7. Simulating Future States

As stated before, in our recurrent positive (i.e., having a stationary distribution) and aperiodic matrix, regardless of the initial probabilities, as the time step increases towards infinity, the chain probability distribution converges. In this case, the chain has a limiting distribution, i.e., a stationary distribution. In Figure 6.7, we inspect a random simulation of $p_{ij}^{(n)}$ through time and show that, taking into account the 1995–2015 probability vector (Figure 6.7a), states converge to two vectors. The probability of being in States 1 (artificial surfaces) and 7 (forest) converges to higher values than the others LULC classes. Over the long term (e.g., after 48 time steps), the increasing probability of becoming forest overcomes the decreasing probability of becoming artificial surfaces.

As for Figure 6.7b, where we focus on the most recent trend (2015–2018), the situation is somewhat similar. While the system evolves to infinity, the probability vector converges to a vector assigning most of the probability to State 3 (permanent crops). In this case, the probability of becoming permanent crops overpasses that of becoming artificial surfaces only after 28 time steps. In addition, the decreasing rate of the probability of becoming artificial surfaces is much higher in b) than in a). Lastly, it was confirmed that the empirical probabilities behave in the same way (Figure 6.8).

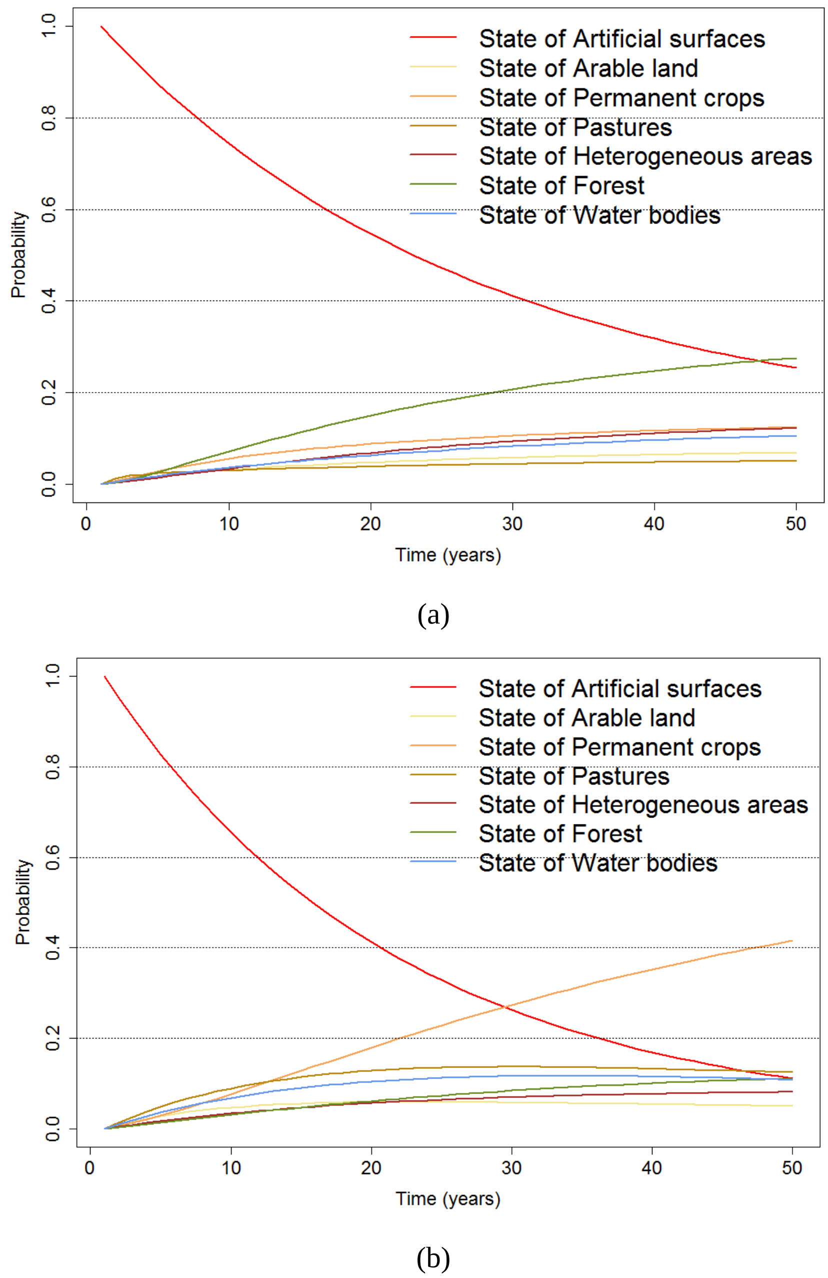
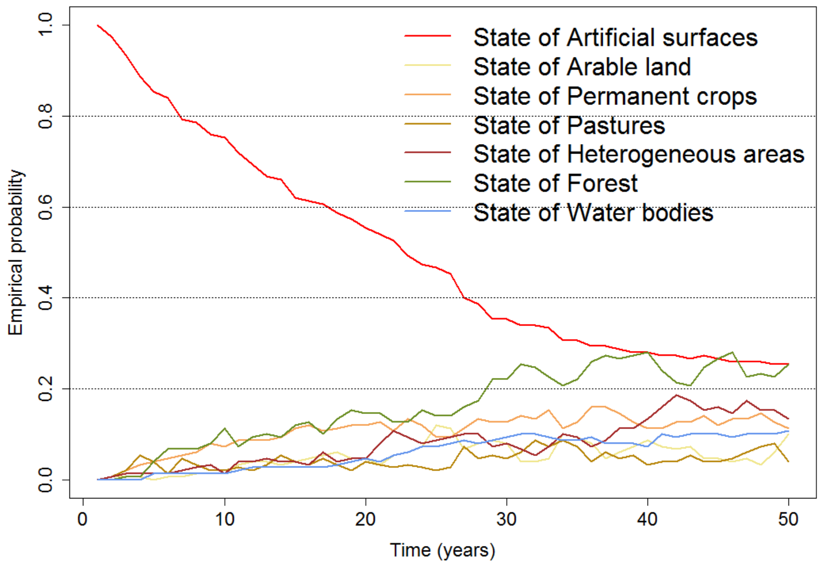
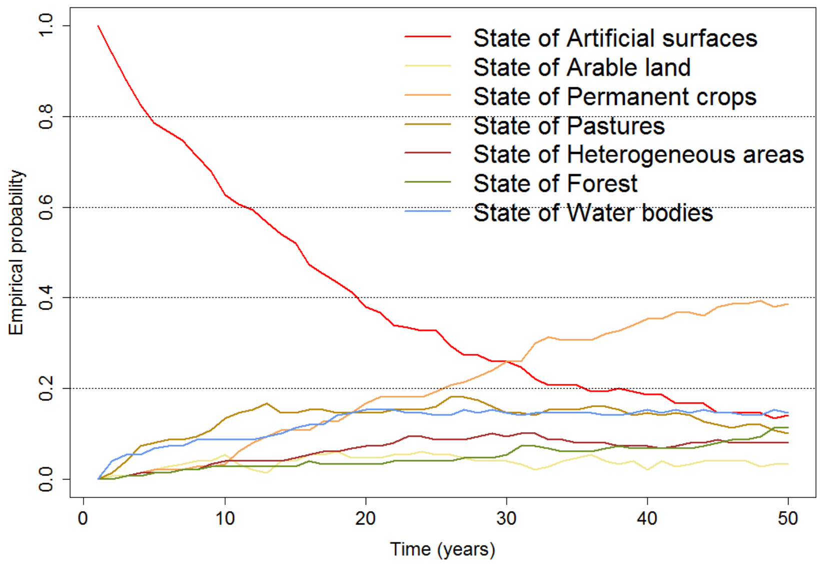


Figure 6.7. LULC class occurrence probability through time: 1995–2015 (a) and 2015–2018 (b)



(a)



(b)

Figure 6.8. LULC class occurrence empirical probability through time: 1995–2018 (a) and 2015–2018 (b)

Therefore, the future state simulation results showed that, in 2040, about 25% of the territory can transition use, and about 21% of that territory corresponds to agricultural land. Accordingly, the major changes concern cropland areas, with a predicted decrease of about −3.67% in 2040 (total 60.3%) (Figure 6.9). However, this situation hides very different behaviors for each cropland class, since arable land (−3.81%), pastures (−1.35%), and heterogeneous areas (−1.02%) diminish. In contrast, and following the trend of past years, permanent crops will see its importance continue to rise (+2.93%). The remaining LULC classes will continue expanding until 2040, where forest class will increase by about 2.60%, artificial surfaces will increase by about

0.40%, and water bodies will increase by about 0.65%. Figure 6.9 shows the spatial distribution of LULC by 2040.

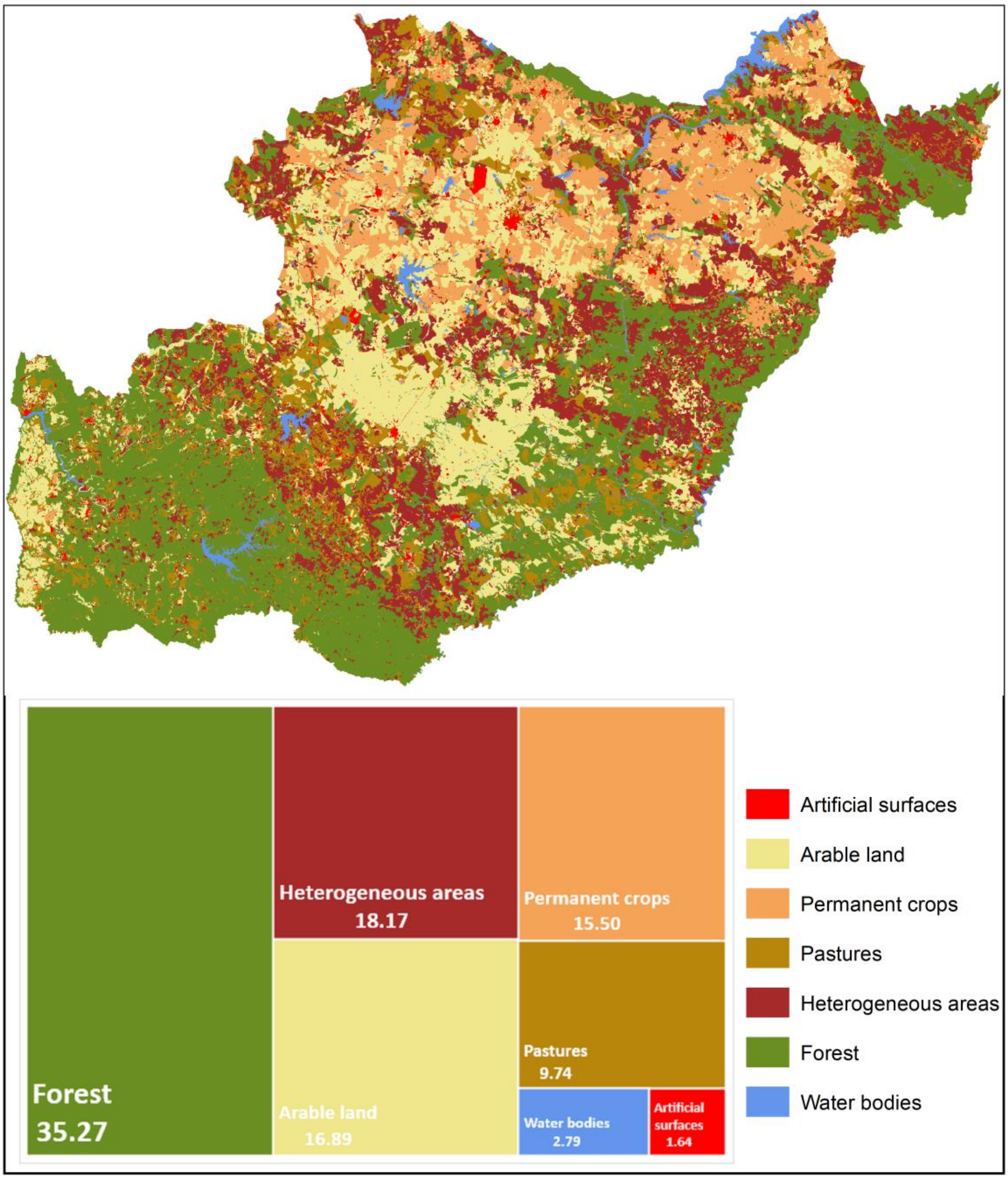
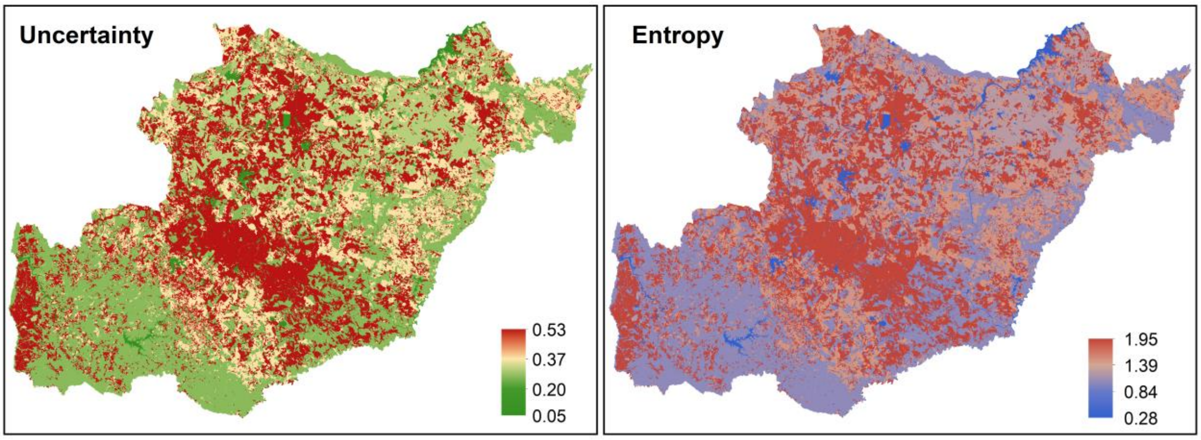


Figure 6.9. LULC simulation for 2040 (values in %)

In Figure 6.10a, we have the uncertainty map for the 2040 LULC simulation, with values ranging from 0.03 to 0.53. Crossing these values with LULC classes shows us where the model performs better. Thus, we have two classes that are very resilient and are therefore modeled with a low degree of uncertainty: artificial surfaces and water bodies, both with an average uncertainty of 0.03. Next, also with low values, are the two most dynamic and important classes of the study area, namely, forest (0.13) and permanent crops (0.17). Finally, three classes remain, which, due to their complex structure and overlapping, achieve the worst values: heterogeneous areas (0.23), arable land (0.38), and pastures (0.53). In Figure 6.10b, we have a spatial notion of the entropy distribution with values oscillating between 0.28 and 1.95. Spatially, both calculations, uncertainty and entropy, show a high rate of agreement.



(a) (b)

Figure 6.10. Uncertainty (a) and entropy (b) of the 2040 LULC map simulation

6.4. Discussion

In recent decades, the abandonment of agricultural land has been an international trend. In Portugal, this has occurred, confirming marked changes in the agricultural landscape (Allen et al., 2018; Feranec et al., 2010). Therefore, in this study, we analyzed the spatiotemporal changes that occurred in the largest district of Portugal, Beja, which has an agricultural economic background, and provided a trend projection for the year of 2040. Accordingly, the monitoring of the dynamics of LULC at the regional level revealed noteworthy LULC dynamics intrinsically associated with distinct agricultural practices of food production and distribution.

6.4.1. About the LULC Changes Detection

The degree of transition between the LULC classes in the Beja district was significantly different during each period under analysis. The following findings are worth noting:

(i) In 23 years, about 28% of the Beja district experienced dozens of inter-class transitions, but the most significant changes were in the croplands. Nevertheless, croplands remain the predominant LULC class in more than half of the Beja district (in 2018 about 64%). Particularly, there was a predominance of arable land that was dispersed in a great part of Beja, despite the significant decrease in this practice between 1995 and 2018 (−5.9%). Despite the decrease, the results confirmed the importance of cereal production in this region, as in the case of wheat (Viana et al., 2019; Viana & Rocha, 2018). Likewise, heterogeneous areas experienced a continuous decline (−2%), but even so this class represents the second largest cropland in the district, which confirms the relevance of the Montado system and the cork oak exploitation in the agroforestry areas of the district.

(ii) Of all the agricultural classes, permanent crops experienced a growing tendency. Thus, this cropland gained prevalence after a certain period and simultaneously experienced some loss (+5.13% between 1995 and 2018). In spite of the reduction in the territory, the substantial increase in permanent crops suggests an emphasis on intensive farming without fallow, proved by the enlargement of the olive grove plantation in this region and the increase in olive oil production (Viana et al., 2019; Viana & Rocha, 2018). These results are consistent with other studies, despite the methodological approach differences (Allen et al., 2018; Roxo & Ventura, 2012; Viana, Girão, et al., 2019).

(iii) As for the overall period, forests experienced more gains than losses (+4.31% between 1995 and 2018). There was a clear tendency of agricultural land loss to the detriment of forests, these results being in line with what is happening on a national level observed in other studies (Feranec et al., 2010; Meneses et al., 2017). Additionally, about 63% of the total changes experienced by forests was attributed to a swap type, which may indicate that the Beja district simultaneously experienced forms of reforestation and deforestation between 1995 and 2018.

(iv) Lastly, artificial surfaces and water bodies experienced more gains than losses and showed a tendency to persist. A main cause for these observed gains is spatial conversions, such as the expansion of the Beja airport area and the construction of new roads in the case of artificial surfaces (Viana et al., 2019), and the Alqueva dam construction in 2002 in the case of the water bodies (Allen et al., 2018). Nonetheless, artificial surfaces only increased about 0.45%, between 1995 and 2018, which indicates that the increase in urban areas was not remarkable in the Beja district, thereby counteracting the Portuguese trend towards an urban area increase (Meneses et al., 2017).

Accordingly, the behavior of the inter-class transitions was significantly different between periods, and explicitly revealed that arable land, pastures, and forest were the most dynamic LULC classes, having experienced both damage and recovery over the years. Hence, the drivers that promote the LULC gains are likely to differ from the drivers that lead to LULC losses (Braimoh, 2006). On the one hand, the main cause for cropland changes could be the competition between LULC classes, since such competition may restrict the use of available land even when this land has suitable determinants for a specific LULC class, which would explain higher observed losses compared with gains (Braimoh, 2006). In addition, the political measures of incentives for the increase in technology, for example, the almost total abandonment of traditional agriculture or the high level of subsidies to agriculture, could be explanatory factors for the agricultural LULC changes observed (Serra et al., 2008). Accordingly, the LULC periodical fluctuations were depolarized by certain drivers, such as political decisions, economic development planning, or even weather conditions, showing the importance of the analysis of the possible drivers of changes in future work (Meneses et al., 2017).

6.4.2. About the Dominant LULC Transitions

The analysis of the differences between the observed LULC and the expected LULC gains and losses under a random process of changes showed that, of the several observed inter-class transitions, few experienced dominant signals of change between 1995 and 2018 (Figure 6.11), namely:

(i) transitions of arable land to permanent crops and to pastures, which accounted for about 5% and 2.9%, respectively;

(ii) transitions of pastures to forest and arable land, which accounted for about 3.5% and 2.7%, respectively;

(iii) transitions of forest to heterogeneous areas, which accounted for about 2.1%.

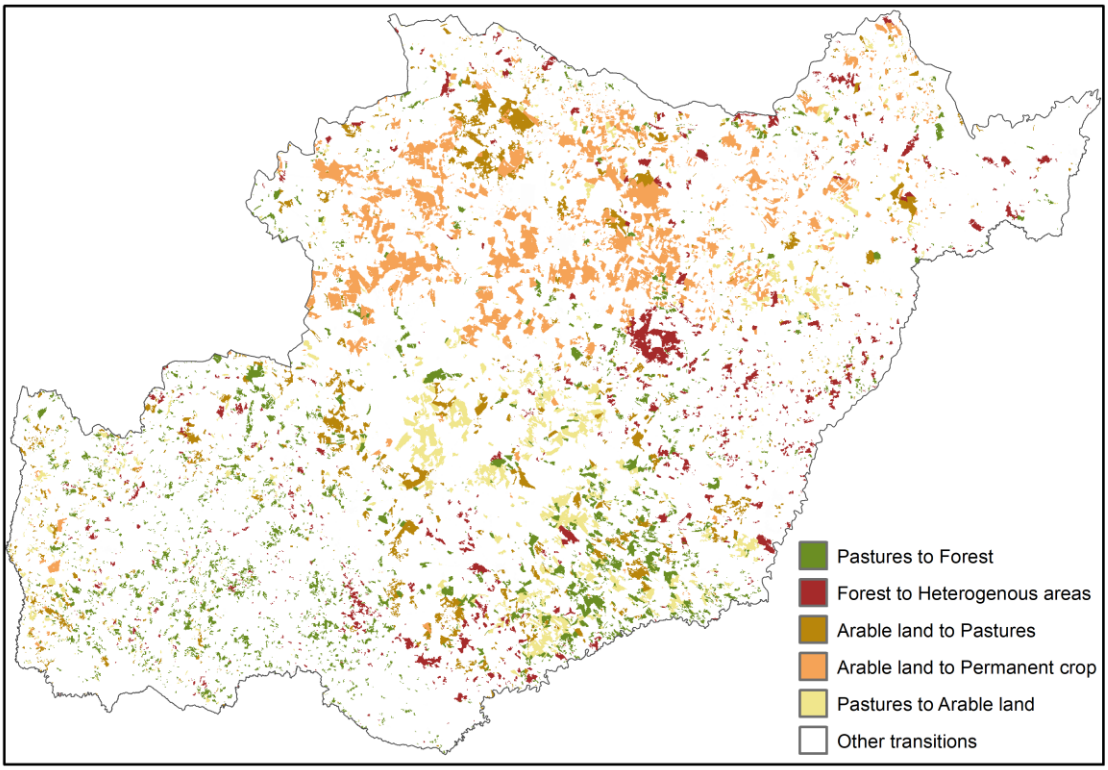


Figure 6.11. Main systematic LULC transitions (1995–2018)

Accordingly, the arable land tendency to lose, especially for the permanent crops class, causes some concerns. The Beja district located in the Mediterranean region is a suitable region to produce olive oil mainly due to favorable weather conditions. Moreover, the positive market environment prevailed with the gradual increase in global olive oil consumption in the past decade, and the super-intensive olive grove plantation also increased to cover the demand (Roxo & Ventura, 2012). Thus, traditional olive grove plantations are becoming unusual, mainly because the super-intensive cropland production profitability is higher and because of Common Agricultural Policy (CAP) incentives (Ferreira, 2001). While this type of agricultural practice will start producing in the 3rd year after planting, the traditional practice can occur after one decade. In addition, mechanization in the super-intensive olive grove is less labor-dependent, meaning a greater capacity for profit. Hence, this all points to one conclusion: the high increase in super-intensive olive crops to produce olive oil is highly associated with the expansion of permanent crops, which is at the expense of arable land.

Thus, the Beja district is experiencing a new landscape dynamic, caused not only by the systematic increase in permanent crops areas but also by the decrease of arable land that produces a historical agricultural product: wheat (the grain of civilization). Indeed, over the last century, Beja consistently showed remarkable importance in wheat production, mainly explained by: (i) the type of soil in Beja, "Barros de Beja" (Beja clays), which is a dense and deep soil that is particularly suitable for this cereal production; (ii) the type of land exploitation (latifundium) (Jones et al., 2011); and (iii) the weather conditions (i.e., a temperate climate with Mediterranean characteristics), as the wheat crop has a relative tolerance to water deficiency (Freire & Lains, 2017).

Thus, these results indicate that, although arable land predominated in the Beja district between 1995 and 2018, this cropland experienced a significant decrease due to the competition between LULC classes. Accordingly, the transition from arable land to permanent crops was relatively constant over time, and it was more pronounced in the last period (2015–2018), presumably triggered by the substantial increase of foreign investment (especially from Spanish investors) and the strong exploitation of water resources, after the construction of the Alqueva dam that filled its reservoir only 10 years after construction began (in 2012) (Roxo & Ventura, 2012). Thus, it seems essential to assess the impact of this systematic transition, since everything points to an increasing super-intensive olive grove cultivation system that will result in a negative impact on this land in the future (García-Ruiz et al., 2013).

In addition, arable land also experienced systematic transitions to pastures. This change was especially strong in the first period (1995–2007), possibly due to changes in the CAP, which led to a decrease in the areas under arable crops, in particular cereal production, and an increase in fallow land maintenance and natural pasture. These policies arose out of a concern related to land preservation in the context of rural planning, resulting, among other actions, in reduced dryland cereal cultivation areas, due to the recognition of the negative impact on soil depletion and erosion resulting from previous agricultural campaigns (Serrano, 2006) and an international acknowledgement of the role of such areas in mitigating climate change (Sevov et al., 2018; Thomas et al., 1992).

On the contrary, Beja also experienced a systematic transition from pastures to arable land. This dynamic LULC change was substantially stronger in the first period (1995–2007) and partly contradicts the prior interpretation of the possible factors responsible for such changes. Therefore, following the abovementioned explanations, we believe this transition can also be linked with the implementation of the CAP, which favors fallow land maintenance. Indeed, the arable land class mainly comprises a land rotation system between annually harvested crops (rain-fed or irrigated) and fallow plots, which in our study area mainly represents dry-fed and rain-fed cereal production. However, according to the COS map nomenclature, this class can also comprise fallow land for a maximum of five years. Additionally, agricultural pasture systems have limitations regarding their production irregularity over a given year due to the irregularity of rainfall (Cosentino et al., 2013; Ferreira, 1992) and may not be economically viable for farmers if not subsidized (Jones et al., 2011; Pearson et al., 1987), which may partly justify these changes in the landscape.

In the Beja district, a systematic transition from pastures to forestland was also found, and this was most significant in the first period (1995–2007) and was probably associated with afforestation policy (the 1992 CAP reform) and EU subsidies (Agenda 2000), which promoted the conversion of pastures in Eucalyptus and *Pinus* forests (Jones et al., 2011) for conservation of the land's structure and functions. Lastly, the systematic transition from forest to heterogeneous areas, which occurred significantly between the third period (2010–2015), is probably related to forest fires, as reforestation took place after this event in the form of *Montado* to combat desertification (Botequim et al., 2017).

For instance, the examination of the systematic and random transition showed that some of the major LULC transitions, such as heterogeneous areas and arable land transitioning to forest (2.53% and 1.78%, respectively) and the heterogeneous areas transitioning to pastures (1.18%), were random or partially systematic. In addition, it is interesting that heterogeneous areas did not experience systematic losses, which indicates a strong signal of persistence in this LULC class. Thus, the Beja district still retains important traditional agricultural areas, such as the *Montado*, an agrosilvopastoral system of global importance.

6.4.3. About the LULC Future Changes

The transition probability simulation between uses for the year 2040 allowed us to understand that imbalances can arise in the future, highlighting that about 25% of the territory is predicted to transition, and about 21% of that territory corresponds to agricultural land. These results are a warning that the Beja district will experience strong landscape dynamics in agricultural land, emphasizing the importance of measures in mitigating the negative effects of such strong signals of LULC transitions (Braimoh, 2006).

Additionally, the DTMC model showed a high agreement value, proving to be a useful mechanism for monitoring and evaluating LULC changes, with a trend projection with high applicability and flexibility. Therefore, the main advantage of using Markov chains is their mathematical logic that enables swift applications, as well as their possible interaction with other techniques such as remote sensing and GIS, both of which commonly work with matrices. However, a disadvantage of this model is that it does not explain the facts, because the process only takes into account the time variable. Thus, predictions made through this model mostly indicate changes in the state of variables in time and not in space. Nevertheless, the stationary probabilities are useful as indicators of the landscape shares of each LULC class and provide highly valuable information in decision-making processes.

6.5. Conclusions

The purpose of this study was to analyze the spatiotemporal evolution of LULC in Beja from 1995 to 2018 and simulate the future development for the year 2040. This analysis indicates that in 23 years there were significant changes that deeply modify the Beja district landscape, mainly due to two opposite phenomena: the intensification of agricultural practices and the afforestation of agricultural lands. The observed LULC patterns reflect the last decade's structural changes in the Portuguese economy and society and confirm the impact of political decisions and different agricultural activity management, which led to the loss of a significant amount of agricultural heritage. In fact, this regional landscape and the traditional features of agricultural land are facing an uneven transformation dynamic that does not take into account either the balance or the environmental enhancement of the local diversity agricultural systems (Ferreira, 2001).

The acquired spatiotemporal knowledge regarding the agricultural LULC changes allowed us to obtain a general picture of the main transitions of a region of Portugal that has more importance in the regional agricultural production sector, particularly in the production of cereals, such as wheat, barley, oats, and other products such as olive oil and cork. An understanding of the systematic transitions and the locations thereof provides useful insight into ecosystem service management and contributes to future sustainability strategies. Nonetheless, this statistical approach is exclusively based on an analysis of the LULC transition matrix, and the interpretation of the LULC transitions is based on a sequence of historical determinism documented in prior investigations (e.g., Ferreira, 1992; Ferreira, 2001; Fonseca, 1996), so any factor mentioned to rationalize the systematic transition is based only on assumptions of historical fact. Therefore, additional research is necessary to assess the many processes that create these LULC changes, which have been linked to dozens of possible drivers and their different levels of feedback. Still, providing an understanding of the inter-class transitions that have occurred as systematic changes has proven to be a straightforward, effective approach that allows us to differentiate the predominating quantity transitions from those that experienced a single abrupt episode process of change influenced by factors that occur suddenly.

The information obtained contributed to sciences related to the earth and human society on a variety of levels, providing effective information for economic and environmental monitoring and evaluation, resulting in informed policy decisions and in the design of anticipatory measures response to change. Furthermore, in this exploratory study, we used COS maps, official Portuguese LULC data produced by a governmental institution (DGT), and these data were advantageous because of their higher cartographic detail and were characterized by temporal continuity. This highlights the importance of having an entity responsible for producing LULC thematic cartography with regularity and reliability and making it available free of charge, as it can be used for multiple purposes, such as territorial planning and management, ecosystem service mapping, and the identification of areas vulnerable to rural fires (e.g., Boavida-Portugal et al., 2016; Gomes, Abrantes, Banos, & Rocha, 2019; Meneses et al., 2018).

References

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P., & Schweik, C. M. (2002). A Review and Assessment of Land-Use Change Models: Dynamics of Space, Time, and Human Choice. Apollo The International Magazine Of Art And Antiques, 62. https://doi.org/10.1289/ehp.6514

Ahmed, B., Ahmed, R., & Zhu, X. (2013). Evaluation of Model Validation Techniques in Land Cover Dynamics. ISPRS International Journal of Geo-Information, 2(3), 577–597. https://doi.org/10.3390/ijgi2030577

Allen, H., Simonson, W., Parham, E., Santos, E. de B. e, & Hotham, P. (2018). Satellite remote sensing of land cover change in a mixed agro-silvo-pastoral landscape in the Alentejo, Portugal. International Journal of Remote Sensing, 00(00), 1–21. https://doi.org/10.1080/01431161.2018.1440095

Basharin, G. P., Langville, A. N., & Naumov, V. A. (2004). The life and work of A.A. Markov. Linear Algebra and Its Applications, 386(1-3 SUPPL.), 3–26. https://doi.org/10.1016/j.laa.2003.12.041

Boavida-Portugal, I., Rocha, J., & Ferreira, C. C. (2016). Exploring the impacts of future tourism development on land use/cover changes. Applied Geography, 77, 82–91. https://doi.org/10.1016/j.apgeog.2016.10.009

Botequim, B., Arias-Rodil, M., Garcia-Gonzalo, J., Silva, A., Marques, S., Borges, J., Oliveira, M., & Tomé, M. (2017). Modeling Post-Fire Mortality in Pure and Mixed Forest Stands in Portugal—A Forest Planning-Oriented Model. Sustainability, 9(3), 390. https://doi.org/10.3390/su9030390

Braimoh, A. K. (2006). Random and systematic land-cover transitions in northern Ghana. Agriculture, Ecosystems and Environment, 113(1–4), 254–263. https://doi.org/10.1016/j.agee.2005.10.019

Correia, T. P. (1993). Threatened landscape in Alentejo, Portugal: the ‘montado’ and other ‘agro-silvo-pastoral’ systems. Landscape and Urban Planning, 24(1–4), 43–48. https://doi.org/10.1016/0169-2046(93)90081-N

Cosentino, S. L., Gresta, F., & Testa, G. (2013). Forage chain arrangement for sustainable livestock systems in a Mediterranean area. Grass and Forage Science, 69(4), 625–634. https://doi.org/10.1111/gfs.12085

Deng, X., Huang, J., Rozelle, S., & Uchida, E. (2008). Growth, population and industrialization, and urban land expansion of China. Journal of Urban Economics, 63(1), 96–115. https://doi.org/10.1016/j.jue.2006.12.006

Deribew, K. T. (n.d.). Spatially explicit statistical modeling of random and systematic land cover transitions in the main grassland plain of Nech Sar National Park, Ethiopia. https://doi.org/10.1186/s13717-019-0199-z

DGT. (2018). Especificações Técnicas da Carta de Uso e Ocupação do Solo (COS) de Portugal Continental para 1995, 2007, 2010 e 2015.

Feranec, J., Soukup, T., Taff, G., Stych, P., & Bicik, I. (2017). Overview of changes in land cover and land use in Eastern Europe. In G. Gutman & V. Radeloff (Eds.), Land-cover and land-use changes in Eastern Europe after the collapse of the Soviet Union in 1991 (Springer).

Feranec, Jan, Jaffrain, G., Soukup, T., & Hazeu, G. (2010). Determining changes and flows in European landscapes 1990-2000 using CORINE land cover data. Applied Geography, 30(1), 19–35. https://doi.org/10.1016/j.apgeog.2009.07.003

Ferreira, D. B. (1992). Problemas florestais no Alentejo: a crise do montado e a eucaliptização. Finisterra, 27(53/54), 25–62. https://doi.org/10.18055/finis1879

Ferreira, D. de B. (2001). Evolução Da Paisagem De Montado No Alentejo Interior Ao Longo Do Século XX: Dinâmica e incidências ambientais. Finisterra, 36(72), 179–193.

Fonseca, H. A. (1996). O Alentejo no Século XIX. Economia e Atitudes Económicas. Imprensa Nacional.

Freire, D., & Lains, P. (2017). An agrarian history of Portugal, 1000-2000: economic development on the European frontier (D. Freire & P. Lains (eds.)). Brill.

Fuchs, R., Herold, M., Verburg, P. H., Clevers, J. G. P. W., & Eberle, J. (2015). Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010. Global Change Biology, 21(1), 299–313. https://doi.org/10.1111/gcb.12714

Gagniuc, P. A. (2017). Markov chains : from theory to implementation and experimentation.

García-Ruiz, J. M., Nadal-Romero, E., Lana-Renault, N., & Beguería, S. (2013). Erosion in Mediterranean landscapes: Changes and future challenges. Geomorphology, 198(September), 20–36. https://doi.org/10.1016/j.geomorph.2013.05.023

Gomes, E., Abrantes, P., Banos, A., & Rocha, J. (2019). Modelling future land use scenarios based on farmers' intentions and a cellular automata approach. Land Use Policy, 85, 142–154. https://doi.org/10.1016/j.landusepol.2019.03.027

Gomes, E., Abrantes, P., Banos, A., Rocha, J., & Buxton, M. (2019). Farming under urban pressure: Farmers' land use and land cover change intentions. Applied Geography, 102, 58–70. https://doi.org/10.1016/j.apgeog.2018.12.009

Hamad, R., Balzter, H., & Kolo, K. (2018). Predicting land use/land cover changes using a CA-Markov model under two different scenarios. Sustainability (Switzerland), 10(10), 3421. https://doi.org/10.3390/su10103421

Han, H., Yang, C., & Song, J. (2015). Scenario simulation and the prediction of land use and land cover change in Beijing, China. Sustainability (Switzerland), 7(4), 4260–4279. https://doi.org/10.3390/su7044260

Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. Remote Sensing of Environment, 122, 66–74. https://doi.org/10.1016/j.rse.2011.08.024

Hathout, S. (2002). The use of GIS for monitoring and predicting urban growth in East and West St Paul, Winnipeg, Manitoba, Canada. Journal of Environmental Management, 66(3), 229–238. https://doi.org/10.1016/S0301-4797(02)90596-7

Hyandye, C., & Martz, L. W. (2017). A Markovian and cellular automata land-use change predictive model of the Usangu Catchment. International Journal of Remote Sensing, 38(1), 64–81. https://doi.org/10.1080/01431161.2016.1259675

Iacono, M., Levinson, D., El-Geneidy, A., & Wasfi, R. (2015). A Markov chain model of land use change in the Twin Cities, 1958-2005. Tema. Journal of Land Use, Mobility and Environment, 8(3), 263–276. https://doi.org/10.6092/1970-9870/2985

INE. (2012). Censos 2011 Resultados Definitivos – Região Alentejo. Instituto Nacional de Estatística.

Jones, N., de Graaff, J., Rodrigo, I., & Duarte, F. (2011). Historical review of land use changes in Portugal (before and after EU integration in 1986) and their implications for land degradation and conservation, with a focus on Centro and Alentejo regions. Applied Geography, 31(3), 1036–1048. https://doi.org/10.1016/j.apgeog.2011.01.024

Koohafkan, P. (Parviz), & Altieri, M. A. (2016). Forgotten agricultural heritage : reconnecting food systems and sustainable development.

Lambin, E. F., Geist, H. J., & Lepers, E. (2003). Dynamics of Land-use and Land-cover change in tropical regions. Annual Review of Environment and Resources, 28(1), 205–241. https://doi.org/10.1146/annurev.energy.28.050302.105459

Li, X., & Yeh, A. G.-O. (2002). Neural-network-based cellular automata for simulating multiple land use changes using GIS. International Journal of Geographical Information Science, 16(4), 323–343. https://doi.org/10.1080/13658810210137004

Liu, Y., & Feng, Y. (2016). Simulating the Impact of Economic and Environmental Strategies on Future Urban Growth Scenarios in Ningbo, China. Sustainability, 8(10), 1045. https://doi.org/10.3390/su8101045

Lu, Q., Liang, F., Bi, X., Duffy, R., & Zhao, Z. (2011). Effects of urbanization and industrialization on agricultural land use in Shandong Peninsula of China. Ecological Indicators, 11(6), 1710–1714. https://doi.org/10.1016/j.ecolind.2011.04.026

Meneses, B. M., Reis, E., Pereira, S., Vale, M. J., & Reis, R. (2017). Understanding driving forces and implications associated with the land use and land cover changes in Portugal. Sustainability (Switzerland), 9(3), 9–14. https://doi.org/10.3390/su9030351

Meneses, B. M., Reis, E., Vale, M. J., & Reis, R. (2018). Modelling the Land Use and Land cover changes in Portugal: A multi-scale and multi-temporal approach. Finisterra, 53(107). https://doi.org/10.18055/Finis12258

Mertens, B., Sunderlin, W. D., Ndoye, O., & Lambin, E. F. (2000). Impact of macroeconomic change on deforestation in South Cameroon: Integration of household survey and remotely-sensed data. World Development, 28(6), 983–999. https://doi.org/10.1016/S0305-750X(00)00007-3

Muñoz-Rojas, J., Pinto-Correia, T., Hvarregaard Thorsoe, M., & Noe, E. (2019). The Portuguese Montado : A Complex System under Tension between Different Land Use Management Paradigms . In Silvicultures - Management and Conservation. IntechOpen. https://doi.org/10.5772/intechopen.86102

Pearson, S., Avillez, F., & Bentley, J. W. (1987). Portuguese agriculture in transition. Cornell University PressIthaca: Cornell University Press.

Petit, C., Scudder, T., & Lambin, E. F. (2001). Quantifying processes of land-cover change by remote sensing: Resettlement and rapid land-cover changes in south-eastern Zambia. International Journal of Remote Sensing, 22(17), 3435–3456. https://doi.org/10.1080/01431160010006881

Pinto-Correia, T. (2000). Future development in Portuguese rural areas: how to manage agricultural support for landscape conservation? Landscape and Urban Planning, 50(1–3), 95–106. https://doi.org/10.1016/S0169-2046(00)00082-7

Pontius Jr, Robert G, Shusas, E., & Mceachern, M. (2004). Detecting important categorical land changes while accounting for persistence. Agriculture, Ecosystems and Environment, 101, 251–268. https://doi.org/10.1016/j.agee.2003.09.008

Pontius Jr, Robert Gilmore, Boersma, W., Castella, J. C., Clarke, K., Nijs, T., Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C. D., McConnell, W., Mohd Sood, A., Pijanowski, B., Pithadia, S., Sweeney, S., Trung, T. N., Veldkamp, A. T., & Verburg, P. H. (2008). Comparing the input, output, and validation maps for several models of land change. Annals of Regional Science, 42(1), 11–37. https://doi.org/10.1007/s00168-007-0138-2

Rousta, I., Sarif, M. O., Gupta, R. D., Olafsson, H., Ranagalage, M., Murayama, Y., Zhang, H., & Mushore, T. D. (2018). Spatiotemporal analysis of land use/land cover and its effects on surface urban heat Island using landsat data: A case study of Metropolitan City Tehran (1988-2018). Sustainability (Switzerland), 10(12), 4433. https://doi.org/10.3390/su10124433

Roxo, M. J., & Ventura, J. E. (2012). Land Use Change in the Lower Alentejo (South of Portugal) and the New Environmental Challenges: The Olive Culture. In R. Efe, M. Ozturk, & S. Ghazanfa (Eds.), Environment and Ecology in the Mediterranean Region (pp. 13–28). Cambridge Scholars Publishing.

Serra, P., Pons, X., & Saurí, D. (2008). Land-cover and land-use change in a Mediterranean landscape: A spatial analysis of driving forces integrating biophysical and human factors. Applied Geography, 28(3), 189–209. https://doi.org/10.1016/j.apgeog.2008.02.001

Serrano, J. E. (2006). Pastagens do Alentejo: bases técnicas sobre caracterização, pastoreio e melhoramento. (U. de É.- ICAM (ed.)). Gráfica Eborense.

Sevov, A., Yancheva, C., & Kazakova, Y. (2018). Sustainable Pasture Management. In New Perspectives in Forage Crops. InTech. https://doi.org/10.5772/intechopen.72310

Shi, W., Fisher, P. F., & Goodchild, M. F. (2002). Spatial Data Quality. Taylor & Francis.

Tavares, A. O., Pato, R. L., & Magalhães, M. C. (2012). Spatial and temporal land use change and occupation over the last half century in a peri-urban area. Applied Geography, 34, 432–444. https://doi.org/10.1016/j.apgeog.2012.01.009

Teferi, E., Bewket, W., Uhlenbrook, S., & Wenninger, J. (2013). Understanding recent land use and land cover dynamics in the source region of the Upper Blue Nile, Ethiopia: Spatially explicit statistical modeling of systematic transitions. Agriculture, Ecosystems and Environment, 165, 98–117. https://doi.org/10.1016/j.agee.2012.11.007

Thomas, R. J., Lascano, C. E., Sanz, J. I., Ara, M. A., Spain, J. M., Vera, R. R., & Fisher, M. J. (1992). The role of pastures in production systems. CIAT.

Turner, B. L., Moss, R. H., & Skole, D. L. (1993). Relating land use and global land-cover change — IGBP Report No. 24, HDP Report No. 5.

Uddin, K., Matin, M. A., & Maharjan, S. (2018). Assessment of land cover change and its impact on changes in soil erosion risk in Nepal. Sustainability (Switzerland), 10(12), 4715. https://doi.org/10.3390/su10124715

Van der Wel, F. J. M., Van der Gaag, L. C., & Gorte, B. G. H. (1998). Visual exploration of uncertainty in remote-sensing classification. Computers and Geosciences, 24(4), 335–343. https://doi.org/10.1016/S0098-3004(97)00120-9

Viana, C. M., Freire, D., & Rocha, J. (2019). Regional agricultural production statistics for 160 years using the geographic information system and the spatial analytical technique. In P. Kyriakidis, D. Hadjimitsis, D. Skarlatos, & A. Mansourian (Eds.), Geospatial Technologies for Local and Regional Development:: Short Papers, Posters and PosterAbstracts of the 22th AGILE Conference on Geographic Information Science (Issue 1989). Cyprus University of Technology: Limassol, Cyprus on 17-20 June, 2019.

Viana, C. M., Girão, I., & Rocha, J. (2019). Long-Term Satellite Image Time-Series for Land Use/Land Cover Change Detection Using Refined Open Source Data in a Rural Region. Remote Sensing, 11(9), 1104. https://doi.org/10.3390/rs11091104

Viana, C. M., Oliveira, S., Oliveira, S. C., & Rocha, J. (2019). Land Use/Land Cover Change Detection and Urban Sprawl Analysis. In H. R. Pourghasemi & C. Gokceoglu (Eds.), Spatial Modeling in GIS and R for Earth and Environmental Sciences (1st ed.). Elsevier.

Viana, C. M., & Rocha, J. (2018). Spatiotemporal analysis and scenario simulation of agricultural land use land cover using GIS and a Markov chain model. In A. Mansourian, P. Pilesjö, L. Harrie, & R. von Lammeren (Eds.), Geospatial Technologies for All : short papers, posters and poster abstracts of the 21th AGILE Conference on Geographic Information Science.

Weng, Q. (2002). Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. Journal of Environmental Management, 64(3), 273–284. https://doi.org/10.1006/jema.2001.0509

Yirsaw, E., Wu, W., Shi, X., Temesgen, H., & Bekele, B. (2017). Land Use/Land Cover Change Modeling and the Prediction of Subsequent Changes in Ecosystem Service Values in a Coastal Area of China, the Su-Xi-Chang Region. Sustainability, 9(7), 1204. https://doi.org/10.3390/su9071204

Chapter 7. Evaluation of the factors explaining the use of agricultural land: a machine learning and model-agnostic approach

Cláudia M. Viana, Maurício Santos, Dulce Freire, Patrícia Abrantes, Jorge Rocha
Article published at the Ecological Indicators (2021), Volume 131, 108200.
https://doi.org/10.1016/j.ecolind.2021.108200
The structure and formatting are slightly adapted.

Abstract

To effectively plan and manage the use of agricultural land, it is crucial to identify and evaluate the multiple human and environmental factors that influence it. In this study, we propose a model framework to identify the factors potentially explaining the use of agricultural land for wheat, maize, and olive grove plantations at the regional level. By developing a machine-learning model coupled with a model-agnostic approach, we provide global and local interpretations of the most influential factors. We collected nearly 140 variables related to biophysical, bioclimatic, and agricultural socioeconomic conditions. Overall, the results indicated that biophysical and bioclimatic conditions were more influential than socioeconomic conditions. At the global interpretation level, the proposed model identified a strong contribution of conditions related to drainage density, slope, and soil type. In contrast, the local interpretation level indicated that socioeconomic conditions such as the degree of mechanisation could be influential in specific parcels of wheat. As demonstrated, the proposed analytical approach has the potential to serve as a decision-making tool instrument to better plan and control the use of agricultural land.

Keywords: Cropland; Interpretability; Artificial intelligence; xAI; LIME

7.1. Introduction

The current global trends of population growth, accelerated urbanisation, and environmental changes, which are associated with the encroachment of agricultural land (Foley et al., 2011; Radwan et al., 2019), agricultural land abandonment (Castillo et al., 2021), and agricultural land fragmentation (Gomes, Banos, et al., 2019; Postek et al., 2019), have an influence on food production and food security (Godfray et al., 2010; W. B. Wu et al., 2014). For the coming decades, enhancing and maintaining food supply will require the efficient use of agricultural land (FAO, 2017; Wu et al., 2014). However, multiple factors (e.g. natural and environmental) that vary both temporally and spatially determine and affect the use of agricultural land (Akpoti et al., 2019; Lambin et al., 2001; Ndamani & Watanabe, 2017). Thus, more studies are needed to better identify and evaluate the factors influencing agricultural land use under different cross-scale and geographical contexts.

Traditionally, empirical and conventional statistical methods such as principal component analysis (PCA), clustering methods, regression, and other linear approaches have been used to better understand the factors influencing land use (Braimoh, 2009; Marcos-Martinez et al., 2017; Santiphop et al., 2012; Velásquez-Milla et al., 2011). While the application of such methods can provide useful information to support effective planning and management measures as well as better-informed decisions concerning efficient land use, they present some analytical limitations. For instance, these statistical methods may not fully capture nonlinear behaviour or discard the effects of heterogeneity and spatial autocorrelation from the analysis (Cartone & Postiglione, 2020; Demšar et al., 2013; Jombart et al., 2008).

Conversely, machine learning (ML), which is a subfield of artificial intelligence (AI), has successfully overcome the limitations of statistical methods. Compared to traditional methods, ML is recognised to achieve superior or at least equivalent accuracy outcomes (Lima et al., 2015; Ren et al., 2020; Shortridge et al., 2016). In turn, ML approaches have many advantages, such as the capability to deal with data of different types, structures, and quantities (i.e. big data) (Molnar, 2019), being non-sensitive to the scale of variables (meaning there is no need for variable normalisation); therefore, it is possible to exploit and combine different data resources to model complex nonlinear relationships that describe agricultural land-use systems.

Owing to the great variety of robust algorithms and flexible model structures (e.g. artificial neural networks (ANNs) and random forests (RFs)), ML models represent a potential solution to the requirements of different land-use modelling applications (Hagenauer et al., 2019). Despite the potential advantages, ML algorithms remain mostly under a 'black box' formulation, which means that without further intervention, it is not possible to directly interpret or retrace how a model performs inference or prediction owing to the many internal weights or structural information (Molnar, 2019). Nevertheless, the explainable AI (xAI) has recently emerged as an important research area, which proposes advanced statistical measures and visualisation tools to enhance the interpretability of ML (Carvalho et al., 2019; Molnar, 2019; Murdoch et al., 2019). For instance, post-hoc techniques such as model-agnostic models have been proposed as interpretability methods to provide explanations about the function underlying the general behaviour of ML models (Molnar, 2019; Murdoch et al., 2019; Ribeiro et al., 2016). The main advantage of a model-agnostic approach is its flexibility as it can deal with the opacity of any kind of black box ML model and gather interpretability, which is a critical aspect when the ML model outcomes are used as a basis for decision making (Ribeiro et al., 2016). Currently, there are some examples of model-agnostic interpretation methods that are global or local in scope, such as the permutation feature importance (PFI), partial dependence plots (PDPs), or local surrogate models (e.g. local interpretable model-agnostic explanations (LIME)) (Molnar, 2019).

To date, ML models have been successfully applied in a wide range of Earth and environmental science studies, for example, in estimating air pollution (Ren et al., 2020), predicting dengue importation (Salami et al., 2020), modelling coastal fish communities (Lehikoinen et al., 2019), and predicting marine fish distributions (Zhang et al., 2019). In the scientific field of land-use modelling, ML has been mostly used for image classification and land use/land cover (LULC) mapping (Abdi, 2020; Raczko & Zagajewski, 2017), as well as to simulate future LULC changes (Gomes, Banos, et al., 2019; Hagenauer et al., 2019). However, the recently developed areas of xAI research and model-agnostic methods, which provide the ML model interpretability needed to enhance scientific consistency (Molnar, 2019), have rarely been introduced to agricultural land modelling studies.

Accordingly, this study explores the use of an ML model coupled with a model-agnostic approach to increase the understanding of human and environmental factors that can explain the use of agricultural land for three cropland plantations relevant to food security and Mediterranean basin ecosystems: wheat, maize, and olive groves (FAO, 2018; Loumou & Giourga, 2003). Thus, we developed an analytical framework using the RF ML algorithm and PFI, PDPs, and LIME model-agnostic methods to provide global and local interpretabilities to understand how bioclimatic, biophysical, and socioeconomic conditions might explain the land used for these three cropland plantations. From a quantitative methodology perspective, this study demonstrates the usefulness of such methods to deal with some of the above-described analytical challenges, and provides novel insights into the use of agricultural land at the regional scale.

7.2. Material and Methods

7.2.1. Case study and cropland context

The case study in which the modelling framework was developed is the Beja district located in southern Portugal, with an area of approximately 10,229.05 km² that covers 11% of Portugal's mainland territories (Figure 7.1). In 2011, the district had a population of 152,758 inhabitants, distributed among 14 municipalities and 75 parishes (INE, 2012). The climate in Beja is influenced by its distance from the coast, with a Mediterranean climate characterised by hot and dry summers and wet and cold winters. This large predominantly agricultural region includes a vast landscape of intermingling cultures, such as wheat, olive groves, vineyards, and cork oak forests. In addition, in this region, we can find an agrosilvopastoral agricultural heritage system named Montado that has been indicated as a globally important agricultural system according to the Food and Agriculture Organization of the United Nations (FAO) (Correia, 1993b; Koohafkan & Altieri, 2016; Muñoz-Rojas et al., 2019). The region's high natural and economic value, in particular to food security and Mediterranean basin ecosystems, emphasizes the case study choice.

Over the last century, Beja consistently produced increasing amounts of wheat due to its ecological–biophysical conditions and parcel structures. The region is also characterised by an enlargement of olive grove plantations, whereby a change from open production to intensive and super-intensive production has been witnessed over the past decade due to the increased exploitation of water resources following the construction of Alqueva Dam (Viana, Girão, et al., 2019; Viana & Rocha, 2020). Although maize plantations were historically confined to the northern regions of Portugal, which are more humid and have a higher water availability than the south, they can now be found in Beja as a result of the construction of irrigation systems during the past decades.

Figure 7.1. Location of the Beja district in Portugal

7.2.2. Experimental design

The model framework was developed to understand the multiple factors explaining the land used for wheat, maize, and olive grove plantations. The framework includes five main stages: (1) collection and pre-processing of spatial data, (2) data multicollinearity diagnosis, (3) ML model building, and (4) application of a model-agnostic approach for interpretability. The workflow of the process is shown in Figure 7.2.

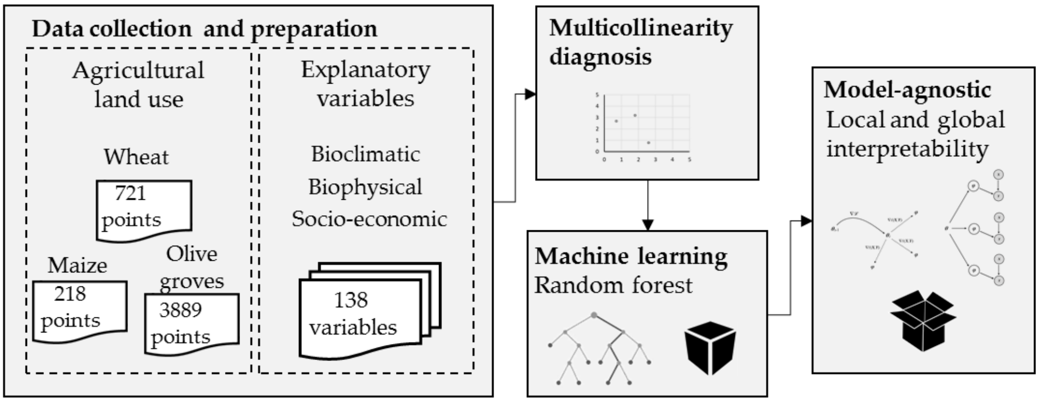


Figure 7.2. Workflow of the modelling process

7.2.2.1 Data collection and pre-processing

7.2.2.1.1. Derivation of the response variable

The spatial locations of wheat, maize, and olive groves were obtained from the Portuguese Institute for Financing Agriculture and Fisheries (IFAP) (https://www.ifap.pt/isip/ows/). The IFAP provides vector 1:10,000 (polygons) structured data concerning the Land Parcel Identification System, which identifies the limit of parcels of national farming systems and classifies agricultural land use (reference data for 2020). This dataset is produced by the Portuguese government for the submission of applications for community aid and the execution of control actions for farmers. For analysis purposes, we generated random points inside the parcel features, with a minimum distance of 500 m between them, to avoid pseudo-replication and to increase the variance of the training data. A total of 721, 218, and 3889 sample presence points were obtained for wheat, maize, and olive groves, respectively (Figure 7.3). Thus, an equal number of real absences were selected and included in the final presence–absence file.

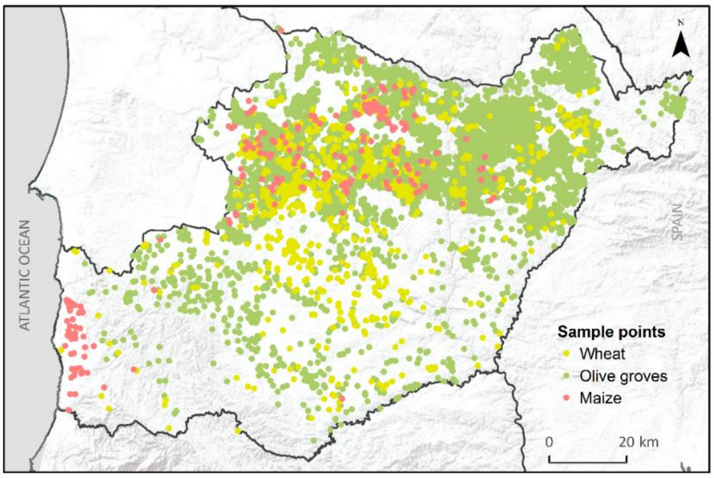


Figure 7.3. Sample points spatial location. The sample size among the three crop plantations depicts the importance in terms of the territorial presence and spatial distribution of each crop plantation

7.2.2.1.2. Derivation of the explanatory variables

Variables (factors) with probable explanatory relevance were selected based on relevant literature (e.g. Akpoti et al., 2019; Kourgialas, 2021; Li et al., 2018; Petit et al., 2011; Valayamkunnath et al., 2020); however, the initial dataset depended on data availability. Briefly, multiple variables were obtained and divided into (i) agricultural socioeconomic statistical data, which provide a current comprehensive information framework for the agricultural sector in the region; and (2) environmental data (bioclimatic and biophysical), which provide basic information on climate and the physical environment. The entire dataset encompassed 138 variables (both categorical and continuous). Table 7.1 presents a summary of the variables included in this study and their metadata. It is worth noting that digital terrain model (DTM) data were used to compute the slope, and drainage network data were used to calculate the drainage density (km/km^2). In addition, land cover data and road networks were used to compute the Euclidian distance to waterbodies, urban areas, and roads. Socio-economic data was collected at the parish level and rasterized at a 100 m resolution. Therefore, all original data were resampled to a common spatial resolution of 100 m to match the climate data resolution and 1:25,000 variable minimum mapping unit (1 ha).

Table 7.1. Summary of the variables included in the model. See Table A1 in the Appendix for further details of the variables

Category	Code	Variable	Scale/Resolution	Year	Data Source
Socio-economic	V1-V55	Agricultural holdings (type of tenure, legal form, type of land use, livestock, with irrigable area)	1:25,000 Parish statistical unit	2019	Statistics Portugal (2019)
	V56-V61	Agricultural types of machinery			
	V62	Familiar agricultural population			
	V63-V89	Sole agricultural holders (Age group, with 65 and more years old, female sex, level of education)			
	V90-V110	Utilised agricultural area			
	V111-V112	Irrigable area (ha) of agricultural holdings and AWU			
Bioclimatic	V113-V114	Mean temperature of the warmest and the coldest month of the year	100 m	1960-1990	Monteiro-Henriques et al. (2016)
	V115-V116	Annual positive temperature and positive precipitation			
	V117-V118	Mean maximum and minimum temperature of the coldest month			
	V119	Simple continentality index, or annual thermal amplitude			
	V120-V121	Thermicity índex and compensated thermicity índex			
	V122	Annual ombrothermic índex			
	V123	Ombrothermic index of the warmest bimonth of the summer quarter			
	V124-V125	Ombrothermic index of the summer quarter and the summer quarter plus the previous month			
	V126-V127	Positive precipitation (for dry and humid years)			
	V128-V129	Ombrothermic index (for dry and humid years)			
	V130-V131	Ombrothermic index anomaly (for dry and humid years)			
Biophysical	V132	Soil type	1:25,000	Static over time	DGADR (https://www.dgadr.gov.pt/)
	V133	Soil capacity			
	V134	Slope			IGEOE (http://www.igeoe.pt/cigeoesig/)
	V136	Drainage density			
	V137-V138-V135	Distance to urban, roads, and water bodies			DGT (2018) and OpenStreetMap (https://www.openstreetmap.org/)

7.2.2.2. Modelling procedures

The first stage of our model approach encompassed a preliminary statistical analysis, during which we diagnostic the variables multicollinearity by calculating the variance inflation factor (VIF) (Dohoo et al., 1997; Lin, 2008). We calculated the VIF using the 'usdm' library in R statistical software, which excluded highly correlated variables from the initial set (n = 138) through a stepwise procedure (Naimi et al., 2014). Any variable with a VIF of >5 was excluded from the model (James et al., 2013; Johnston et al., 2018). The VIF was calculated for each cropland individually and the final number of included variables was 42 for wheat, 25 for maize, and 44 for olive groves (see Table A2 in the Appendix).

The second stage involved building the ML model using the RF algorithm. Several studies have demonstrated that RF exhibits very similar performance or performs better than other ML algorithms (see Al-Fugara et al., 2020; Li et al., 2016; Wu et al., 2019; Yang et al., 2016). Furthermore, the RF algorithm has been described as robust in terms of fitting capacity during training and validation procedures, even with a small number of sample points (Luan et al., 2020; Moghaddam et al., 2020; Y. Qi, 2012). We performed RF using the 'RandomForest' library in R statistical software (Liaw & Wiener, 2002) without pre-tuning the number of trees (n = 500) and setting the number of variables in the subset of each node to $\sqrt[2]{n}$ (Probst & Boulesteix, 2018). Each cropland was modelled separately. Due the RF model structures is important to tune and validate it. Therefore, we used k-fold cross-validation (k = 10) to enhance the model reliability and avoid overfitting (Meyer et al., 2018). We chose a data split of 70% for training and 30% for testing instead of an 80%–20% split due to the low number of presences for maize (218) and wheat (721) (Meyer et al., 2018).

7.2.2.3. Model-agnostic approach for interpretability

The third stage involved interpretability of the ML model by using the model-agnostic approach to extract the post-hoc explanations. The model-agnostic approach provides an explanation based on the different behaviours of fitted complex models (e.g. RF algorithm), presenting information at the global level (i.e. what the model learned from the input variables) and at the local level (i.e. the rationales that the model provides for each estimate). Depending on the purpose of the analysis, different methods can be used jointly for global or local interpretability of the same model. In this study, we applied two methods for global interpretability and one method for local interpretability. The global interpretation was implemented using the following methods:

(a) The PFI method, which is commonly used to measure the increase in model performance error after a variable is permuted (i.e. randomly shuffled) (Molnar, 2019; A. Winkler et al., 2015). Specifically, this method allowed us to understand which variables contributed to the underlying ML model outcomes and quantify their importance scores. In this study, we computed both the area under the curve (AUC) and the R^2 value, and used the latter for the variable importance analysis.

(b) The PDPs method, which depicts the explanatory variables' overall relationship with the response variable (variable explanation probability) by imposing all occurrences to have the same variable value and measuring the marginal or average effect for this value on the model response (Apley & Zhu, 2016; Goldstein et al., 2013; Molnar, 2019).

Therefore, it indicates the marginal effects of variables on the model outcomes, thereby identifying the threshold value at which these variables are likely to explain the land used for a specific cropland plantation.

In addition, local interpretation was implemented via:

(a) The LIME method, which trains a local surrogate model (a simple model such as a decision tree model) to reconstruct the inter logic workings around the individual observation (i.e. the local potential interactions occurring) (Ribeiro et al., 2016b). Formally, the LIME interpretability constraint is defined as:

$$explanation(x) = argmin_{g \in G} L(f, g, \pi x) + \Omega(g) \quad (7.1)$$

where the explanation model (x) is the model (g), for example, a generalised linear model, which minimises the loss (L) using the mean squared error, which measures how close the explanation is to the prediction of the original model (f), for example, a RF model while keeping the model complexity $\Omega(g)$ low (e.g. using the minimum of features). Value (G) is the number of possible explanations, for example, all possible general linear models. The proximity measure πx defines the size of the neighbourhood considered for the explanation around instance x. In practice, LIME only optimises the loss part, and the user has to determine the complexity, for example, by selecting the maximum number of features that the linear regression model may use. Therefore, it is possible to understand whether variables that increase the explanation probability either support or contradict the explanation for a given parcel. Locally, the behaviour of the ML model might be different because the outcomes rely linearly or monotonically on some variables, instead of having a complex dependence on them; therefore, local interpretability might be more accurate than global interpretability (Ribeiro et al., 2016b).

In this study, PFI and PDPs were performed using the 'varim' and 'pdp' packages (Greenwell, 2017; Probst & Janitza, 2020) of R version 4.0.2 statistical software, respectively, while the 'lime' package was used to implement the LIME method (Pedersen & Benesty, 2019).

7.3. Results

The RF modelling outputs were examined for global interpretation using the PFI method. Figure 7.4a–c presents the 10 most influential variables in explaining the land used for wheat, maize, and olive groves. For each cropland, the list of variable importance was distinctive. Although 5 of the 10 variables were the same for the explanation, they presented importance scores quite differently. Overall, the model results indicate that bioclimatic and biophysical variables provided a more significant explanation, while socioeconomic variables were less important and did not seem to affect the model significantly.

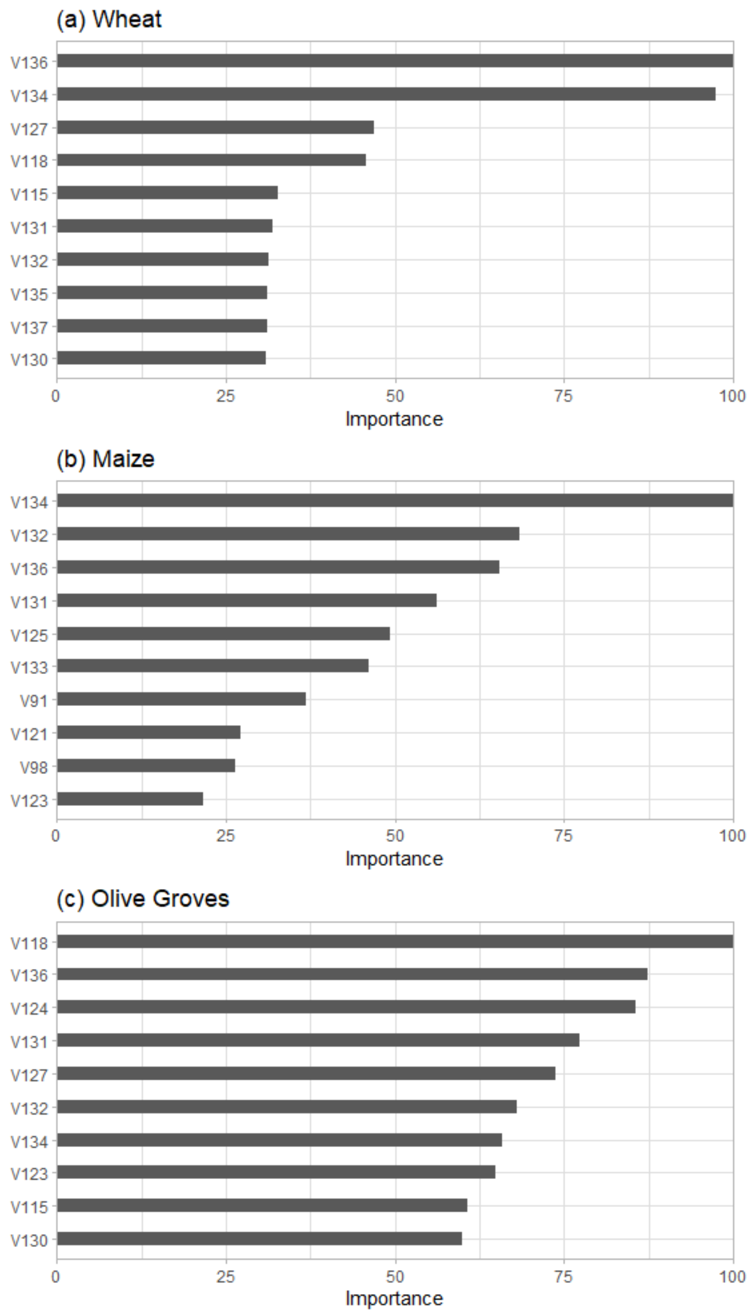


Figure 7.4. Ten most influential variables and respective importance score in explaining wheat (a), maize (b), and olive grove (c) plantations. See Table A1 in the Appendix for further details of the variables

The most influential variables (with an importance score of >50) in explaining land use for wheat plantations were drainage density (V136) and slope (V134), while those for maize plantations were slope, soil type (V132), drainage density, and the ombrothermic index anomaly for humid years (V131). In the case of land used for olive grove plantations, all 10 most influential variables had an importance score of >50, but the mean minimum temperature of the coldest month (V118), draining density, and ombrothermic index of the summer quarter (V124) were the top three most important.

The PDP method was also used to provide a global interpretation. Figure 7.5 presents the response curves for the first six most influential variables and their probability of explaining the land used for wheat plantations. In particular, each plot shows that the probability of land being used for wheat plantations increased on average: (i) by 0.4 as the drainage density increased up to 5 km/km^2, after which the probability did not change; (ii) by 0.4 as the slope increased up to 30%, after which the probability did not change; (iii) by 0.25 as positive precipitation for humid years increased up to 800 mm, after which the probability did not change; (iv) by 0.20 as the mean temperature of the coldest month of the year increased, with a positive impact up to 4.5 ºC, whereas the variable effect reduced quickly at >5 ºC before the probability increased again by 0.20 at >5.5 ºC; (v) by 0.18 as the annual positive temperature increased (the variable effect reduced quickly after 1900 ºCx10); and (vi) by 0.16 as the ombrothermic index anomaly for humid years increased up to 0.36, after which the variable affect was negligible.

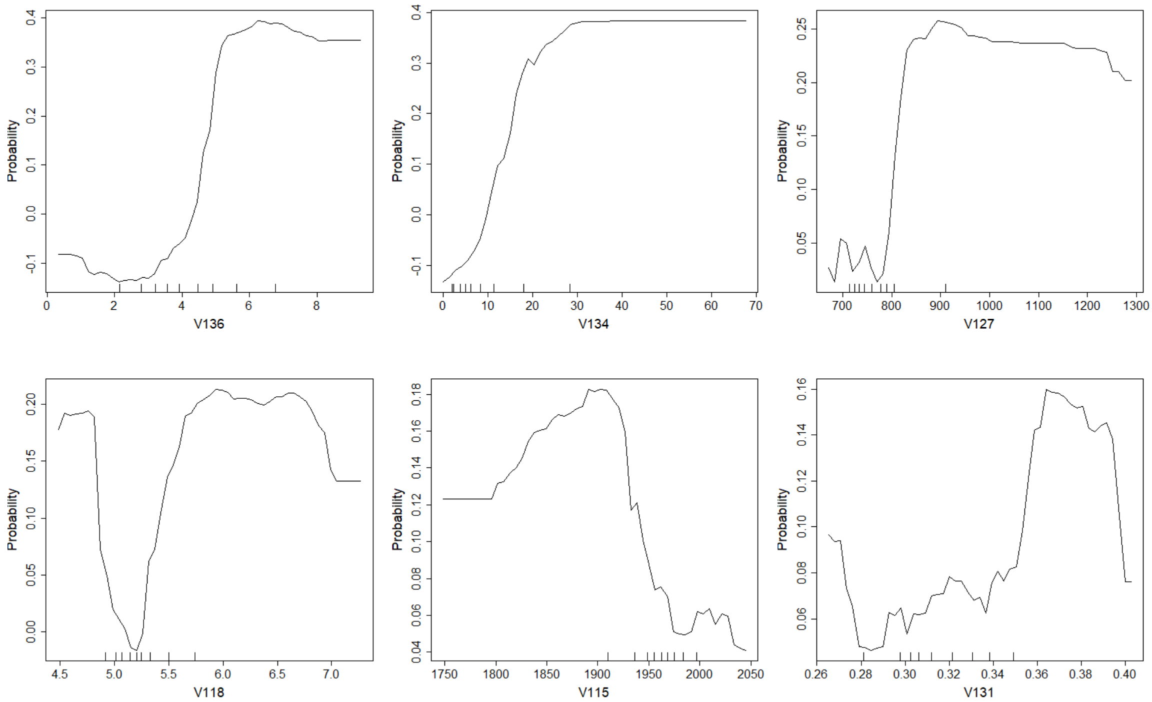


Figure 7.5. Response curves for the six most influential variables and their probability of explaining the land used for wheat plantations. See Table A1 in the Appendix for further details of the variables

Figure 7.6 displays the response curves for the first six most influential variables and their probability of explaining the land used for maize plantations. In particular, each plot shows that the probability of land being used for maize plantation increased on average: (i) by 0.8 as the slope increase up to 10%, after which the probability did not change; (ii) by 0.5 when the soil type was classified as "A" (incipient soils - modern non-limestone alluvisols of medium texture); (iii) by 0.6 as the drainage density increased up to 8 km/km^2, after which the probability did not change; (iv) by 0.6 as the ombrothermic index anomaly for humid years increased up to 0.34, after which the probability did not change; (v) by 0.4 as the ombrothermic index of the summer quarter plus the previous month increased up to 0.6, after which the variable effect reduced quickly, and the probability increases by 0.2 at >0.7 °C; and (vi) by 0.3 when the soil capacity was classified as "Ee" (very severe limitations).

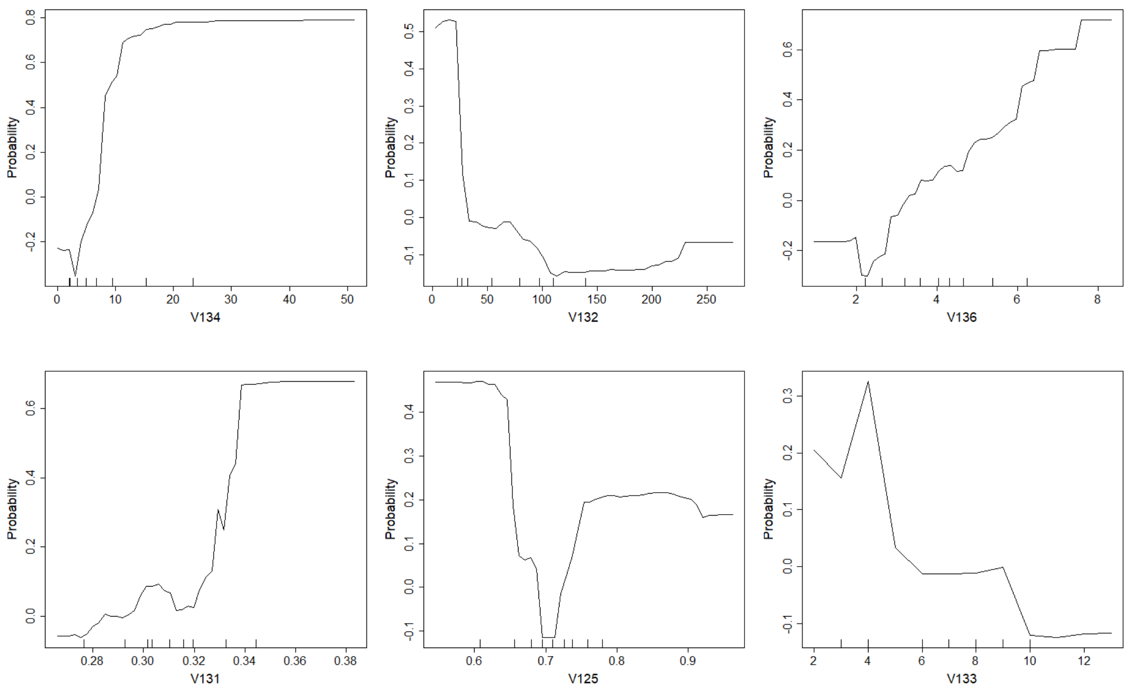


Figure 7.6. Response curves for the five most influential variables and their probability of explaining the land used for maize plantations. See Table A1 in the Appendix for further details of the variables

Figure 7.7 presents the response curves for the first six most influential variables and their threshold value probability of explaining the land used for olive grove plantations. In particular, each plot shows that the probability of land being used for olive grove plantations increased on average: (i) by 0.3 as the mean minimum temperature of the coldest month increases up to 6 ºC, after which the probability did not change; (ii) by 0.3 as the draining density increased up to 8 km/km^2, after which the variable affect reduced; (iii) by 0.25 as the ombrothermic index of the summer quarter increased up to 0.3, after which the variable affect reduced quickly; (iv) by 0.16 as the ombrothermic index anomaly for humid years increased up to 0.35, after which the variable affect reduced; (v) by 0.2 as positive precipitation for humid years increased up to 1000 mm, after which the probability did not change; and (vi) by 0.15 when the soil type was classified as “Ex” (incipient soils - lithosols of xeric regime climates, of schist or greywacke).

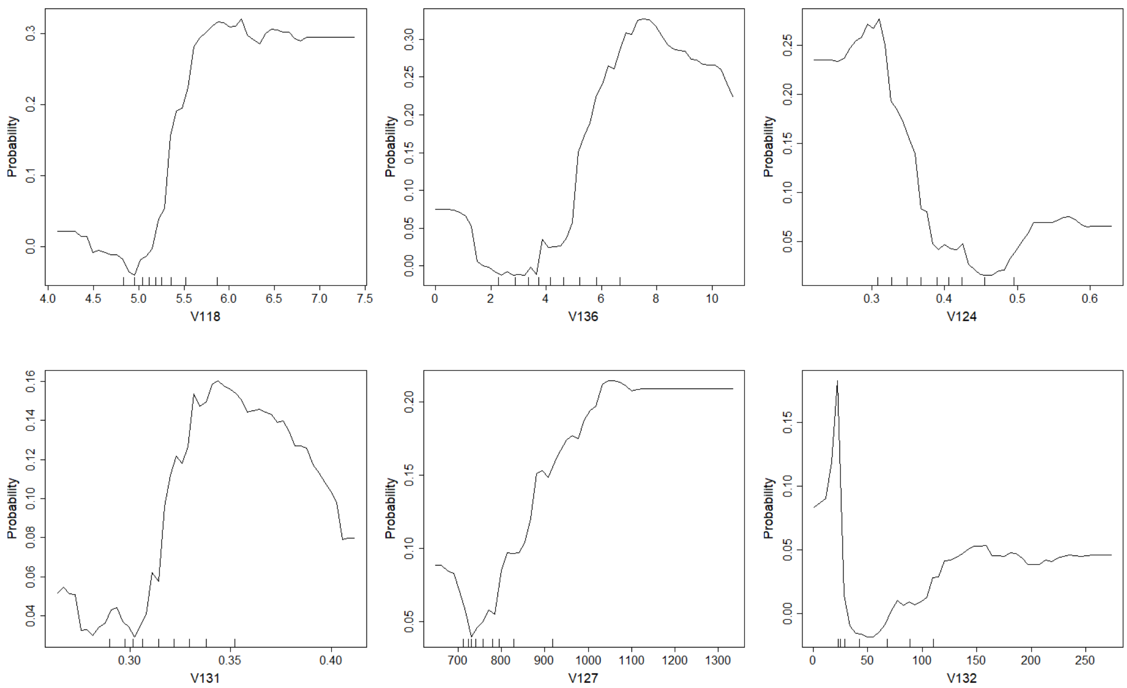


Figure 7.7. Response curves for the five most influential variables and their probability of explaining the land used for olive groves. See Table A1 in the Appendix for further details of the variables

The LIME method was employed to provide local interpretability (an explanation at the parcel level). Table 7.2 lists the best probability of four cases for each cropland (each case is a single parcel) and the top five variables that supported or contradicted the local explanation. Although most of the variables supported the explanation for single parcels used for wheat and olive groves, some variables contradicted the explanation (in parcel case #424 for wheat, and parcel cases #2138, #2534, and #2608 for olive groves). For maize, none of the top five variables contradicted this explanation. In general, for each cropland in the analysis, the top five variable sets were similar, although they weighted quite differently.

Table 7.2. Four principal cases of each cropland (the four with the best probability) and the respective five most influential variables that increased (supported) or decreased (contradicted) the explanation. See Table A1 in the Appendix for further details of the variables

Cropland	Case#	Probability	Top five variables from LIME
Wheat	539	0.99	Supports: V134, V136, V132, V53, V118
Wheat	424	0.98	Supports: V132, V136, V53, V134; Contradicts: V115
Wheat	283	0.97	Supports: V134, V132, V53, V115, V98
Wheat	406	0.96	Supports: V134, V136, V53, V132, V118
Maize	185	1.00	Supports:V134, V98, V91, V132, V131
Maize	152	0.99	Supports:V98, V134, V91, V136, V132
Maize	176	0.99	Supports:V98, V134, V91, V132, V121
Maize	153	0.99	Supports:V134, V98, V91, V132, V136
Olive groves	2138	1.00	Supports: V118, V97, V127, V124; Contradicts: V6
Olive groves	6531	0.99	Supports: V118, V136, V133, V124, V127
Olive groves	2534	0.99	Supports: V118, V136, V124, V95; Contradicts: V79
Olive groves	2608	0.99	Supports: V118, V97, V124, V132; Contradicts: V6

Figure 7.8 displays eight cases, whereby the explanation probability of a parcel was used for wheat plantations. It can be seen that slope (V134), soil type (V132), agricultural holdings with agricultural combine harvester machinery (V53), and drainage density (V136) supported the highest (largest positive weight) explanation probability. In particular, when the slope was <3.35% and the drainage density was between 2.85 km/km^2 and 3.93 km/km^2, there was a higher probability that a parcel was used for wheat plantation. At the local level, agricultural holdings with agricultural combine harvester machinery increased the probability, specifically >28 pieces of combine harvester machinery increased the probability of wheat plantations in all eight single parcels.

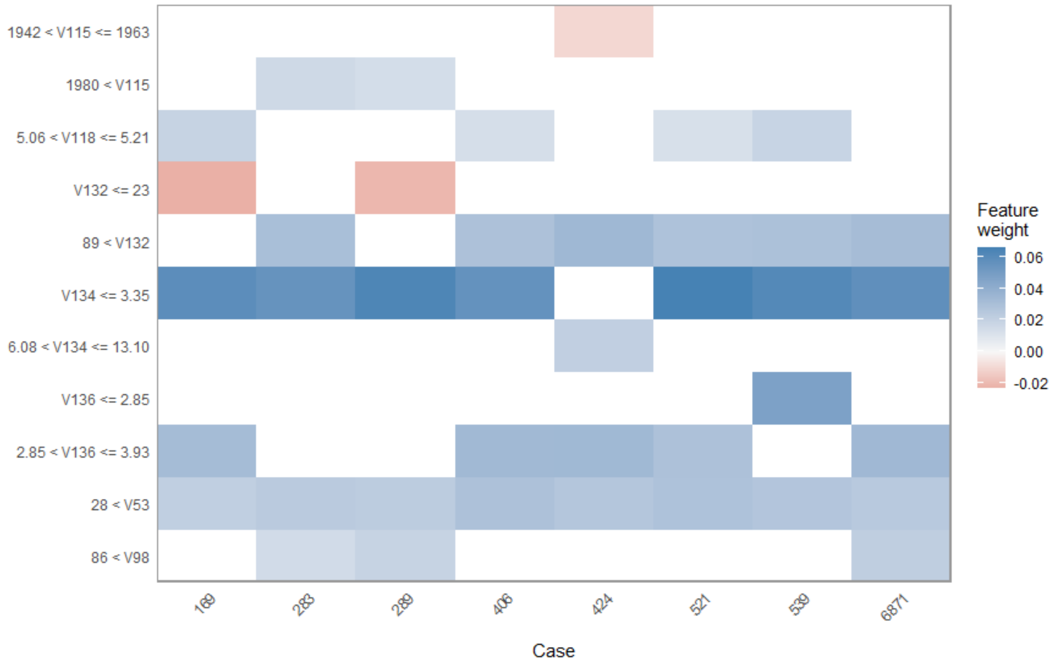


Figure 7.8. Most influential variables explaining the parcels used for wheat plantations. See Table A1 in the Appendix for further details of the variables

Figure 7.9 presents eight cases and the five most influential variables explaining the parcels used for maize plantations. It can be seen that slope (V134) and drainage density (V136) supported the explanation probability. In particular, when the slope was <2.24% and the drainage density was <2.64 km/km^2, there was a higher probability that a parcel was used for maize plantations. In addition, autonomous (legal form) utilised agricultural areas (V91) of ≤3213 ha and areas of permanent crops with fresh fruit plantations (V98) of <86 ha increased the probability of maize plantations in all eight single parcels in the analysis.

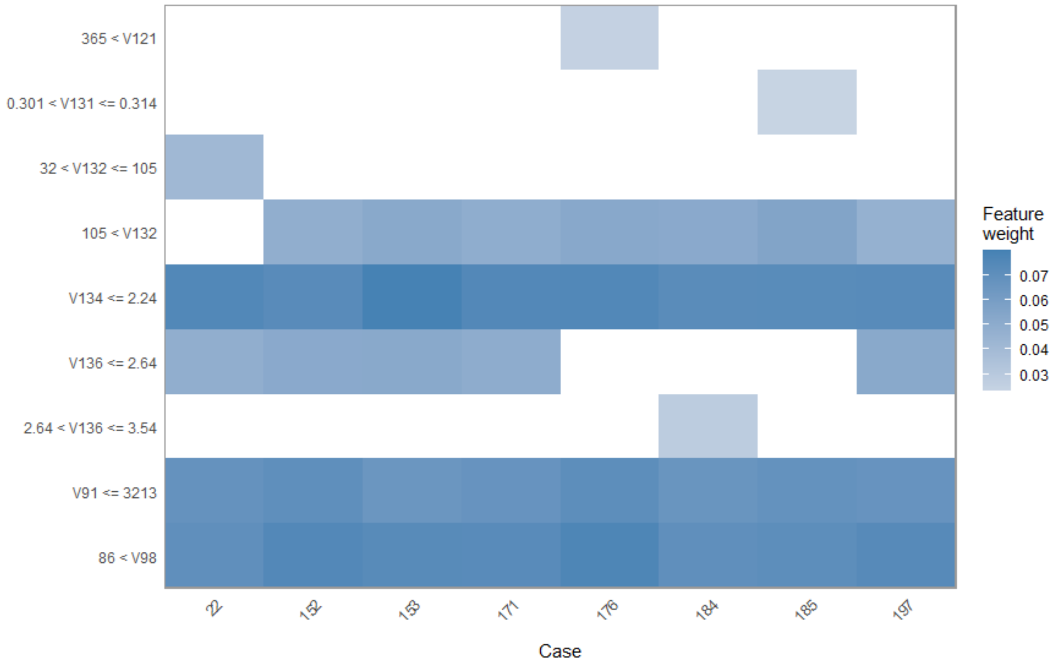


Figure 7.9. Most influential variables explaining the parcels used for maize plantations. See Table A1 in the Appendix for further details of the variables

Figure 7.10 shows eight cases and the most influential variables explaining the parcels used for olive groves plantation. It can be seen that when the mean minimum temperature of the coldest month (V118) was ≤4.99 ºC and the drainage density (V136) was between 3.09 km/km^2 and 4.10 km/km^2, the probability of a parcel being used for olive grove plantations increased. In addition, the area of permanent crops (V97) of >5568 ha increased the probability explanation. However, agricultural holdings with <662 (number) poultry (V6) and ≤10% of sole agricultural holders with a level of education outside the agricultural/forestry field (V79) contradicted the explanation. The variables explaining the parcels used for olive groves did not remain constant in all eight parcel cases in the analysis.

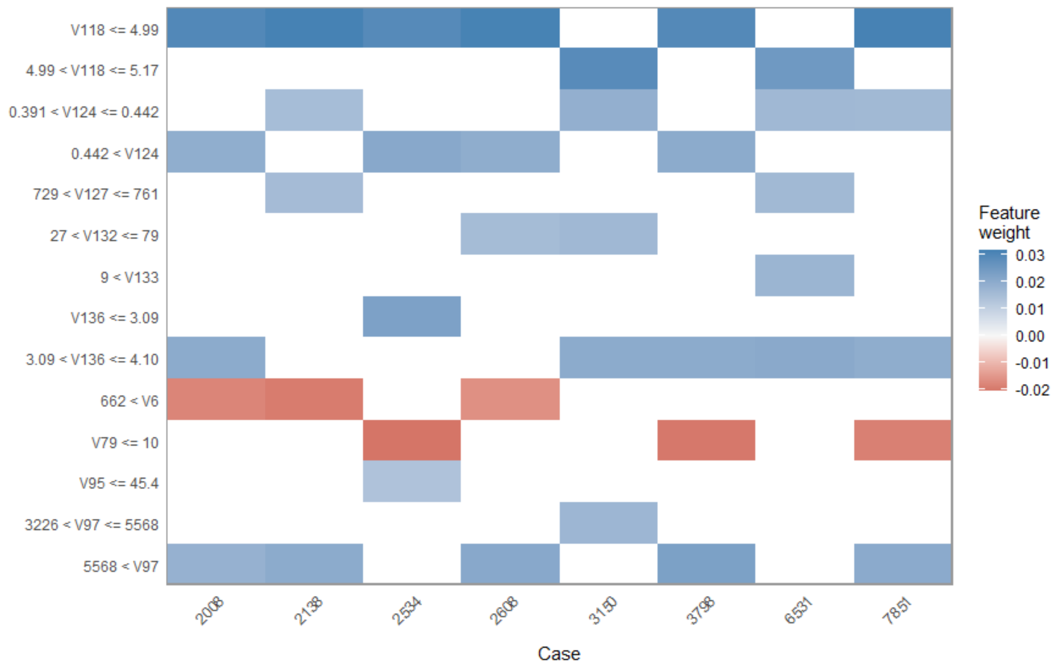


Figure 7.10. Most influential variables explaining the parcels used for olive groves plantation. See Table A1 in the Appendix for further details of the variables

7.4. Discussion

7.4.1. An ML and model-agnostic approach for agricultural land modelling

The modelling outcomes revealed the relationship between the use of agricultural land and environmental and socioeconomic conditions, showing that for each cropland under analysis, the explanatory factors varied significantly. At the global interpretability level, the results showed a highly dominant explanation of drainage density to the land used for wheat, maize, and olive grove plantations. However, the same variable not only exhibited different importance scores for the model interpretability of each cropland, but also presented different threshold values. For instance, in wheat plantations, the probability increased for a drainage density threshold value up to 5 km/km^2, whereas in maize and olive grove plantations, the explanation increased for a threshold up to 8 km/km^2. Overall, the model results emphasise that a high drainage density (>3.5 km/km^2) (Shankar & Mohan, 2006) is an important condition explaining the land used for these three crop plantations. These findings agree with those of other studies, which highlighted the importance of highly drained soils for the rooting depth of crops (Akpoti et al., 2019).

In addition, the slope could increase the explanation regarding wheat and maize plantations; however, the threshold values were substantially different (up to 30% for wheat and 10% for maize). Indeed, the slope of a plantation is a crucial factor for crop growth because it not only affects the vegetation structure but also the internal soil water drainage (Akpoti et al., 2019; Marcos-Martinez et al., 2017). Moreover, the mean minimum temperature of the coldest month explained most of the land used for olive grove plantations. For instance, exposure to cold temperatures is linked to the optimal differentiation of flower buds and the reduction of parasites and pathogens in olive trees (De Melo-Abreu et al., 2004; Rallo & Cuevas, 2017).

At the local interpretability level, the outcomes highlighted that socioeconomic factors became relevant, with differences observed with regard to variable explanation probability. For instance, the degree of mechanisation had a significant probability of supporting the explanation regarding wheat plantation in the eight parcels analysed (Ismail & Abdel-Mageed, 2010). Therefore, the outcome of the analysis indicates that, while environmental factors such as drainage density or slope were important for globally explaining the plantations of the three croplands in the study area, socioeconomic factors became equally important at the parcel level. These findings are consistent with those of other studies (Akpoti et al., 2019; Marcos-Martinez et al., 2017; Santiphop et al., 2012; Thenkabail, 2003).

The results of this study showed that the ML and model-agnostic methods could capture the complex interactions between the human–environmental processes influencing agricultural land use. The identification of the main variables that can explain the use of agricultural land for wheat, maize, and olive grove plantations helps to fill the knowledge gap for modelling these croplands, especially in southern Portugal. Therefore, in geographical regions with conditions similar to those of Mediterranean basins, such factors could be used to better characterise the suitable agricultural areas (Akpoti et al., 2019; Marcos-Martinez et al., 2017).

From a methodological point of view, our study suggests that the developed approach presents a high potential for use as an analytical method in the field of agricultural land systems. For example, the level of interpretability provided by the applied approach provides a reference for land suitability analysis (Akpoti et al., 2019). Certainly, the potential of such an approach deserves further development and testing for other spatiotemporal phenomena of land use to support planning strategies and more efficient and targeted land policies (e.g. research on the driving forces of LULC changes (Aburas et al., 2019)), and/or to anticipate and manage upcoming land changes due to variations in environmental and socioeconomic factors (Baessler & Klotz, 2006; Santiphop et al., 2012).

7.4.2. Limitations and recommendations

Based on our results and the considerations discussed above, we recognise that the development and implementation of the proposed approach had some limitations. First, some skills and knowledge of statistical software and methods were required. Second, although data availability has increased in many scientific fields, in this study, data related to the multiple factors affecting the use of agricultural land were limited to biophysical, bioclimatic, and socioeconomic factors. As such, the lacking of data related to for example political or cultural factors were a limitation of this study. In fact, having a large amount of data is an important element to strengthen the capability of ML modelling because it ensures adequate training and validation, thereby preventing generalisation problems (Carvalho et al., 2019). However, engaging scientific research with spatially explicit data depends on data availability, which can limit the use of ML models and must be a decisive factor to be considered in subsequent studies using such methods. Third, not having sufficient and representative sample sizes for the three analysed crop plantations can be detrimental to the model-agnostic results. This was, in fact, a second major constraint of this study, and it depicts the limitations associated with such an approach. Moreover, by being a model-agnostic approach, different types of explanations and degrees of interpretability regarding the factors potentially explaining agricultural land use may be obtained (Alvarez-Melis & Jaakkola, 2018; Carvalho et al., 2019; Murdoch et al., 2019; Slack et al., 2019). Therefore, the results should be interpreted critically by field experts who have improved knowledge regarding the underlying functions of agricultural land-use systems. Moreover, our findings need to be used at a regional scale because of their high spatial variability. Fourth, while this study builds upon timely and recent data to provide insights into the agricultural land used for wheat, maize, and olive grove plantations at a regional scale in a Portuguese district, we acknowledge that more studies should be carried out at different scales and across different geographic contexts to gain a deeper understanding of the underlying factors explaining other important and relevant croplands for global food security and ecosystem services (FAO, 2018). Fifth, and last, we used the RF ML algorithm, PFI, PDPs, and LIME model-agnostic methods, but substantial efforts have already been made in AI and xAI research fields, and different algorithms and methods are readily available (Carvalho et al., 2019; Molnar, 2019). Therefore, future research should focus on comparative studies that could guide new information and improve interpretation (Brun et al., 2020).

7.5. Conclusions

To comprehensively evaluate the factors that can potentially explain the use of agricultural land for wheat, maize, and olive grove plantations, this study implemented an ML and agnostic-model approach based on agricultural parcel-data sampled points and biophysical, bioclimatic, and socioeconomic variables. We applied global interpretable methods and identified that drainage density, slope, soil type, and the ombrothermic index anomaly (for humid and dry years) were the five most common important variables explaining the use of agricultural land for the three crop plantations in the study area. Using the local interpretable method, we explored spatial variations and found that socioeconomic variables became relevant at the parcel level. For instance, factors such as the degree of mechanisation could influence specific parcels of wheat. Overall, the analysis outcomes indicated that biophysical and bioclimatic conditions were more influential than socioeconomic conditions. As demonstrated, the proposed analytical approach may be particularly important for research on agricultural land use because it can capture the complex behaviours and underlying functions of agricultural land-use systems, thus providing crucial insights that can help solve several problems related to food production, social stability, and sustainable land use. We believe that this approach is a step towards providing comprehensive assessments of agricultural land use, and has the potential to serve as a decision-making tool to better plan and control the use of agricultural land. Despite the results we have achieved, more studies should be conducted at different scales and across different geographic contexts to gain a deeper understanding. Further research can be used to analyse the underlying factors explaining other important and relevant croplands for global food security and ecosystem services.

References

Abdi, A. M. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience and Remote Sensing, 57(1), 1–20. https://doi.org/10.1080/15481603.2019.1650447

Aburas, M. M., Ahamad, M. S. S., & Omar, N. Q. (2019). Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: a review. Environmental Monitoring and Assessment, 191(4), 1–28. https://doi.org/10.1007/s10661-019-7330-6

Akpoti, K., Kabo-bah, A. T., & Zwart, S. J. (2019). Agricultural land suitability analysis: State-of-the-art and outlooks for integration of climate change analysis. In Agricultural Systems (Vol. 173, pp. 172–208). Elsevier Ltd. https://doi.org/10.1016/j.agsy.2019.02.013

Al-Fugara, A., Pourghasemi, H. R., Al-Shabeeb, A. R., Habib, M., Al-Adamat, R., Al-Amoush, H., & Collins, A. L. (2020). A comparison of machine learning models for the mapping of groundwater spring potential. Environmental Earth Sciences 2020 79:10, 79(10), 1–19. https://doi.org/10.1007/S12665-020-08944-1

Alvarez-Melis, D., & Jaakkola, T. S. (2018, June 21). On the Robustness of Interpretability Methods. ICML Workshop on Human Interpretability in Machine Learning (WHI 2018).

Apley, D. W., & Zhu, J. (2016). Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 82(4), 1059–1086.

Baessler, C., & Klotz, S. (2006). Effects of changes in agricultural land-use on landscape structure and arable weed vegetation over the last 50 years. Agriculture, Ecosystems and Environment, 115(1–4), 43–50. https://doi.org/10.1016/j.agee.2005.12.007

Braimoh, A. K. (2009). Agricultural land-use change during economic reforms in Ghana. Land Use Policy, 26(3), 763–771. https://doi.org/10.1016/j.landusepol.2008.10.006

Brun, P., Thuiller, W., Chauvier, Y., Pellissier, L., Wüest, R. O., Wang, Z., & Zimmermann, N. E. (2020). Model complexity affects species distribution projections under climate change. Journal of Biogeography, 47(1), 130–142. https://doi.org/10.1111/JBI.13734

Cartone, A., & Postiglione, P. (2020). Principal component analysis for geographical data: the role of spatial effects in the definition of composite indicators. Spatial Economic Analysis, 1–22. https://doi.org/10.1080/17421772.2020.1775876

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine Learning Interpretability: A Survey on Methods and Metrics. Electronics, 8(8), 832. https://doi.org/10.3390/electronics8080832

Castillo, C. P., Jacobs-Crisioni, C., Diogo, V., & Lavalle, C. (2021). Modelling agricultural land abandonment in a fine spatial resolution multi-level land-use model: An application for the EU. Environmental Modelling and Software, 136, 104946. https://doi.org/10.1016/j.envsoft.2020.104946

Correia, T. P. (1993). Threatened landscape in Alentejo, Portugal: the ‘montado’ and other ‘agro-silvo-pastoral’ systems. Landscape and Urban Planning, 24(1–4), 43–48. https://doi.org/10.1016/0169-2046(93)90081-N

De Melo-Abreu, J. P., Barranco, D., Cordeiro, A. M., Tous, J., Rogado, B. M., & Villalobos, F. J. (2004). Modelling olive flowering date using chilling for dormancy release and thermal time. Agricultural and Forest Meteorology, 125(1–2), 117–127. https://doi.org/10.1016/j.agrformet.2004.02.009

Demšar, U., Harris, P., Brunsdon, C., Fotheringham, A. S., & McLoone, S. (2013). Principal Component Analysis on Spatial Data: An Overview. Annals of the Association of American Geographers, 103(1), 106–128. https://doi.org/10.1080/00045608.2012.689236

DGT. (2018). Especificações Técnicas da Carta de Uso e Ocupação do Solo (COS) de Portugal Continental para 1995, 2007, 2010 e 2015.

Dohoo, I. R., Ducrot, C., Fourichon, C., Donald, A., & Hurnik, D. (1997). An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. Preventive Veterinary Medicine, 29(3), 221–239. https://doi.org/10.1016/S0167-5877(96)01074-4

FAO. (2017). The Future of Food and Agriculture - Trends and Challenges.

FAO. (2018). Crop Prospects and Food Situation.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O’Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., … Zaks, D. P. M. (2011). Solutions for a cultivated planet. Nature, 478(7369), 337–342. https://doi.org/10.1038/nature10452

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. In Science (Vol. 327, Issue 5967, pp. 812–818). American Association for the Advancement of Science. https://doi.org/10.1126/science.1185383

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2013). Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation. Journal of Computational and Graphical Statistics, 24(1), 44–65.

Gomes, E., Banos, A., Abrantes, P., Rocha, J., Kristensen, S. B. P., & Busck, A. (2019). Agricultural land fragmentation analysis in a peri-urban context: From the past into the future. Ecological Indicators, 97, 380–388. https://doi.org/10.1016/j.ecolind.2018.10.025

Greenwell, B. (2017). pdp: An R Package for Constructing Partial Dependence Plots. The R Journal, 9(1), 421–436.

Hagenauer, J., Omrani, H., & Helbich, M. (2019). Assessing the performance of 38 machine learning models: the case of land consumption rates in Bavaria, Germany. International Journal of Geographical Information Science, 33(7), 1399–1419. https://doi.org/10.1080/13658816.2019.1579333

INE. (2012). Censos 2011 Resultados Definitivos – Região Alentejo. Instituto Nacional de Estatística.

Ismail, Z. E., & Abdel-Mageed, A. E. (2010). WORKABILITY AND MACHINERY PERFORMANCE FOR WHEAT HARVESTING. Misr Journal of Agricultural Engineering, 27(1), 90–103. https://doi.org/10.21608/mjae.2010.106860

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to Statistical Learning: With Applications in R. In Current medicinal chemistry (Corr. 7th printing, Vol. 7, Issue 10). Springer. https://doi.org/10.1007/978-1-4614-7138-7

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. Quality and Quantity, 52(4), 1957–1976. https://doi.org/10.1007/s11135-017-0584-6

Jombart, T., Devillard, S., Dufour, A. B., & Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity, 101(1), 92–103. https://doi.org/10.1038/hdy.2008.34

Koohafkan, P. (Parviz), & Altieri, M. A. (2016). Forgotten agricultural heritage : reconnecting food systems and sustainable development.

Kourgialas, N. N. (2021). A critical review of water resources in Greece: The key role of agricultural adaptation to climate-water effects. In Science of the Total Environment (Vol. 775, p. 145857). Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2021.145857

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Folke, C., Bruce, J. W., Coomes, O. T., Dirzo, R., George, P. S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E. F., Mortimore, M., Ramakrishnan, P. S., Richards, J. F., Steffen, W., … Veldkamp, T. A. (2001). The causes of land-use and land-cover change : moving beyond the myths. 11, 261–269.

Lehikoinen, A., Olsson, J., Bergström, L., Bergström, U., Bryhn, A., Fredriksson, R., & Uusitalo, L. (2019). Evaluating complex relationships between ecological indicators and environmental factors in the Baltic Sea: A machine learning approach. Ecological Indicators, 101, 117–125. https://doi.org/10.1016/j.ecolind.2018.12.053

Li, S., Juhász-Horváth, L., Pintér, L., Rounsevell, M. D. A., & Harrison, P. A. (2018). Modelling regional cropping patterns under scenarios of climate and socio-economic change in Hungary. Science of the Total Environment, 622–623, 1611–1620. https://doi.org/10.1016/j.scitotenv.2017.10.038

Li, X., Chen, W., Cheng, X., & Wang, L. (2016). A Comparison of Machine Learning Algorithms for Mapping of Complex Surface-Mined and Agricultural Landscapes Using ZiYuan-3 Stereo Satellite Imagery. In Remote Sensing (Vol. 8, Issue 6). https://doi.org/10.3390/rs8060514

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. R News, 2(3), 18–22.

Lima, A. R., Cannon, A. J., & Hsieh, W. W. (2015). Nonlinear regression in environmental sciences using extreme learning machines: A comparative evaluation. Environmental Modelling and Software, 73, 175–188. https://doi.org/10.1016/j.envsoft.2015.08.002

Lin, F. J. (2008). Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. Quality and Quantity, 42(3), 417–426. https://doi.org/10.1007/s11135-006-9055-1

Loumou, A., & Giourga, C. (2003). Olive groves: 'The life and identity of the Mediterranean'. Agriculture and Human Values, 20(1), 87–95. https://doi.org/10.1023/A:1022444005336

Luan, J., Zhang, C., Xu, B., Xue, Y., & Ren, Y. (2020). The predictive performances of random forest models with limited sample size and different species traits. Fisheries Research, 227, 105534. https://doi.org/10.1016/J.FISHRES.2020.105534

Marcos-Martinez, R., Bryan, B. A., Connor, J. D., & King, D. (2017). Agricultural land-use dynamics: Assessing the relative importance of socioeconomic and biophysical drivers for more targeted policy. Land Use Policy, 63, 53–66. https://doi.org/10.1016/j.landusepol.2017.01.011

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environmental Modelling and Software, 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001

Moghaddam, D. D., Rahmati, O., Panahi, M., Tiefenbacher, J., Darabi, H., Haghizadeh, A., Haghighi, A. T., Nalivan, O. A., & Tien Bui, D. (2020). The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers. CATENA, 187, 104421. https://doi.org/10.1016/J.CATENA.2019.104421

Molnar, C. (2019). Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. Available online: https://christophm.github.io/interpretable-ml-book/ (accessed on 22 January 2021).

Monteiro-Henriques, T., Martins, M. J., Cerdeira, J. O., Silva, P., Arsénio, P., Silva, Á., Bellu, A., & Costa, J. C. (2016). Bioclimatological mapping tackling uncertainty propagation: application to mainland Portugal. International Journal of Climatology, 36(1), 400–411. https://doi.org/10.1002/joc.4357

Muñoz-Rojas, J., Pinto-Correia, T., Hvarregaard Thorsoe, M., & Noe, E. (2019). The Portuguese Montado : A Complex System under Tension between Different Land Use Management Paradigms . In Silvicultures - Management and Conservation. IntechOpen. https://doi.org/10.5772/intechopen.86102

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. Proceedings of the National Academy of Sciences of the United States of America, 116(44), 22071–22080. https://doi.org/10.1073/pnas.1900654116

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? Ecography, 37(2), 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Ndamani, F., & Watanabe, T. (2017). Developing indicators for adaptation decision-making under climate change in agriculture: A proposed evaluation model. Ecological Indicators, 76, 366–375. https://doi.org/10.1016/j.ecolind.2016.12.012

Pedersen, T. L., & Benesty, M. (2019). lime: Local interpretable model-agnostic explanations.

Petit, C., Aubry, C., & Rémy-Hall, E. (2011). Agriculture and proximity to roads: How should farmers and retailers adapt? Examples from the Ile-de-France region. Land Use Policy, 28(4), 867–876. https://doi.org/10.1016/j.landusepol.2011.03.001

Portugal, S. (2019). Recenseamento Agrícola. https://ra09.ine.pt/

Postek, P., Leń, P., & Stręk, Ż. (2019). The proposed indicator of fragmentation of agricultural land. Ecological Indicators, 103, 581–588. https://doi.org/10.1016/j.ecolind.2019.04.023

Probst, P., & Boulesteix, A.-L. (2018). To Tune or Not to Tune the Number of Trees in Random Forest. In Journal of Machine Learning Research (Vol. 18). https://doi.org/10.5555/3122009.3242038

Probst, P., & Janitza, S. (2020). varImp: RF Variable Importance for Arbitrary Measures.

Qi, Y. (2012). Random Forest for Bioinformatics. Ensemble Machine Learning, 307–323. https://doi.org/10.1007/978-1-4419-9326-7_11

Raczko, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. European Journal of Remote Sensing, 50(1), 144–154. https://doi.org/10.1080/22797254.2017.1299557

Radwan, T. M., Blackburn, G. A., Whyatt, J. D., & Atkinson, P. M. (2019). Dramatic Loss of Agricultural Land Due to Urban Expansion Threatens Food Security in the Nile Delta, Egypt. Remote Sensing, 11(3), 332. https://doi.org/10.3390/rs11030332

Rallo, L., & Cuevas, J. (2017). Fructificación y producción (Fruiting and production). In D. Barranco, R. Fernandez-Escobar, & L. Rallo (Eds.), El Cultivo del Olivo (Olive Growing) (pp. 145–186). Mundi-Prensa.

Ren, X., Mi, Z., & Georgopoulos, P. G. (2020). Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution: Modeling ozone concentrations across the contiguous United States. Environment International, 142, 105827. https://doi.org/10.1016/j.envint.2020.105827

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016a, June 16). Model-Agnostic Interpretability of Machine Learning. ICML Workshop on Human Interpretability in Machine Learning.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016b). 'Why should i trust you?' Explaining the predictions of any classifier. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016, 1135–1144. https://doi.org/10.1145/2939672.2939778

Salami, D., Sousa, C. A., Martins, M. do R. O., & Capinha, C. (2020). Predicting dengue importation into Europe, using machine learning and model-agnostic methods. Scientific Reports, 10(1), 1–13. https://doi.org/10.1038/s41598-020-66650-1

Santiphop, T., Shrestha, R. P., & Hazarika, M. K. (2012). An analysis of factors affecting agricultural land use patterns and livelihood strategies of farm households in Kanchanaburi Province, Thailand. Journal of Land Use Science, 7(3), 331–348. https://doi.org/10.1080/1747423X.2011.587208

Shankar, M. N. R., & Mohan, G. (2006). Assessment of the groundwater potential and quality in Bhatsa and Kalu river basins of Thane district, western Deccan Volcanic Province of India. Environmental Geology, 49(7), 990–998. https://doi.org/10.1007/s00254-005-0137-5

Shortridge, J. E., Guikema, S. D., & Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrology and Earth System Sciences, 20(7), 2611–2628. https://doi.org/10.5194/hess-20-2611-2016

Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2019). Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods. AIES 2020 - Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 180–186.

Thenkabail, P. S. (2003). Biophysical and yield information for precision farming from near-real-time and historical Landsat TM images. International Journal of Remote Sensing, 24(14), 2879–2904. https://doi.org/10.1080/01431160710155974

Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D. J., & Franz, K. J. (2020). Mapping of 30-meter resolution tile-drained croplands using a geospatial modeling approach. Scientific Data, 7(1), 1–10. https://doi.org/10.1038/s41597-020-00596-x

Velásquez-Milla, D., Casas, A., Torres-Guevara, J., & Cruz-Soriano, A. (2011). Ecological and socio-cultural factors influencing in situ conservation of crop diversity by traditional Andean households in Peru. Journal of Ethnobiology and Ethnomedicine, 7, 40. https://doi.org/10.1186/1746-4269-7-40

Viana, C. M., Girão, I., & Rocha, J. (2019). Long-Term Satellite Image Time-Series for Land Use/Land Cover Change Detection Using Refined Open Source Data in a Rural Region. Remote Sensing, 11(9), 1104. https://doi.org/10.3390/rs11091104

Viana, C. M., & Rocha, J. (2020). Evaluating Dominant Land Use/Land Cover Changes and Predicting Future Scenario in a Rural Region Using a Memoryless Stochastic Method. Sustainability, 12(10), 4332. https://doi.org/10.3390/su12104332

Winkler, A. M., Webster, M. A., Vidaurre, D., Nichols, T. E., & Smith, S. M. (2015). Multi-level block permutation. NeuroImage, 123, 253–268. https://doi.org/10.1016/j.neuroimage.2015.05.092

Wu, L., Zhu, X., Lawes, R., Dunkerley, D., & Zhang, H. (2019). Comparison of machine learning algorithms for classification of LiDAR points for characterization of canola canopy structure. International Journal of Remote Sensing, 40(15), 5973–5991. https://doi.org/10.1080/01431161.2019.1584929

Wu, W. B., Yu, Q. Y., Verburg, P. H., You, L. Z., Yang, P., & Tang, H. J. (2014). How could agricultural land systems contribute to raise food production under global change? Journal of Integrative Agriculture, 13(7), 1432–1442. https://doi.org/10.1016/S2095-3119(14)60819-4

Yang, R. M., Zhang, G. L., Liu, F., Lu, Y. Y., Yang, F., Yang, F., Yang, M., Zhao, Y. G., & Li, D. C. (2016). Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. Ecological Indicators, 60, 870–878. https://doi.org/10.1016/J.ECOLIND.2015.08.036

Zhang, Z., Xu, S., Capinha, C., Weterings, R., & Gao, T. (2019). Using species distribution model to predict the impact of climate change on the potential distribution of Japanese whiting Sillago japonica. Ecological Indicators, 104, 333–340. https://doi.org/10.1016/j.ecolind.2019.05.023

Appendix

Table A1. Variables included in the model

Code	Variable	Unit	Year	Source
Socioeconomic				
V1	Livestock of agricultural holdings by animal species (Cattle)	Nº		
V2	Livestock of agricultural holdings by animal species (Pigs)	Nº		
V3	Livestock of agricultural holdings by animal species (Sheep)	Nº		
V4	Livestock of agricultural holdings by animal species (Goats)	Nº		
V5	Livestock of agricultural holdings by animal species (Equidae)	Nº		
V6	Livestock of agricultural holdings by animal species (Poultry)	Nº		
V7	Livestock of agricultural holdings by animal species (Rabbits)	Nº		
V8	Livestock of agricultural holdings by animal species (Inhabited hives and traditional cork hives)	Nº		
V9	Agricultural holdings by type of tenure (total utilised agricultural area)	Nº		
V10	Agricultural holdings by type of tenure (Owner farming	Nº		
V11	Agricultural holdings by type of tenure (Tenant farming	Nº		
V12	Agricultural holdings by type of tenure (Other modes of tenure	Nº		
V13	Agricultural holdings by legal form (Natural person (sole holder))	Nº		
V14	Agricultural holdings by legal form (Autonomous)	Nº		
V15	Agricultural holdings by legal form (Businessman)	Nº		
V16	Agricultural holdings by legal form (group holding)	Nº		
V17	Agricultural holdings by type of land use (Total)	Nº		
V18	Agricultural holdings by type of land use (Utilised agriculture area)	Nº	2019	Statistics
V19	Agricultural holdings by type of land use (Wooded area)	Nº		Portugal (2019)
V20	Agricultural holdings by type of land use (Unutilised agricultural land (UAL))	Nº		
V21	Agricultural holdings by type of land use (Other land)	Nº		
V22	Agricultural holdings with permanent crops by type (Total)	Nº		
V23	Agricultural holdings with permanent crops by type (Fresh fruit plantations (excluding citrus plantations))	Nº		
V24	Agricultural holdings with permanent crops by type (Citrus plantations)	Nº		
V25	Agricultural holdings with permanent crops by type (Fruit plantations (subtropical climate zones))	Nº		
V26	Agricultural holdings with permanent crops by type (Nuts plantations)	Nº		
V27	Agricultural holdings with permanent crops by type (Olive plantations)	Nº		
V28	Agricultural holdings with permanent crops by type (Vineyards)	Nº		
V29	Agricultural holdings with permanent crops by type (Other permanent crops)	Nº		
V30	Agricultural holdings with temporary crops by type (Total)	Nº		
V31	Agricultural holdings with temporary crops by type (Cereals)	Nº		
V32	Agricultural holdings with temporary crops by type (Dried pulses)	Nº		

V33	Agricultural holdings with temporary crops by type (Temporary grasses and grazings)	Nº
V34	Agricultural holdings with temporary crops by type (Fodder plants)	Nº
V35	Agricultural holdings with temporary crops by type (Potatoes)	Nº
V36	Agricultural holdings with temporary crops by type (Industrial crops)	Nº
V37	Agricultural holdings with temporary crops by type (Fresh vegetables)	Nº
V38	Agricultural holdings with temporary crops by type (Flowers and ornamental plants)	Nº
V39	Agricultural holdings with temporary crops by type (Other temporary crops)	Nº
V40	Agricultural holdings with livestock by animal species (Cattle)	Nº
V41	Agricultural holdings with livestock by animal species (Pigs)	Nº
V42	Agricultural holdings with livestock by animal species (Sheep)	Nº
V43	Agricultural holdings with livestock by animal species (Goats)	Nº
V44	Agricultural holdings with livestock by animal species (Equidae)	Nº
V45	Agricultural holdings with livestock by animal species (Poultry)	Nº
V46	Agricultural holdings with livestock by animal species (Rabbits)	Nº
V47	Agricultural holdings with livestock by animal species (Inhabited hives and traditional cork hives)	Nº
V48	Agricultural holdings with agricultural machineries by type of agricultural machinery (Total)	Nº
V49	Agricultural holdings with agricultural machineries by type of agricultural machinery (Tractors)	Nº
V50	Agricultural holdings with agricultural machineries by type of agricultural machinery (Walking cultivators)	Nº
V51	Agricultural holdings with agricultural machineries by type of agricultural machinery (Motor hoes (rotovators))	Nº
V52	Agricultural holdings with agricultural machineries by type of agricultural machinery (Motormowers)	Nº
V53	Agricultural holdings with agricultural machineries by type of agricultural machinery (Combine harvester)	Nº
V54	Agricultural holdings with permanent grassland and meadow (Total)	Nº
V55	Agricultural holdings with irrigable area (Total)	Nº
V56	Agricultural machineries by type of agricultural machinery (Total)	Nº
V57	Agricultural machineries by type of agricultural machinery (Tractors)	Nº
V58	Agricultural machineries by type of agricultural machinery (Walking cultivators)	Nº
V59	Agricultural machineries by type of agricultural machinery (Motor hoes (rotovators))	Nº
V60	Agricultural machineries by type of agricultural machinery (Motormowers)	Nº
V61	Agricultural machineries by type of agricultural machinery (Combine harvester)	Nº
V62	Familiar agricultural population (Total)	Nº
V63	Sole agricultural holders by Age group (Total)	Nº
V64	Sole agricultural holders by Age group (15 - 24 years)	Nº
V65	Sole agricultural holders by Age group (25 - 34 years)	Nº
V66	Sole agricultural holders by Age group (35 - 44 years)	Nº
V67	Sole agricultural holders by Age group (45 - 54 years)	Nº
V68	Sole agricultural holders by Age group (55 - 64 years)	Nº
V69	Sole agricultural holders by Age group (65 and more years)	Nº
V70	Sole agricultural holders with 65 and more years old (Total)	Nº
V71	Sole agricultural holders with remunerate activities outside agricultural holding (Total)	Nº
V72	Sole agricultural holders of female sex (Total)	Nº
V73	Proportion of sole agricultural holders by level of education (None)	%
V74	Proportion of sole agricultural holders by level of education (Do not know to read and write)	%
V75	Proportion of sole agricultural holders by level of education (Can read and write)	%
V76	Proportion of sole agricultural holders by level of education (Basic)	%
V77	Proportion of sole agricultural holders by level of education (Secondary / post-secondary)	%
V78	Proportion of sole agricultural holders by level of education (Agricultural/Forestry)	%
V79	Proportion of sole agricultural holders by level of education (Non-agricultural/Non-forestry)	%

V80	Proportion of sole agricultural holders by level of education (Superior)	%		
V81	Proportion of sole agricultural holders by level of education (Agricultural/Forestry)	%		
V82	Proportion of sole agricultural holders by level of education (Non-agricultural/Non-forestry)	%		
V83	Proportion of sole agricultural holders by time of agricultural activity on agricultural holding (225 days/ 1800 hours/year)	%		
V84	Proportion of sole agricultural holders by time of agricultural activity on agricultural holding (full time)	%		
V85	Proportion of sole agricultural holders by time of agricultural activity on agricultural holding (parcial time)	%		
V86	Proportion of sole agricultural holders by time of agricultural activity on agricultural holding (> 0 - < 25%)	%		
V87	Proportion of sole agricultural holders by time of agricultural activity on agricultural holding (25 - < 50%)	%		
V88	Proportion of sole agricultural holders by time of agricultural activity on agricultural holding (50 - < 75%)	%		
V89	Proportion of sole agricultural holders by time of agricultural activity on agricultural holding (75 - < 100%)	%		
V90	Utilized agricultural area by legal form (Natural person (sole holder))	ha		
V91	Utilized agricultural area by legal form (Autonomous)	ha		
V92	Utilized agricultural area by legal form (Businessman)	ha		
V93	Utilized agricultural area by legal form (Group of natural persons (group holding))	ha		
V94	Utilized agricultural area by legal form (Group of natural persons (others group holding))	ha		
V95	Average utilized agricultural area per holding (Total)	ha		
V96	Utilized agricultural area per annual work unit (Total)	ha		
V97	Area of permanent crops by type (Total)	ha		
V98	Area of permanent crops by type (Fresh fruit plantations (excluding citrus plantations))	ha		
V99	Area of permanent crops by type (Citrus plantations)	ha		
V100	Area of permanent crops by type (Olive plantations)	ha		
V101	Area of permanent crops by type (Vineyards)	ha		
V102	Area of temporary crops by type (Total)	ha		
V103	Area of temporary crops by type (Cereals)	ha		
V104	Area of temporary crops by type (Dried pulses)	ha		
V105	Area of temporary crops by type (Temporary grasses and grazings)	ha		
V106	Area of temporary crops by type (Fodder plants)	ha		
V107	Area of temporary crops by type (Potatoes)	ha		
V108	Area of temporary crops by type (Industrial crops)	ha		
V109	Area of temporary crops by type (Fresh vegetables)	ha		
V110	Area of permanent grassland and meadow (Total)	ha		
V111	Irrigable area of agricultural holdings (Total)	ha		
V112	Average annual work unit by agricultural holding (AWU -Total)	ha		
Bioclimatic				
V113	Mean temperature of the warmest month of the year	ºC		
V114	Mean temperature of the coldest month of the year	ºC		
V115	Annual positive temperature	ºCx10		
V116	Positive precipitation	mm		
V117	Mean maximum temperature of the coldest month	ºC		
V118	Mean minimum temperature of the coldest month	ºC		
V119	Simple continentality index, or annual thermal amplitude	index		
V120	Thermicity index	index		
V121	Compensated thermicity index	index		
V122	Annual ombrothermic index	index	1960-1990	Monteiro-Henriques et al. (2016)
V123	Ombrothermic index of the warmest bimonth of the summer quarter	index		
V124	Ombrothermic index of the summer quarter	index		
V125	Ombrothermic index of the summer quarter plus the previous month	index		
V126	Positive precipitation for dry year	mm		
V127	Positive precipitation for humid year	mm		
V128	Ombrothermic index for dry year	index		
V129	Ombrothermic index for humid year	index		
V130	Ombrothermic index anomaly for dry year	index		
V131	Ombrothermic index anomaly for humid year	index		
Biophysical				
V132	Soil type	NA		DGADR

V133	Soil capacity	NA	Static over time	DGADR
V134	Slope	%		IGEOE
V135	Distance to water bodies	m		DGT
V136	Draining density	km/km2		IGEOE
V137	Distance to urban	m		DGT
V138	Distance to roads	m		OpenStreetMap

Table A2. Included variables with VIF <5

Wheat		Maize		Olive groves	
V4	V92	V8	V136	V4	V84
V5	V95	V75	V137	V5	V87
V7	V96	V77	V138	V6	V88
V8	V97	V78		V7	V89
V39	V98	V82		V8	V95
V41	V99	V87		V9	V96
V47	V115	V88		V33	V97
V50	V118	V89		V35	V99
V52	V123	V91		V36	V115
V53	V124	V92		V38	V118
V64	V127	V95		V39	V123
V65	V130	V98		V47	V124
V74	V131	V116		V50	V127
V75	V132	V121		V59	V130
V77	V133	V123		V60	V131
V78	V134	V125		V64	V132
V82	V135	V130		V65	V133
V84	V136	V131		V74	V134
V86	V137	V132		V75	V135
V87	V138	V133		V78	V136
V88		V134		V79	V137
V89		V135		V82	V138

Part V. Dynamics of agricultural production: modelling spatial distributions across time

Chapter 8. Evolution of agricultural production in Portugal during 1850-2018: a geographical and historical perspective

Cláudia M. Viana, Dulce Freire, Patrícia Abrantes, Jorge Rocha

Article published at the Land (2021), Volume 10, Part 8, 776.

https://doi.org/10.3390/land10080776

The structure and formatting are slightly adapted.

Abstract

Agricultural statistical data enable the detection and interpretation of the development of agriculture and the food supply situation over time, which is essential for food security evaluation in any country. Based on the historical agricultural statistics, this study produces a long spatial time-series with annual production values of three cereals relevant to global food security—wheat, maize, and rice, aiming to provide geographical and historical perspectives. Therefore, we reconstructed past and current production patterns and trends at the district level over 169 years, which supported a space–time cross-reading of the general characteristics of the regional agricultural production value distributions and relative densities in Portugal. Particularly, the production trends of wheat, maize, and rice showed three different situations: growth (maize), stability (rice), and decline (wheat). For decades, maize and wheat production alternated, depending on agricultural years and political aspects, such as the Wheat Campaign (1929–1938). The changes over time presented a pattern that, in the case of these three cereals, enabled a clear division of the country into major regions according to cereal production. Overall, maize and rice, both grown on irrigated croplands, presented a similar pattern in some regions of Portugal, mainly the central region. In this study, a preliminary analysis was presented and related to successive public policies; however, notably, there are more lessons to be learned from this long spatial time-series.

Keywords: historical data; census; agricultural statistics; spatial analysis; time-series; cereal production

8.1. Introduction

Since at least the first half of the nineteenth century, agricultural statistics have been collected and reported by most countries in Western Europe (Mitchell, 1998, 2011). This data play an important role in academic discussions and in planning a sustainable future as they provide insights into the development of agriculture and food supply situation over time (Calleja et al., 2012; Eurostat., 2020; Westlund & Nilsson, 2019), which is essential for food security evaluation in any country. However, agricultural statistics data of many countries present limitations due to the variations in administrative units (i.e., geographic boundary changes for the attribution of data), which hampers in-depth long-term interpretations of agricultural development and also prevents comparison in terms of the evolution of regional productions, that is, what was produced, in which locations, and in which amount.

Considering the example of Portugal, public administration services have been collecting data every year since the mid-1850s. In terms of quantitative data, currentl, there are 140 years (1850–1989) of agricultural production information available at the district level. However, agricultural statistical data for the period after 1989 were collected and published linked to two different administrative units: from 1990 to 2005 at the agrarian region level, and since 2006 at the Nomenclature of Territorial Units for Statistics (NUTS) II level. Due to the incompatibility of these administrative units, a long-term analysis that also includes the last 30 years, reflecting the regional production dynamics and oscillations, was not possible to reach. To date, several authors have been conditioned to use aggregated statistical series at the national level (Freire & Lains, 2017; Justino, 1988; Lains & Silveira E Sousa, 1998; Pereira, 1983).

As phenomena occur at multiple spatial and temporal scales, the creation of a long spatial time-series is essential, and can strengthen the ability to assess past occurrences via scientific analysis and provide insights in order to form more effective solutions for current and future phenomena faced by society (Boivin & Crowther, 2021; Gregory & Ell, 2005; Murrieta-Flores & Martins, 2019). Therefore, for different scientific subjects, it is crucial to contour the limitations of statistical data that hamper the creation of a long-time series, and the ability of cross-timescale analysis to detect spatiotemporal patterns and trends, as well as to forecast future events (Gregory et al., 2001; Gregory & Ell, 2005; Murrieta-Flores & Martins, 2019). Accordingly, considering the importance of agricultural statistical data in reinterpreting both past and current agriculture and food supply situations, this study focused on the Portuguese agricultural statistics data to find solutions in order to homogenize the data at the upper scales and to create a long spatial time-series to provide geographical and historical perspectives.

Why analyze the Portuguese context? Because Portugal was one of the first European countries to published reliable agricultural statistical data since at least the last century and a half, at regular intervals, and linked only to three different administrative units. In addition, agriculture was the most important economic sector in Portugal until the 1960s, and despite being a relatively small European country, the presence of very distinct edaphoclimatic and agroecological conditions in the northern, central, and southern regions enables the cultivation of various agricultural products (e.g., wheat, maize, rice), with each one having an important role in the primary sector. Thus, the main objectives of the present study were threefold: (i) apply a method to overcome the variation in the administrative units of the Portuguese agricultural statistical data to create a long spatial time-series with annual agricultural production values at the district level for three cereals crops (wheat, maize, and rice); (ii) provide geographical and historical perspectives on each cereal production value distributions and relative density from 1850 to 2018; and (iii) analyze the historical spatiotemporal patterns and trends in regional cereal production over 169 years.

8.2. Materials and Methods

8.2.1. Study Area

Mainland Portugal is located on the western coast of continental Europe, sharing its North and East borders with Spain and being bounded by the Atlantic Ocean to the West and South. In the territory of Portugal, to the northwest, the landscape is mountainous and is characterized by the abundance of water and existence of fertile soils, and the property is structured around the minifundium. In the southern interior, in the direction of Algarve, open rolling plains and granite hills characterize the relief, with water scarcity, poor soils, and agriculture that developed around the latifundium (Ribeiro et al., 1988). For statistical purposes, the Portuguese government uses the NUTS system (NUTS is a geographical nomenclature subdividing the economic territory of the European Union into three different levels, NUTS I, II, and III, respectively, moving from larger to smaller territorial units) and, informally, the district system (established in 1835 and used until European integration). This study was conducted for the 18 districts of mainland Portugal (Figure 8.1).

A considerable part of mainland Portugal is dedicated to agriculture, although the importance of agriculture varies between the different regions due to the edaphoclimatic and agroecological conditions (climate, soil fertility, water availability, and land structure, among others). In terms of the main products, the northern region is known for vineyards in the Douro Valley, while the southern region, specifically in Alentejo, developed an extensive monoculture of cereals and olive trees. Other important productions includes fruits, namely oranges from Algarve, cherries from the Central region, and pêra rocha from the Oeste region, and a wide variety of crops, namely green vegetables, oilseeds, nuts, and cork. Nonetheless, the primary sector in Portugal was the most important until the 1960s, when the population working in agriculture fell rapidly from 42% in 1960 to 32% in 1970. This decline continues to the present day, with just about 12% in agriculture (Lains, 2003).

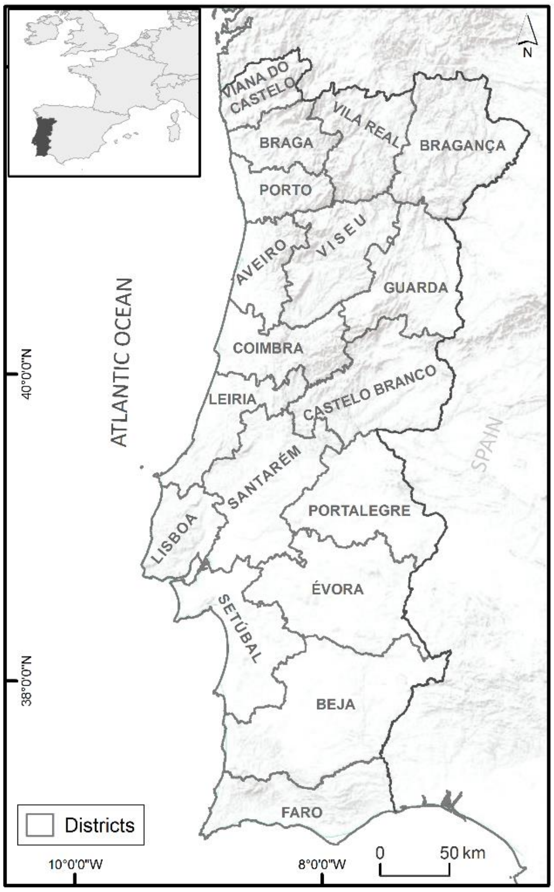


Figure 8.1. The study area (Portuguese districts)

8.2.2. Portuguese agricultural statistical data

From the second half of the nineteenth century until the mid-1960s the Portuguese Civil Governors started annually reporting quantitative information on agriculture statistical data. Thereafter, the Statistics Office, and subsequent organizations, were responsible for the publication of an official statistics repository: the Portuguese Statistical Yearbook. In 1935 Statistics Portugal (INE) was founded and assumed these functions, and from 1943 to now, the INE has published quantitative information on agriculture statistical data annually in the Agricultural Statistics book. Thus, when consulting various Portuguese statistical yearbooks and the annual Agricultural Statistics, it is possible to gather about 70 years of long-term continuous time-series at the district level (1916 to 1989) (Lains & Silveira E Sousa, 1998).

However, agricultural statistical data for the period before 1916 and for the period after 1989 present two distinct problems: firstly, the data before 1916 were scattered across the central and regional Portuguese public archives. As such, to gather the data, an interdisciplinary research team of the 'Agriculture in Portugal: food, development, and sustainability (1870-2010)' project, funded by the Portuguese Foundation for science and technology (FCT), had to work for three years by looking at thousands of documents deposited in the National Archives of Torre do Tombo (formerly known as the General Archive of the Kingdom, an organic nucleus of the General Directorate of Books, Archives and Libraries, which has been the central archive of the Portuguese State since the Middle Ages being more than 600 years old), and the Arquivos Distritais (the Arquivos Distritais of Portugal has the mission of preserving and valuing the archival heritage of historical interest). The research team was able to gather agricultural data at the district level with annual production values from 1850 to 1915 for different agricultural products. It should be noted the first version of the database had an absence of certain annual production values for specific products. Since the database had a sufficient temporal continuity, which allowed a decrease in the margin of error and did not affect the attributes of the data (Thomas, 1980; van Zanden & van Leeuwen, 2012), the team applied the linear interpolation technique to fill the gaps. Secondly, the agricultural statistical data for the period after 1989 were published at different administrative units over the years: from 1990 to 2005 at the agrarian region level (AR) (the agrarian regions are agricultural statistical regions, there are seven agrarian units in Portugal mainland), and from 2006 until 2018 (last year with agricultural statistics data already published) at the NUTS II level (Figure 8.2).

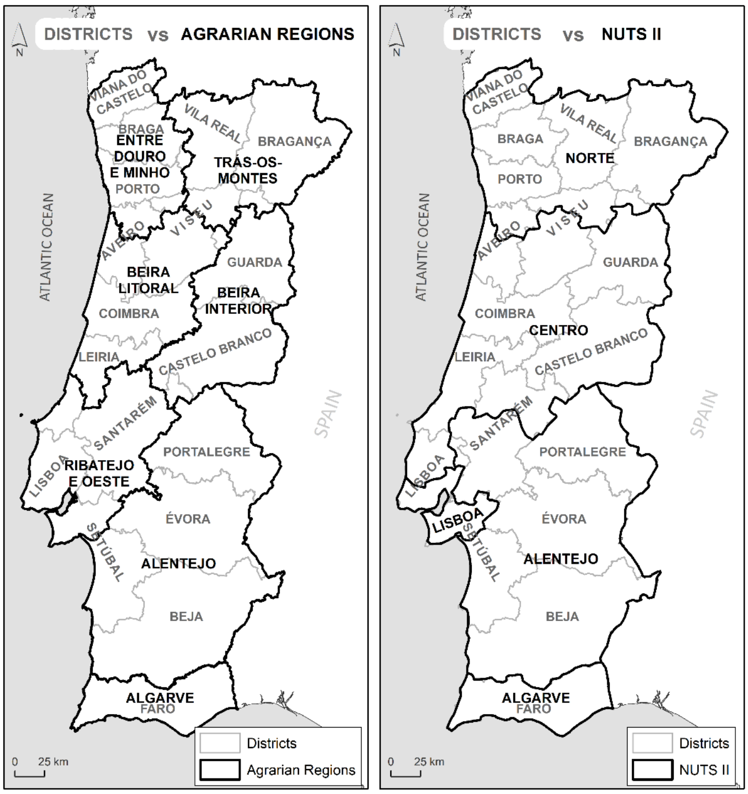


Figure 8.2. Differences between each administrative level (district vs agrarian regions and district vs NUTS II)

In summary, we used two different data sources to gather agricultural statistics from 1850 to 2018. The first source is the data exploited by the 'Agriculture in Portugal…' project at the district level from 1850 to 1915, being freely available for download on the project website (http://www.ruralportugal.ics.ul.pt/). The second source is the Statistical Yearbooks and Agricultural Statistics published by the statistics public administration services (INE) at the district level (1916 to 1989), at the agrarian level (1990 to 2005), and at the NUTS II level (2006 forwards), being freely available for download on the INE website (https://ine.pt/).

8.2.2.1 Data homogenization

To create a continuous time-series database of the annual agricultural production at the district level from 1850 to 2018 it was necessary to overcome the administrative unit variation of data from 1990 to 2018 (at the agrarian and NUTS II levels). Therefore, we proposed the use of the traditional method of areal weighting enhanced by an annual growth production coefficient to disaggregate the statistical data to the district level (Figure 8.3). The district level was used because it was the smallest administrative unit which can consistently compare the available official space–time data. In addition, for the three administrative levels, this was the level with the longest reporting.

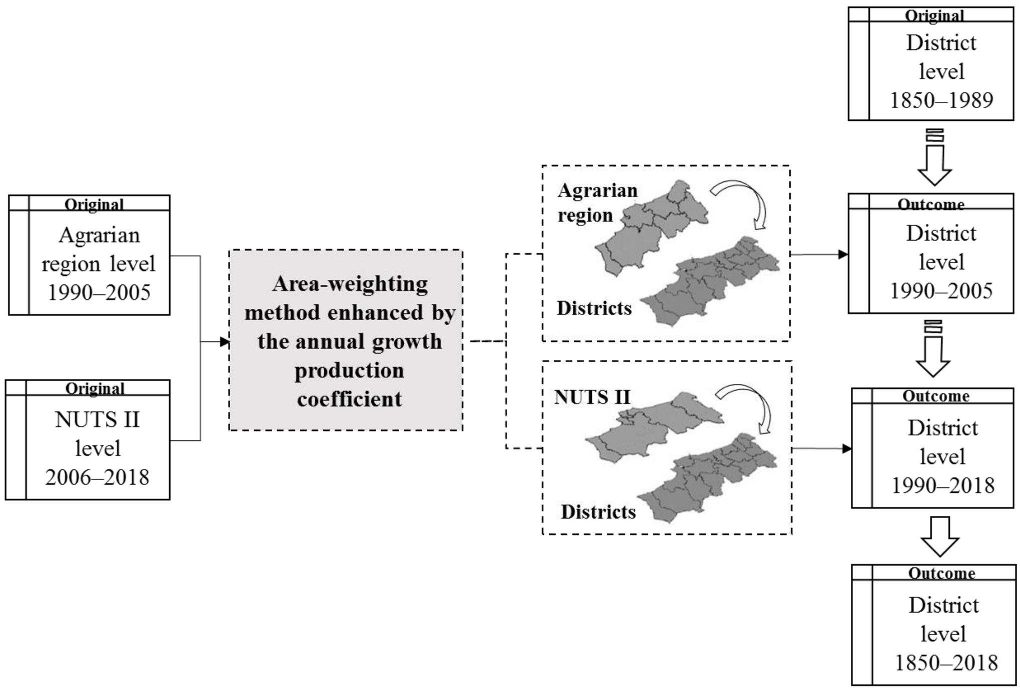


Figure 8.3. Flowchart of the statistical data homogenization methodology

The GIS-based technique of area-weighting has been applied in many studies because it enables different geographical units of analysis containing attribute data to be standardized

(Comber & Zeng, 2019; Gale, 1996; Penghui et al., 2021). Langford et al. (1991, p. 56) defines the area-weighting as a 'transfer of data from one set (source units) to a second set (target units) of overlapping, non-hierarchical, areal units'. Thus, the collected variable (source unit) is interpolated onto a target unit (the unit the analyst requires for the analysis) weighted by each source unit area within the target unit area (Goodchild & Lam, 1980; Gregory, 2002; Gregory & Ell, 2005). Accordingly, in this study, the area-weighting method is as follows:

$$\hat{y}_t = \sum_s \left(\frac{A_{st}}{A_s} \times y_s \right) \tag{8.1}$$

where $\hat{y}_t$ is the estimated variable of the target zone, y_s is the variable of the source zone, A_s is the area of the source zone, and A_{st} is the area of the zone of intersection between the source and target zones (Goodchild & Lam, 1980). This method assumes a spatially homogeneous relationship between the source and target units, which means that the variable is equally distributed across the source unit. However, this assumption is unrealistic in the real world (Fisher & Langford, 1995; Mitchel Langford, 2006). Thus, to overcome the spatially homogeneous relationship and improve the areal interpolation accuracy, we applied the annual growth production coefficient:

$$a^1 = \frac{P_t}{P_{t-1}} \tag{8.2}$$

where P_t corresponds to the first known agricultural production value at the agrarian region level for the period from 1990 to 2005 and at the NUTS II level for the period from 2006 to 2018 (N1), and P_{t-1} corresponds to the sum of the agricultural production value of each district unit area that overlaps the AR level or the NUTS II level (N1*). Thus, the evolution of agricultural production in each district is given by the annual production growth of the AR or NUTS II, where the district mainly overlaps.

We adopted this approach for three main reasons: (1) we were unable to use ancillary data due to their scarcity (Gregory, 2002; Gregory & Southall, 2005; Hallisey et al., 2017; Schroeder, 2007); (2) we could not simply assume the source unit area as the spatial conversion criterion to disaggregate the data, as there is no spatially homogeneous relationship between the source and target units, that is, both administrative units (agrarian regions and NUTS II) have large areas regarding each district level unit (Figure 2); and (3) we aimed to control significant increases in annual production values (suspicious spike values) (for more information regarding spike values see Gregory and Ell (2006)).

Therefore, the proposed method disaggregates the statistical data based on the geometric intersection of the source unit and target unit weighted by the production annual growth of the original collected data. As such, the evolution of production in each district is given by the production annual growth of the original collected data. Figure 8.4 presents an example of homogenized statistical data of the wheat production values from 1990 at the AR "Entre Douro e Minho". Thus, to disaggregate the wheat production values in 1990 at the district level, we first selected all the districts that completely or partially overlaid the AR "Entre Douro e Minho" and recorded their last know production values (at district-level, 1989). There were six districts that overlaid this region, but only three of them were spatially completely contained inside, namely "Viana do Castelo", "Braga", and "Porto". The other three districts ("Aveiro", "Vila Real", and "Viseu") areas overlapped not only the AR "Entre Douro e Minho" but also the AR "Trás os Montes" and "Beira Litoral". For these three districts, a GIS overlay operation was performed in order to obtain the areas that overlapped each agrarian region, and, accordingly, the production values of these three districts were interpolated using the areal-weighting method. Then, we calculated the annual growth production coefficient. Lastly, the production values were multiplied by the annual growth production coefficient. This procedure was applied for each AR ("Entre Douro e Minho", "Trás os Montes", "Beira Litoral", "Beira Interior", "Ribatejo e Oeste", "Alentejo", and "Algarve") from 1990 to 2005 and for the three cereals. Finally, the same procedure was applied for each NUTS II ("Norte", "Centro", "Lisboa", "Alentejo", and "Algarve") from 2006 to 2018 for the three cereals. Figure 8.5 presents an example of standardized statistical data for wheat production values from 2006 at the NUTS II "Norte" that were disaggregated at the district level.

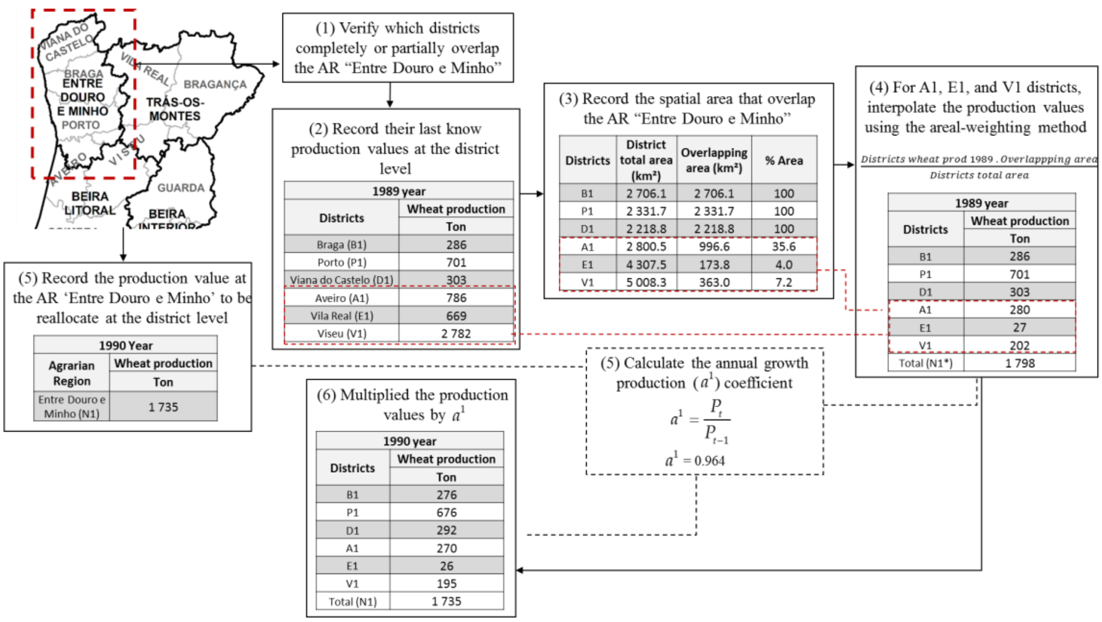


Figure 8.4. Conceptual diagram of wheat production values interpolation (1990) from the AR "Entre Douro e Minho" to the district level

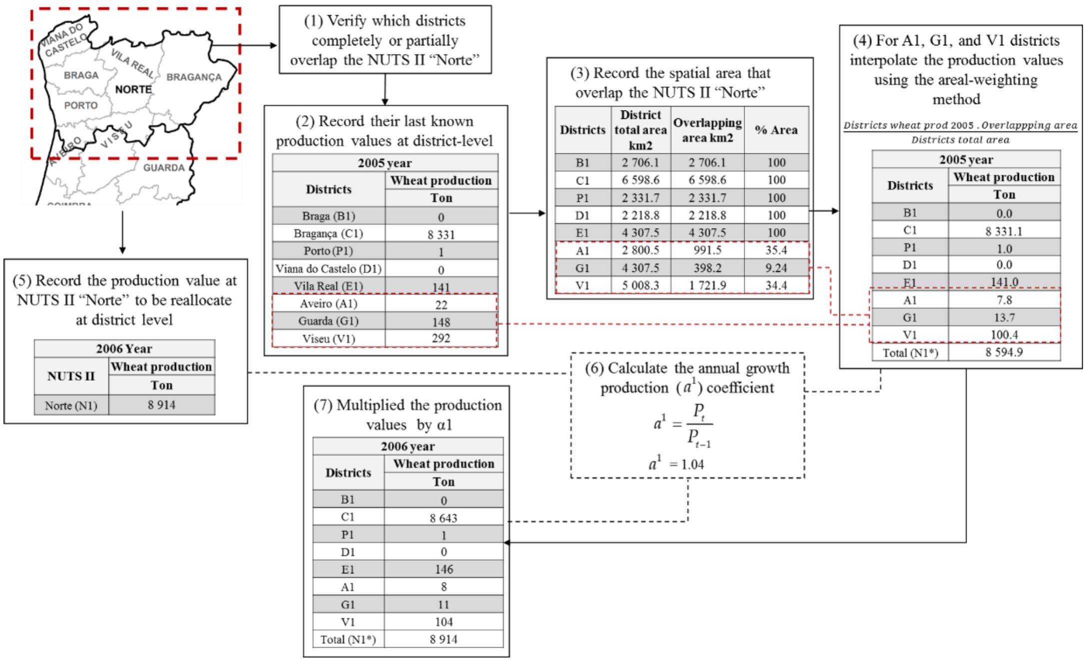


Figure 8.5. Conceptual diagram of wheat production values interpolation (2006) from NUTS II "Norte" to the district level

8.2.3. Long-term spatiotemporal interpretation and trend analysis

To provide a long-term spatiotemporal analysis of the annual agricultural production, we created annual maps by rescaling the values within a range of 0 to 100 using linear min–max normalization (Han et al., 2012). This procedure was necessary because the range of agricultural production varied widely. The normalized production value (PNV) is presented as a percentage as follows:

$$PNV = \frac{x_{i,t} - \min_t}{\max_t - \min_t} \times 100 \tag{8.3}$$

where $x_{i,t}$ is the original production value in district (i) at a given time (t); and min_t and max_t are the overall minimum and maximum production values at a given time, respectively. An index of 100 indicates the district that produced the most in each year. This procedure was applied for the production values of the wheat, maize, and rice cereals from 1850 to 2018.

Thereafter, to identify the spatiotemporal trends over 169 years, we explored the 'Emerging Hot Spot Analysis' tool from ArcGIS Pro software. This tool combines two statistical measures, namely the Mann–Kendall (MK) trend test (Yue et al., 2002) and the Getis–Ord Gi statistic (Getis & Ord, 2010) and aggregates the data into a time cube to identify the trends over time by generating space–time intervals for every location with data. The MK test searches for a trend in a time series without specifying whether it is linear or non-linear (Khaliq et al., 2009). In this study, the partition of data across time was defined in one-year intervals, from 1850 to 2018, for each Portugal district and for the absolute values. The hot and cold spot trends were evaluated using the MK trend test for each location, and one of eight categories was identified: new, consecutive, intensifying, persistent, diminishing, sporadic, oscillating, and historical hot and cold spots.

The number of time-step periods that determine which features are analyzed together to assess local space–time clustering was set to 20 years, which was considered as the most appropriate regular interval considering the evolutionary phases of regional agricultural production in Portugal (e.g. wheat campaign phase from 1929 to 1938 and world wars) (Freire & Lains, 2017). Therefore, the final analysis provides the general trend of the 20-year cluster in recent time, revealing the latest production tendency for wheat, maize, and rice.

8.3. Results

8.3.1. Overview of wheat, maize, and rice national production

From a national perspective, the production trends of wheat, maize, and rice showed three different situations: growth (maize), stability (rice), and decline (wheat) (Figure 8.6). The distinct amount of production for the three cereals is also clear. Overall, the largest production of cereals occurred in the 1950–1970 period, followed by a major production crisis in the 1970s and 1980s. Notably, for decades, maize and wheat productions has alternated. In fact, maize production was only lower than wheat in some years of the 1930s, 1950s, and 1970s. In addition, higher wheat production occurred in 1958, with approximately 805,000 tons produced. However, since the 1980s, maize production has increased compared to that of the other two cereals, registering an accelerated growth and reaching its absolute maximum production value in 1998 (>1000000 tons). Simultaneously, wheat started to present smaller production values, and only in the 1990s did wheat production increased. Since the 1940s, rice production has presented a growing trend, in which higher production occurred in 1970, but only to approximately 227,000 tons.

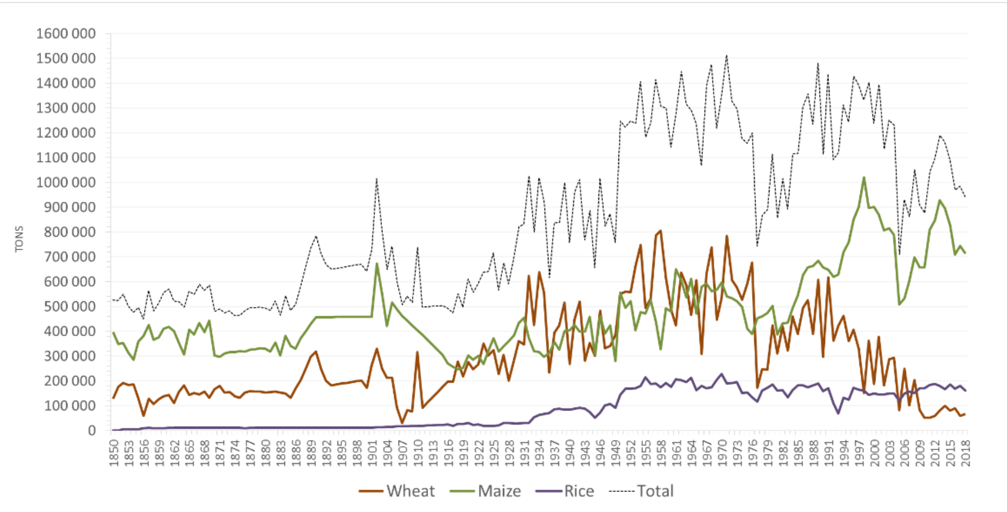


Figure 8.6. Wheat, maize, and rice national production during 1850–2018

8.3.2. Regional long-term spatiotemporal analysis

Chronologically, wheat was mainly cultivated in the southern region of the country, more specifically in the Alentejo region (Figure 8.7). Over the 169-year period under study, the importance of the districts that are part of the Alentejo region (Portalegre, Évora, Beja) to wheat production is remarkable. Since 1984, the spatial pattern remains, and Beja and Évora are the districts that produce the most in Portugal, followed by Portalegre. In particular, Beja consistently produced more wheat between 1850 and 2018, except for 1977 and 1981, in which Évora produced more wheat than Beja.

Figure 7 shows the existence of two other regions situated in the south-central part of Portugal, with a strong impact on wheat production. The first is the Lisbon District, which has a relevant role in wheat production. During poor wheat production in Alentejo region (1906 and 1907), the production in Lisbon was significant. However, in the following years, the Lisbon District started to lose importance in wheat production and became irrelevant from a regional perspective. Historically, the Lisbon district heavily lost wheat production due to urban development that took the good land used to produce this cereal.

Another prominent region is the Ribatejo and Lezíria do Tejo, which mainly incorporates the municipalities of the Santarém District. Specifically, wheat production presented several oscillations over the 169-year period under study, with a strongly decreasing trend after 1936. It is also particularly relevant that the northwest of the country always had a low tendency for wheat production, which will be better explained in the analysis of maize production in mainland Portugal. Lastly, we emphasize the oscillating relevance over the 169 years of the Bragança District, which is situated in the northeast region.

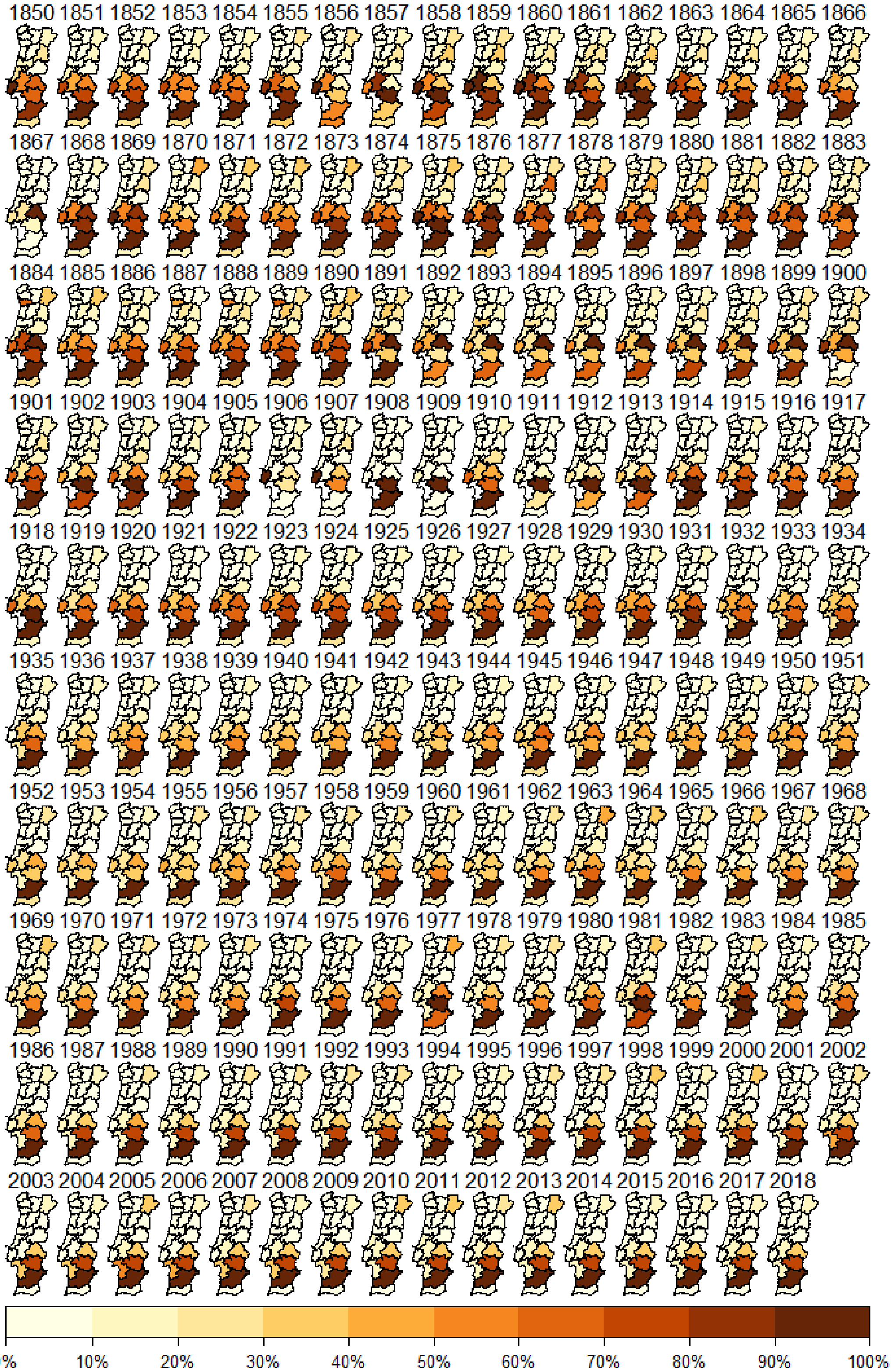


Figure 8.7. Spatiotemporal pattern of wheat production (1850–2018)

Figure 8.8 shows that regions which produce wheat do not produce maize. In the 19th century, maize production was concentrated in the opposite direction of wheat production, particularly in the northern and coastal regions. This cereal expanded from northwest to south along the coastal regions of small agricultural parcels. By the beginning of the 20th century, maize production reached its maximum expansion in the Portuguese territory.

Figure 8.8. Spatiotemporal pattern of maize production (1850–2018)

However, during the late 1980s, with the entry of Portugal into the European Economic Community (in 1986), maize production moved to the south-central part of the country. Presently, the main maize production region is the south-central region of the country, more specifically the Ribatejo zone (Santarém District), which has emphasized maize production since 1991. In fact, for the last 25 years, Santarém has consistently produced more maize in the national context. Cereal production has also increased slowly in the Alentejo region. Notably, the regions historically associated with maize production are no longer focused on this crop.

In the second half of the 19th century, the main rice-producing regions were the districts of Beja, Évora, and Portalegre (Figure 8.9). However, the rice production distributions rapidly changed, turning the emphasis to different regions, such as the districts of Lisbon, Santarém, and Aveiro. The main production regions of this cereal shifted to the coastal area north of Lisbon District up to the area of Aveiro District (Vagos), passing through the Coimbra and Mondego valleys. This distribution continued until the beginning of the 20th century. In addition, in the 1970s, the rice reached its maximum expansion in the Portuguese territory, reaching Alentejo. Recent decades highlight the importance of Setúbal and Santarém regions.

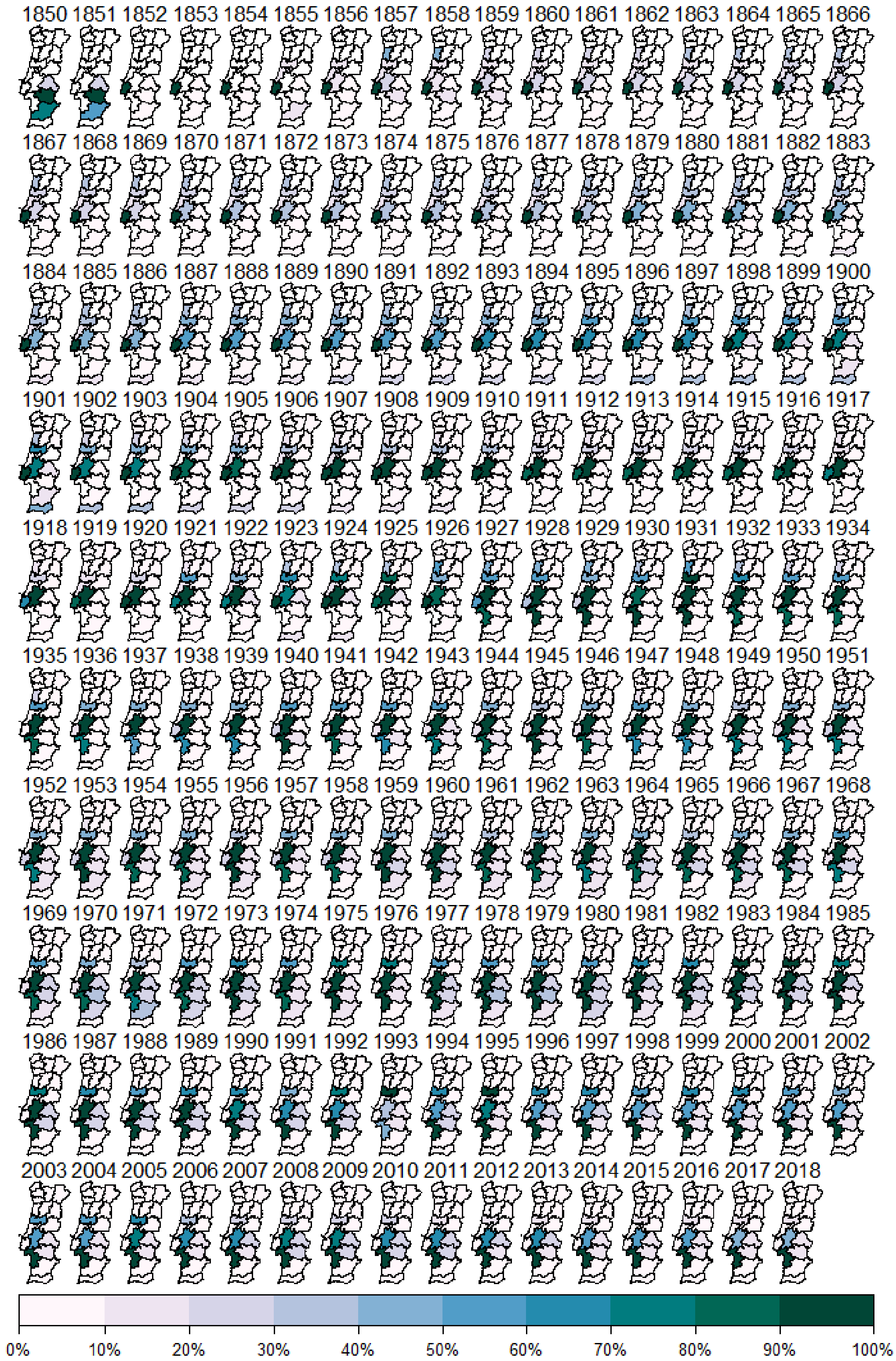


Figure 8.9. Spatiotemporal pattern of rice production (1850–2018)

8.3.3. Spatiotemporal pattern and trends analysis

The three cereals exhibited distinct production evolution trends (Figure 8.10). In the case of wheat production patterns, two main regions can be identified: the north center and the south of

Portugal. The first one is a cold area that includes the central (Lisboa, Santarém, Leiria, Coimbra, and Castelo Branco) and northern regions (Viana do Castelo, Braga, Porto, Aveiro, Viseu, Guarda, Vila Real, and Bragança). Most of these district' production was statistically significant cold spots for 90% of the time-step periods, including the final time step. The sporadic cold spot districts indicated an on-again then off-again production trend, as < 90% of the time-step periods consisted of statistically significant cold spots, and none was considered a hot spot of statistical significance. Coimbra is of particular interest because it is the only district that presented a persistent production trend among cold spots, which indicates that this region has no discernible production trend, showing a rise or fall in the magnitude of the clustering of counts over time. The oscillating pattern of Lisboa and Santarém shows that the production trend was significantly cold at the last time-step period, but was also a statistically significant hot spot during previous time steps. The second main region was the southern region (Setúbal, Évora, Beja, and Faro). These district productions exhibited continuous hot spot bins with statistical values in the last time-step periods. Finally, no pattern was observed for the Portalegre District.

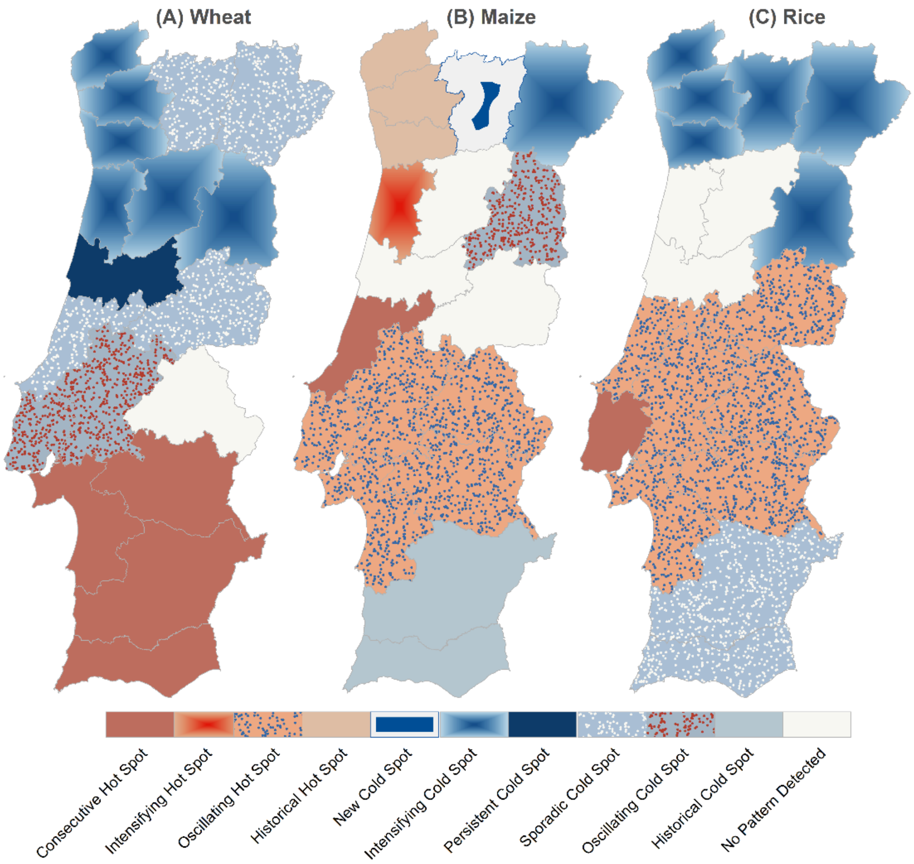


Figure 8.10. Wheat (A), maize (B), and rice (C) production trend patterns

The results of the maize production trend presented distinct patterns in the north of Portugal. On the northern coastline, the production of three districts (Viana do Castelo, Braga, and Porto) in the most recent time period was not hot; however, in at least 90% of the time-step periods, there have been important hot spots. Guarda was considerably cold during the last time-step period but was considered a hot spot of statistical significance in previous time steps. In contrast, Vila Real and Bragança presented cold spot patterns. In the first case, a new cold spot was statistically significant in the most recent period; however, no statistically significant cold spot occurred before. In the second case, this district was an important cold spot for 90% of the time-step periods, considering also the final one.

In contrast, the production trend of Aveiro increased overall, and this increase was statistically significant for 90% of the time-step periods, considering also the last one. We observed two different production hotspots in the central and southern parts of Portugal. One is the Leiria District, which exhibited continuous hot spot bins with statistical significance in the last time-step periods. The second is marked by an oscillating hot spot, which indicates that although these districts (Lisboa, Santarém, Portalegre, Setúbal, and Évora) were not hot production spots in the most recent period, they were important hot spots for at least 90% of the time-step periods. In contrast, for half of the center region (Viseu, Coimbra, and Castelo Branco), no pattern was observed. Finally, in the most recent period, the southern region of Portugal (Faro and Beja) exhibited no cold spots; however, at least 90% of the time-step periods were statistically significant cold production spots.

The trend pattern for rice production allows us to divide Portugal into four main regions: two hot spots and two cold spots. The first is the north (Viana do Castelo, Braga, Porto, Vila Real, Bragança, and Guarda), which presents a consistent production pattern as a statistically significant cold spot for 90% of the time-step periods, considering also the final one. The second is half of the central and southern regions (Castelo Branco, Santarém, Leiria, Setúbal, Évora, and Portalegre). Similar to the maize pattern of production regions, these districts in the most recent period were not hot; however, they were statistically significant hot spots in at least 90% of the time-step periods. The Lisbon District is distinguished for having continuous hot spot bins with statistical significance in the last time-step periods. Finally, Beja and Faro presented a sporadic pattern, which means that these districts were on-again than off-again cold production spots, with < 90% of the time-step periods being statistically significant cold spots, with no statistically significant hot spots in previous time-step periods.

8.4. Discussion

8.4.1. Visualisation, description, and interpretation of the regional productions

Historically, the Portuguese agricultural sector has different regional specializations because of the considerable diversity of natural and economic–social conditions, which produced different reactions to the impacts of public policies, markets, and technological changes over time (Avillez, 2015; Freire & Lains, 2017). In this study, notably, for decades, maize and wheat production alternated, depending on agricultural years and political aspects such as the Wheat Campaign (an initiative launched during 1929–1938 in order to achieve productive self-sufficiency) (Freire & Lains, 2017). In fact, we were able to confirm that the regional production was influenced by two types of public policies, with different incentives that impacted the rural areas and agricultural activities and affected the cereals. One of the policy orientations was to stimulate the production of specific dryland (non-irrigated) crops. In the group of cereals, wheat benefited most.

Specifically, in Portugal, wheat production, typically in dryland, was introduced by the Romans (in the years before Christ). Therefore, wheat is one of the most important cereals in Portugal, and historically, there was a goal to extend its production to all regions of the country. However, from an agroecological or economic perspective, this was not viable in some regions (Pais et al., 1976). In particular, wheat production only increased in regions in which there was an intersection of public policies (wheat campaign with the 1889 legislation, and the wheat policy that started in 1929 and ended in the 1960s) with agroecological and biophysical conditions, such as the Alentejo, Ribatejo, Lezíria do Tejo (most fertile lands in the country), and Lisbon regions (Rosas, 1991).

Chronologically, wheat was mainly cultivated in the southern region of the country, more specifically in the Alentejo region, due to not only ecological and biophysical conditions, but also the parcel structures. It is a cereal that grows well in large parcels, exactly the type of field geometry of this region. However, Beja was the district that consistently produced more wheat, mainly explained by the type of soil existing in this district, 'Barros de Beja' (Beja clays), which is clayey soil rich in humus and very fertile, and so particularly suitable for wheat production. In the minifundium districts in the northern part of the country, wheat production has only been marginally altered, and in contrast, Évora, Beja, Portalegre, and Santarém districts of great exploitation (latifundium) saw their production substantially increased (Jones et al., 2011).

Historically, the Lisbon district heavily lost wheat production due to the urban development of the city that took the land used to produce this cereal. In addition, wheat production in the Ribatejo and Lezíria do Tejo regions had several oscillations over the 169-year period, with a strong decreasing trend after 1936, mainly because in the late 19th century these agricultural fields began to be exploited for other productions such as wine (Rosas, 1991). In this period, wine competed greatly for the same resources as wheat. The reason why the spreading of vineyards and other crops did not occur in the great region of Alentejo was only because of the public policies that gave special attention to the protection of wheat in this region until the 20th century (Jones et al., 2011). In summary, when the campaign finished, due to the poor monetary credit conditions and the manner in which the land was being exploited (e.g. left alone, tenancy agreement, trading partnership), large wheat production imbalances occurred, and only in the 1990s did they increase due to the exceptional weather conditions.

The other set of public policies that supported crops enabled irrigation production through the development of infrastructure. However, at first, this set of public policies positively impacted the production of rice, as the first interventions in irrigated land were initially implemented in regions where rice was already produced. Therefore, in the 1970s, rice reached its maximum expansion in the Portuguese territory, reaching the Alentejo region, due to the public work interventions in this region. Recent decades have highlighted the importance of the Setúbal and Santarém regions, which considerably benefited from public investment in irrigation infrastructures. Overall, mainly because of the manner in which it was collected by the public agencies, the evolution of rice production has not been thoroughly documented from the perspective of historical knowledge, which opens a set of questions that forces us to seek a more in-depth analysis of this cereal.

In the second stage, since the 1960s, public policies have benefited dryland regions that were liable to be converted to irrigated land through water dam construction (Rosas, 1991). Therefore, the abrupt change in maize production in the Ribatejo zone (Santarém District) started when the protection measures regarding wheat stopped, and the focus turned to the irrigation of crop production, which benefited from the developed irrigation infrastructures in the previous years. Since the second half of the 1970s, maize production predominantly increased due to the construction of irrigation systems by the state. It is interesting to see that this cereal, which was only introduced in the 16th century after the first Cristóvão Colombo trip (in 1492), has always had more considerable importance in the agriculture context than wheat.

Historically, maize has been a recent cereal in the Iberian Peninsula, and it progressed from 1905 due to the biophysical conditions that the Portuguese territory offered to produce this cereal. This cereal expanded from northwest to south along the coastal regions of small agricultural parcels, particularly since at least the second half of the 19th century until the 1930s (Pereira, 1983). In fact, maize production was lower than wheat only in some years of the 1930s, 1950s, and 1970s, when there was a combination of strong political wheat protectionism with ‘good’ agricultural years.

Chronologically, the distributions of wheat and maize in Portugal since at least the mid-19th century were evidently regional. Maize was confined to the northern regions of Portugal, which are more humid and have higher availability of water, whereas wheat was grown mainly in Alentejo (Pais et al., 1976). Notably, the regions historically associated with maize production ceased to be so, mainly due to irrigation policy incentives. In fact, our results provide some clues to solve questions that have not yet been clarified, stimulating geographical and historical cross-knowledge and providing a national overview and a regional prospect that was lacking.

Finally, the space–time pattern trend analysis enabled the interpretation of complex trends that occurred over 169 years. The changes over time presented a pattern that, in the case of these three cereals, enabled the clear division of the country into its main regions according to the cereal production. The identification of hot and cold spots helped to determine which regions are facing increasing or decreasing threats to agricultural production. Maize and rice, both irrigated croplands, presented a similar pattern in some regions of Portugal, mainly the central region. The differences we observed between both crops were mainly related to the agroecological conditions. The wheat production trend presented a completely different pattern when compared to the other two cereals, mainly because this cereal is a dryland crop. Thus, considering the size of the statistical series, we suggest it would be important to select different parameters in the future that can also provide good results and explain different trends present in the data (Harris et al., 2017).

8.4.2. Data homogenisation

Currently, in terms of quantitative information on agriculture production for the period prior to 1916, there was a major advance when the interdisciplinary research team of the "Agriculture in Portugal…" project collected many of the scattered data stored in the central and regional Portuguese public archives. These data together with the data published by INE, gave 140 years (1850-1989) of agriculture production information at the district level. While at this stage we already had absolute production values from 1850 to 2018 at the national level, we continued to struggle to obtain a clear understanding of the individual dynamics and oscillations of regional agriculture production over the last 30 years. Accordingly, in this study, we used a GIS-based spatial analytical approach to homogenize the Portuguese agricultural statistical data, with the purpose of creating a long spatial time-series with annual and regional agricultural production during 1850-2018.

Although the use of the area-weighting GIS-based technique to homogenize the different system data units has long been recognized in the literature (Gregory, 2002; Gregory & Ell, 2005; Huyen Do et al., 2015; Kounadi et al., 2018; Murakami & Tsutsumi, 2011), errors may exist in our proposed methodology. Thus, a major limitation of this study is that we could not attest the veracity of the interpolated production values. Nonetheless, quantifying the error using real-world data is typically challenging, particularly with regard to historical data (Gregory & Ell, 2006; Hallisey et al., 2017). In the future, an alternative solution should be developed to evaluate the interpolation error of the proposed methodology in order to attest the data accuracy (e.g. the approaches of Geddes et al. (2013) or Thevenin, et al. (2016)). Ultimately, the developed methodology avoids time-consuming and/or computationally intensive complex GIS-based methods, and it can be tested for agricultural statistics data from other countries or other statistical data with administrative unit variations, such as demographic census data (Gregory, 2002).

8.4.3. Data inconsistencies

The long spatial time-series presented two inconsistencies. One is regarding rice production data, as there were some value gaps in the first years for the period under study, specifically from 1850 to 1900 for 7 of the 18 districts (Braga, Bragança, Castelo Branco, Guarda, Porto, Vila Real, and Viseu). Nevertheless, we decided to use the rice production data because, although these districts only had available data starting around the 1900s, the production values of the following years were zero. Hence, we assumed that the gap in data would not affect the findings.

The second limitation of our statistical data are the production values of Lisbon and Setúbal districts. Until 1927, Setúbal District was an administrative part of the Lisbon District, that is, both districts were a single administrative unit, which means the values of Lisbon from 1850 to 1926 are a sum of the Lisbon and Setúbal productions. This event is, however, explained by the way data were collected until 1927. Thus, the description and interpretation of the production values of Lisbon and Setúbal districts until 1926 should be made with precaution, specifically for rice production, as the Setúbal District emerged as the main producer of this cereal, particularly after 1927, whereas Lisbon ceased to be an important producer, a conclusion that was hidden in the maps until 1927. It is particularly important for future studies to disaggregate the production values of Lisbon from 1850 to 1926 into two administrative units to allow a more thorough analysis.

8.5. Conclusions

This study advances understanding and analysis of past and current regional agricultural production patterns and trends in Portugal of three cereal crops (wheat, maize and rice). Although we acknowledge that there may be some precautions taken regarding the use of these data due to the limitations associated with the absence of the possible error identification caused by our proposed methodology and how some partial data were collected, this was a qualitative advance for Portuguese modern economical historiography. The analysis of Portuguese agriculture now has a new contemporary quantitative dimension, allowing comprehensive space–time cross-reading of the general characteristics of the regional distribution of agricultural production for a 169-year period. Additionally, we can now consolidate important historic information on Portuguese agriculture with how it has developed to date, articulating spatial and temporal aspects of certain events and historical conjunctures in the Portuguese agriculture field.

In this study, a preliminary analysis was presented and related to successive public policies; however, notably, there are more lessons to be learned from this long spatial time-series. Furthermore, we have the opportunity to expand the discussion relating to European and global trends by including the Portuguese case in academic discussions and further contribute to the integration of long spatial time-series in scientific analysis that can strengthen the ability to assess past and current regional food production situations, which are essential for a country's food security evaluation. One major benefit of having such data is the possibility of conducting an in-depth analysis of the past to truly understand the present (Gregory & Ell, 2005) and forecast probable future events (Murrieta-Flores & Martins, 2019). A direction for future work would be to combine these statistical data with long-term meteorological data (e.g., ECWMF or GLDAS), as climate change is known to directly affect the variation of crop cultivation. Furthermore, the proposed methodological approach is a direct solution that further contributes to the integration of long spatial time-series in scientific analysis to evaluate past and current changes.

References

Avillez, F. (2015). A Agricultura Portuguesa. Fundação Francisco Manuel dos Santos.

Boivin, N., & Crowther, A. (2021). Mobilizing the past to shape a better Anthropocene. Nature Ecology & Evolution 2021, 1–12. https://doi.org/10.1038/s41559-020-01361-4

Calleja, E. J., Ilbery, B., & Mills, P. R. (2012). Agricultural change and the rise of the British strawberry industry, 1920-2009. Journal of Rural Studies, 28(4), 603–611. https://doi.org/10.1016/j.jrurstud.2012.07.005

Comber, A., & Zeng, W. (2019). Spatial interpolation using areal features: A review of methods and opportunities using new forms of data with coded illustrations. Geography Compass, 13(10), 1–23. https://doi.org/10.1111/gec3.12465

Eurostat. (2020). Regional yearbook: 2020 Edition. Publications Office of the European Union. https://doi.org/10.2785/764915

Fisher, P. F., & Langford, M. (1995). Modelling the Errors in Areal Interpolation between Zonal Systems by Monte Carlo Simulation. Environment and Planning A: Economy and Space, 27(2), 211–224. https://doi.org/10.1068/a270211

Freire, D., & Lains, P. (2017). An agrarian history of Portugal, 1000-2000: economic development on the European frontier (D. Freire & P. Lains (eds.)). Brill.

Gale, H. F. (1996). Age cohort analysis of the 20th century decline in U.S. farm numbers. Journal of Rural Studies, 12(1), 15–25. https://doi.org/10.1016/0743-0167(95)00049-6

Geddes, A., Elston, D. A., Hodgson, M. E. A., & Birnie, R. V. (2013). Stochastic model-based methods for handling uncertainty in areal interpolation. International Journal of Geographical Information Science, 27(4), 785–803. https://doi.org/10.1080/13658816.2012.722636

Getis, A., & Ord, J. K. (2010). The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis, 24(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Goodchild, M. F., & Lam, N. (1980). Areal Interpolation: A Variant of the Traditional Spatial Problem. Geo-Processing, I, 297–312.

Gregory, I. (2002). The accuracy of areal interpolation techniques: standardising 19th and 20th century census data to allow long-term comparisons. Computers, Environment and Urban Systems, 26(4), 293–314. https://doi.org/10.1016/S0198-9715(01)00013-8

Gregory, I., Dorling, D., & Southall, H. (2001). A century of inequality in England and Wales using standardized geographical units. Area, 33(3), 297–311. https://doi.org/10.1111/1475-4762.00033

Gregory, I., & Ell, P. (2005). Breaking the boundaries: geographical approaches to integrating 200 years of the census. Journal of the Royal Statistical Society: Series A (Statistics in Society), 168(2), 419–437. https://doi.org/10.1111/j.1467-985X.2005.00356.x

Gregory, I., & Ell, P. (2006). Error-sensitive historical GIS: Identifying areal interpolation errors in time-series data. International Journal of Geographical Information Science, 20(2), 135–152. https://doi.org/10.1080/13658810500399589

Gregory, I., & Southall, H. (2005). The Great Britain Historical GIS. Historical Geography 33, 132–134.

Hallisey, E., Tai, E., Berens, A., Wilt, G., Peipins, L., Lewis, B., Graham, S., Flanagan, B., & Lunsford, N. B. (2017). Transforming geographic scale: a comparison of combined population and areal weighting to other interpolation methods. International Journal of Health Geographics, 16, 29. https://doi.org/10.1186/s12942-017-0102-z

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In Data Mining: Concepts and Techniques. Elsevier Inc. https://doi.org/10.1016/C2009-0-61819-5

Harris, N. L., Goldman, E., Gabris, C., Nordling, J., Minnemeyer, S., Ansari, S., Lippmann, M., Bennett, L., Raad, M., Hansen, M., & Potapov, P. (2017). Using spatial statistics to identify emerging hot spots of forest loss. Environmental Research Letters, 12(2), 024012. https://doi.org/10.1088/1748-9326/aa5a2f

Huyen Do, V., Thomas-Agnan, C., & Vanhems, A. (2015). Spatial reallocation of areal data – another look at basic methods. Revue d'Économie Régionale & Urbaine, 1, 27–58. https://doi.org/10.3917/reru.151.0027

Jones, N., de Graaff, J., Rodrigo, I., & Duarte, F. (2011). Historical review of land use changes in Portugal (before and after EU integration in 1986) and their implications for land degradation and conservation, with a focus on Centro and Alentejo regions. Applied Geography, 31(3), 1036–1048. https://doi.org/10.1016/j.apgeog.2011.01.024

Justino, D. (1988). A formação do espaço económico nacional. Portugal: 1810-1913, 2 vols. Vega.

Khaliq, M. N., Ouarda, T. B. M. J., Gachon, P., Sushama, L., & St-Hilaire, A. (2009). Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers. Journal of Hydrology, 368(1–4), 117–130. https://doi.org/10.1016/j.jhydrol.2009.01.035

Kounadi, O., Ristea, A., Leitner, M., & Langford, C. (2018). Population at risk: using areal interpolation and Twitter messages to create population models for burglaries and robberies. Cartography and Geographic Information Science, 45(3), 205–220. https://doi.org/10.1080/15230406.2017.1304243

Lains, P. (2003). Os progressos do atraso : uma nova história económica de Portugal, 1842-1992. Instituto Ciencias Sociais da Universidade de Lisboa.

Lains, P., & Silveira E Sousa, P. (1998). Estatística e produção agrícola em Portugal, 1848-1913. Análise Social, XXXIII(149), 935–968.

Langford, M., Maguire, D., & Unwin, D. J. (1991). The areal interpolation problem: estimating population using remote sensing in a GIS framework. In I. Masser & B. Blakemore (Eds.), Handling geographical information: methodology and potential applications (pp. 55–77). Longman Scientific & Technical.

Langford, Mitchel. (2006). Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. Computers, Environment and Urban Systems, 30(2), 161–180. https://doi.org/10.1016/j.compenvurbsys.2004.07.001

Mitchell, B. R. (1998). International Historical Statistics. In International Historical Statistics. Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-14735-9

Mitchell, B. R. (2011). British historical statistics.

Murakami, D., & Tsutsumi, M. (2011). A New Areal Interpolation Method Based on Spatial Statistics. Procedia - Social and Behavioral Sciences, 21, 230–239. https://doi.org/10.1016/J.SBSPRO.2011.07.034

Murrieta-Flores, P., & Martins, B. (2019). The geospatial humanities: past, present and future. In International Journal of Geographical Information Science (Vol. 33, Issue 12, pp. 2424–2429). Taylor and Francis Ltd. https://doi.org/10.1080/13658816.2019.1645336

Pais, J. M., Lima, A. M. V. De, Baptista, J. F., Jesus, M. F. M. de, & Gameiro, M. M. (1976). Elementos para a história do fascismo nos campos: A «Campanha do Trigo»: 1928-38 (I). Análise Social, XII(43), 401–473.

Penghui, J., Dengshuai, C., & Manchun, L. (2021). Farmland landscape fragmentation evolution and its driving mechanism from rural to urban: A case study of Changzhou City. Journal of Rural Studies, 82, 1–18. https://doi.org/10.1016/j.jrurstud.2021.01.004

Pereira, M. H. (1983). Livre-câmbio e desenvolvimento económico. Sá da Costa.

Ribeiro, O., Lautensach, H., & Daveau, S. (1988). Geografia de Portugal. Sá da Costa.

Rosas, F. (1991). Rafael Duque e a política agrária do Estado Novo (1934-44). Análise Social, xxvi(3), 771–790.

Schroeder, J. P. (2007). Target-Density Weighting Interpolation and Uncertainty Evaluation for Temporal Analysis of Census Data. Geographical Analysis, 39(3), 311–335. https://doi.org/10.1111/j.1538-4632.2007.00706.x

Thevenin, T., Mimeur, C., Schwartz, R., & Sapet, L. (2016). Measuring one century of railway accessibility and population change in France. A historical GIS approach. Journal of Transport Geography, 56, 62–76. https://doi.org/10.1016/j.jtrangeo.2016.08.017

Thomas, R. W. (1980). Modelling in geography : a mathematical approach. Harper & Row.

van Zanden, J. L., & van Leeuwen, B. (2012). Persistent but not consistent: The growth of national income in Holland 1347-1807. Explorations in Economic History, 49(2), 119–130. https://doi.org/10.1016/j.eeh.2011.11.002

Westlund, H., & Nilsson, P. (2019). Agriculture's transformation and land-use change in a post-urban world: A case study of the Stockholm region. Journal of Rural Studies. https://doi.org/10.1016/j.jrurstud.2019.07.002

Yue, S., Pilon, P., & Cavadias, G. (2002). Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. Journal of Hydrology, 259(1–4), 254–271. https://doi.org/10.1016/S0022-1694(01)00594-7

Part VI. General Conclusion

Chapter 9. General Conclusion

This thesis arises from understanding how the integration of concepts, tools, techniques, and methods from GIS can provide a formalised knowledge base for agricultural land systems at multiple spatial and temporal scales and geographical contexts in response to future agricultural and food system challenges. Four research questions associated with the overall objective were addressed in Parts II–V of the thesis in seven chapters. To achieve the overall objective, each chapter presents the theory and practical innovation by applying a variety of GIS-based approaches and spatial data sources, with different cross-disciplinary solutions, cross-scale frameworks, and comprehensive assessments. This final chapter provides a general reflection of the main findings of the research conducted in Chapters 2–8 in the context of the stated research questions, highlights several practical implications, weaknesses, and opportunities, with recommendations for further research.

Part II, Chapter 2 responds to the need to situate the context of the contribution of this thesis within the scientific domain of agricultural land systems. As a result, this chapter revises and systematises the heterogeneity of the published research studies conducted from an agricultural land systems perspective to summarise the body of knowledge and state the answer to RQ. 1. The chapter expands the current knowledge on the research fields, methodological directions, and data sources, making a substantial contribution to current agricultural land systems research by identifying gaps and highlighting insights that can assist both scholars, practitioners, and early career researchers to acquaint themselves with research trends. Furthermore, this chapter has shown the multi-scalar complexities of agricultural land systems and their link to SDGs that should be addressed holistically with an interdisciplinary approach incorporating knowledge of life, social, earth, environmental, and sustainability sciences and many other disciplines to support a variety of methodological approaches.

This finding reflects the multi-dimensionality of agricultural land system research, the importance of different science standpoints, improvements in data collection, and robust models that are essential to provide new knowledge and likely present new questions that must be addressed in different cross-scale frameworks and geographical contexts. Of particular relevance is the finding that six main research fields summarise diverse and highly important topics for long-term global development through the integration of research and innovation activities with systemic theories, multitemporal and multisensory technologies, GIS approaches, scenario development and analysis, land-use models, and spatial data. Notably, the research conducted in Chapters 3–8 that forms part of this thesis has significantly contributed to the current state of knowledge on the research field of *the dynamics of agricultural land systems*.

In Part III, Chapters 3 and 4 explore remotely sensed and VGI data, multitemporal and multisensory approaches, as well as a variety of statistical methods for mapping, quantifying, and assessing regional agricultural land dynamics in the Beja district. By providing practical and theoretical insights regarding the availability of potential spatial data sources, and the advantages and limitations in their use, the two chapters substantially contribute to integrating knowledge and providing answers to RQ. 2. In Chapter 3, we explore a methodological approach using the K-means clustering technique to refine sample sources for training the TWDTW classifier algorithm to support long-term classification and multi-temporal land-use change detection from Landsat imagery time series. In Chapter 4, we evaluate the potential of OSM data as an auxiliary data source to support rural multi-temporal land-use mapping at the regional scale, as well as for training purposes in supervised multi-temporal remote sensing classifications. In Part IV, Chapters 5, 6, and 7 explore the CA-Markov model, discrete-time Markov chain model, machine learning, and model-agnostic approach, as well as a set of spatial metrics and statistical methods for modelling the factors and spatiotemporal changes of agricultural land-use in the Beja district. The three studies contribute to the practical and theoretical knowledge of well-established and recently developed GIS-based approaches by highlighting their capacities and limitations, providing answers to RQ. 3. In Chapter 5, a framework is developed to better understand how to evaluate a simulation model with an illustration of CA-Markov and insightful validation methods. In Chapter 6, a set of spatial metrics and statistical methods, and the Markov model, are explored to map, quantify, assess, and predict spatiotemporal changes in agricultural land use. In Chapter 7, an analytical framework combining machine learning and model-agnostic methods is integrated to increase the understanding of the factors potentially explaining the use of agricultural land for three crop plantations. In Part V, Chapter 8 explores an area-weighting GIS-based technique, a spatiotemporal data cube, and statistical methods to model the spatial distribution over time of regional agricultural production in Portugal. The study contributes to practical and theoretical knowledge of potential GIS-based approaches and statistical data sources by highlighting their capacities and limitations, providing answers to RQ. 4. In Chapter 8, a set of GIS-based approaches are explored to produce a long spatial time-series of agricultural production data at the regional level to map, quantify, and assess the temporal evolution and spatial distribution of cereal production.

On a more general note, along with the research conducted in this thesis, there was a consistent finding that in terms of available data, there is a lack of comprehensive and representative data for mapping and analysing agricultural land dynamics in Portugal (RQ. 2). Even in the age of ‘big data', satellites, and VGI platforms, we confronted several data weaknesses related to lack of spatial, thematic, or temporal detail. For example, remotely sensed data, such as Landsat products, provide high spatial resolution and temporal coverage; however, the case study in Chapter 3 showed that there are limitations to its use. Most of these limitations are related to long-term application in rural regions. First, even with available refined training samples that improved the overall accuracy of image classification, we were not able to distinguish and accurately map the type of crop in detail, for example, olive groves or wheat cereal parcels, partly because of the complex spectral properties of the different vegetation types present in the study area. Second, despite the regular coverage of almost 40 years of Landsat images, the temporal coverage of the analysis was limited to a 21-year period (1995-2018) because it was not possible to gather samples for previous years to train the classifier algorithm. Third, the study was conducted at the municipality level, because at the district level, the chosen classifier was not time-efficient, having increased the processing time to several weeks based on our available hardware. Furthermore, the case study in Chapter 4 showed that despite the wide use of OSM data for mapping and visualisation, there are limitations when using this data source to support regional and rural multi-temporal agricultural land mapping. For example, the lack of available OSM data for the years prior to 2011 restricted the temporal coverage of the analysis. In addition, the lack of thematic detail in OSM data does not allow us to distinguish the type of crop in detail. Thus, using either satellite or VGI platforms, each data source lacked one or more critical components (space, time, or theme). These findings, which are strongly conditioned by our study area characteristics, gradually dictated the type of data sources used in subsequent studies.

From a scientific and methodological perspective, the thesis case studies demonstrated the strengths and limitations of several methods, tools, and techniques from GIS, reflecting in part the state-of-art of land-use modelling and contributing significantly to the introduction of advances in agricultural system modelling research and land system science. Nevertheless, the GIS-based approaches explored in Chapters 5–8 were not meant to be exhaustive in terms of the large variety of approaches currently available. Instead, we aimed to demonstrate the underlying principles that underpin each empirical GIS-based approach in a way that allows us to infer their potentiality and appropriateness for modelling regional agricultural land-use and production dynamics, stimulating further research along this line. Thus, in terms of practical implications, Chapters 5–8 demonstrate that application of GIS-based approaches necessarily varies according to objectives and intended use, limitations due to model design, data availability, and timely, accurate spatial data (RQ. 3 and RQ. 4). For instance, with a practical example of a well-established land-change model, Chapter 5 illustrates a research limitation owing to the model design. As demonstrated in this chapter, the IDRISI CA-Markov model was not suitable for modelling agricultural land changes and extrapolating future states in the Beja district, mainly because we were not able to understand and control all aspects of the model behaviour. Considering the importance of this finding, the follow-up of this chapter was redesigned to help users more clearly identify the characteristics and limitations of a simulation model before they are immersed in the details of the analysis, to prevent misinterpretation of the outputs. Thus, the framework presented in this chapter forms an important and necessary contribution to land-use modelling and land system science by supporting both scholars and practitioners in becoming acquainted with proper evaluation and validation methods for land-change modelling applications. Given these challenges, alternative GIS-based approaches were explored for modelling agricultural land-use dynamics in the Beja district. Thus, Chapter 6 explores a set of spatial metrics and statistical methods as well as the Markov model, which was demonstrated to be a useful approach with high applicability and flexibility according to the research objectives. This chapter provides more instructive insights on the inter-class transitions than the most often used measurements usually provide and revealed a potential future scenario of agricultural land use in a region of Portugal that has great importance in the regional agricultural production sector. Chapter 7 presents an alternative and/or complementary approach to address some of the challenges in using empirical and conventional statistical methods and provided comprehensive assessments of the underlying factors explaining agricultural land-use. Together, the explored GIS-based approaches shed light on the pathways for improving our interpretation of agricultural land-use dynamics and provide formalised knowledge on promising approaches for analysing agricultural land systems.

Generally, the conducted research, combining both the spatial and temporal components as well as a long-term perspective, provided historical explanation with empirical geographical evidence under a space–time cross-reading. Nevertheless, the only way to provide long-term spatiotemporal perspectives is to use alternative data sources such as historical datasets relying on local, regional, or national statistics. However, such data sources encompassing long-time spans are commonly bound to different administrative units and, therefore, lack spatial detail. Chapter 8 illustrates such limitations owing to the nature of the data. Nevertheless, the proposed GIS-based approach has shown potential as a solution to overcome this limitation and integrate long spatial time-series in scientific analysis to provide a long-term spatiotemporal perspective and consolidate existing historical knowledge of agricultural production dynamics at the regional scale across different years, decades, and even centuries. Thus, this chapter breaks new ground by interpreting, quantifying, and visualising Portuguese regional agricultural production, conditioned by the nature of the statistical data. Moreover, this chapter provides a next step to expand our knowledge based on the ability to assess the past to understand the present and forecast probable future events. Also, demonstrates the possibility of new analyses that can strengthen and expand our knowledge when data with spatial, thematic, and temporal details are available.

Although the research in Chapters 3–8 was conducted within the Portuguese agricultural context, generally, most of the findings regarding the scientific and methodological approaches can make a significant addition to the current state of knowledge and contribute to the advancement of the field of agricultural land systems, land-use modelling, and land-use science. The insights contribute to sciences related to the earth and human society on a variety of levels, providing effective information for economic and environmental monitoring and evaluation, resulting in informed policy decisions, guiding toward sustainable and profitable pathways, and in the design of anticipatory measures to respond to change, which is of specific interest in tackling global societal challenges such as meeting global food production and food security challenges, sustainable development, and climate change. Nevertheless, research in this direction is not complete, as several issues remain open to further investigation. For instance, the research described in this thesis could be integrated during the formulation of long-term policies, planning, and management, as well as for cooperation among the public and private sectors. Nonetheless, basic aspects of model building, such as the selection of calibration data or statistical models, should continue to be explored to ensure sufficient quality. As the findings of each simulation yield a different result, future research should aim to test the advantages and limitations of each explored GIS-based approach as well as their complementarities. Moreover, agricultural land systems research requires cross-timescale analysis; thus, more detailed fine-scale analysis at the household level, based on, for example, farm-level data, is crucial for more sustainable agriculture or framing the best political options for agricultural land-use. In addition, it would be interesting to investigate the determinants of agricultural land-use dynamics by considering, for example, farmers' land-use decisions and their management strategies and motivations. If more detailed data regarding crop situation becomes available in the future, we recommend conducting similar analyses at different geographical contexts and spatial and temporal scales. Lastly, engaging geographical and historical perceptions has not been approached sufficiently in scientific analysis; thus, there are technical limitations concerning the availability of methods, tools, and data that necessitate more effective solutions.