INSTITUTO SUPERIOR DE AGRONOMIA
Universidade de Lisboa

Measuring Human Mobility in Rural Areas

Emerging Approaches, Methods, and Applications

Nuno Filipe Lopes Nunes

Dissertação para a obtenção do Grau de Mestre em

Engenharia Agronómica

Orientador: Doutora Maria João Prudêncio Rafael Canadas

Júri:

Presidente: Doutor José Carlos Santos Silva, Professor Associado, Instituto Superior de Agronomia da Universidade de Lisboa

Vogal: Doutora Maria João Prudêncio Rafael Canadas, Professora Auxiliar, Instituto Superior de Agronomia da Universidade de Lisboa

Vogal: Doutora Ana Maria Contente de Vinhas Novais, Professora Auxiliar, Instituto Superior de Agronomia da Universidade de Lisboa

2023

Universidade de Lisboa

Abstract

This study explores the viability of alternative methods for measuring human mobility in rural settings, with a focus on Portugal. It aims to assess whether these methods can effectively substitute or complement official statistics for modeling migration and mobility. Considering the variables from two rural case studies, the study evaluates emerging applications in the context of direct and indirect migration, commuting mobility, and recreational mobility. Based on a systematic literature review, the study underscores that high-resolution gridded datasets can offer precise population density estimates, however they may be better calibrated against urban areas. Direct migration measurements through gridded net migration datasets exhibit significant discrepancies from official statistics, highlighting a possible absence of current solutions. The availability of commuting and recreational mobility models is also limited geographically, and their adaptation to general rural studies requires considerable effort, which may not be feasible as secondary development for rural research applications. Ethical considerations concerning data privacy, responsible use of AI and potential dual use of technology have been sparsely noted or not significantly discussed in most studies reviewed. Looking forward, the current investment in the improvement of gridded net migration estimates may increasingly result in more accurate and fine estimates. Gradual gains in the understanding of commuting and recreational mobility in rural areas may arise from model transferability.

Keywords

Rural Mobility Modeling, Migration Patterns, Commuting Analysis, Gridded Datasets, Data Privacy in Mobility Research

Resumo

Este estudo explora a viabilidade de métodos alternativos em medições de mobilidade humana em ambientes rurais, com foco em Portugal. O estudo tem como objetivo avaliar a possibilidade de novos métodos poderem efetivamente substituir ou complementar estatísticas oficiais relativas a migração e mobilidade. Considerando as variáveis de dois estudos de caso rurais, o estudo avalia aplicações emergentes no contexto de migração direta e indireta, mobilidade pendular e mobilidade recreativa. Baseado em uma revisão sistemática da literatura, o estudo sublinha que grelhas populacionais de alta resolução podem oferecer estimativas precisas de densidade populacional, apresentando, no entanto, a possibilidade de estarem melhor calibradas para usos em áreas urbanas. Medições de migração direta através grelhas de migração líquida exibem discrepâncias significativas em relação a estatísticas oficiais, indicando uma possível ausência de alternativas correntes. A disponibilidade de modelos de mobilidade pendular e recreativa também se demonstrou limitada geograficamente, e a sua adaptação para estudos rurais requer esforço considerável, o que pode não ser viável como desenvolvimento secundário em aplicações de estudos rurais. Considerações éticas sobre privacidade de dados, uso responsável de IA e possível dupla intenção no uso de tecnologia foram apenas esparsamente notadas na maioria dos estudos incluídos na revisão. Olhando para o futuro, o atual investimento para melhorar grelhas de migração líquida pode resultar cada vez mais em estimativas precisas e detalhadas. Ganhos graduais na compreensão da mobilidade pendular e recreativa em áreas rurais podem também surgir da transferibilidade de modelos.

Palavras-chave

Modelos de Mobilidade Rural, Padrões de Migração, Análise de Deslocamento Pendular, Grelhas Populacionais, Privacidade em Pesquisa de Mobilidade

Resumo Expandido

Este estudo investiga a viabilidade de métodos alternativos para calcular variáveis de migração e mobilidade humana em contextos rurais, com foco específico em Portugal. O objetivo é o de avaliar métodos emergentes que apresentem potencial de substituir ou complementar estatísticas oficiais relativas a migração e mobilidade. Considerando as variáveis de dois estudos de caso rurais, o estudo avalia modelos e aplicações emergentes no contexto da migração direta e indireta, mobilidade pendular e mobilidade recreativa. Com base numa revisão sistemática da literatura, os resultados indicam que grelhas de estimativas populacionais de alta resolução podem oferecer estimativas precisas de densidade populacional. Cálculos de migração direta através de estimativas de migração líquida em grelha de 1000 m de resolução mostram discrepâncias significativas em relação a estatísticas oficiais. A disponibilidade de modelos de mobilidade pendular e recreativa encontra-se geograficamente limitada. Os ajustes que requerem para utilização em estudos rurais fora das áreas geográficas para as quais foram desenvolvidos requer um nível de esforço de programação e aquisição de dados que pode não ser viável como desenvolvimento secundário durante trabalhos de investigação rural. Várias considerações éticas relativas à privacidade de dados, ao uso responsável de inteligência artificial e potenciais usos diferenciais da mesma tecnologia para fins diferentes, mostraram-se pouco realçadas ou sem discussão significativa na maioria dos estudos revistos.

Durante este estudo, as dificuldades encontradas na aplicação de modelos recentes servem para realçar as múltiplas facetas da mobilidade humana. Se o seu entendimento é essencial para a compreensão de dinâmicas populacionais e para questões como seja o planeamento de políticas de desenvolvimento, também a sua natureza permanece em parte intangível. Uma primeira abordagem permite uma diferenciação em traços gerais entre conceitos de migração e mobilidade. Mobilidade tende a ser associada a todo o movimento de indivíduos, enquanto migração é normalmente utilizada como referência a movimentos de natureza mais permanente ou de longo prazo. Análises quantitativas de ambos os fenómenos apresentam desafios relativos à dificuldade de executar medições objetivas. Migração e mobilidade tornam-se complexas devido a uma grande diversidade e imprevisibilidade de comportamentos, a limitações em disponibilidade de dados, mas também a preocupações com privacidade e a desafios metodológicos. Assim mesmo, no meio de tal complexidade, a emergência de novas fontes de dados cada vez mais ricos em detalhe, bem como a rápida evolução de técnicas de modelação, têm contribuído para uma precisão crescente em modelos de migração e mobilidade.

Para estudar estes recentes avanços em modelação, a metodologia desenvolveu-se ao redor da necessidade de uma revisão sistemática da literatura e de testes de aplicabilidade para avaliar a relevância dos modelos identificados. Num primeiro passo foram selecionados dois estudos de caso relativos a investigação de aspetos do meio rural. Abreu et al. (2019) e Canadas et al. (2023), forneceram a base para a seleção de variáveis de estudo e meio para condução de testes de aplicabilidade. As variáveis de estudo foram selecionadas com base na sua capacidade de representar migração e mobilidade. De uma inicial extração de variáveis, surgiu a necessidade de agrupá-las em quatro categorias distintas: medida direta e medida indireta de migração, mobilidade pendular e mobilidade recreativa. Esta seleção de variáveis permitiu direcionar a revisão sistemática de literatura, através da pesquisa, identificação e avaliação de publicações relevantes. Testes de aplicabilidade envolveram a avaliação de modelos emergentes em relação às variáveis do estudo, usando software GIS para comparar estimativas obtidas através de modelos recentes com dados censitários.

Cada modelo foi descrito numa base de dados criada para o projeto com nome do modelo, fontes de dados, acesso às fontes de dados, variáveis produzidas pelos resultados, resolução espacial, resolução temporal, acesso aos resultados, métricas de precisão, metodologia, cobertura geográfica, casos de uso, vantagens, limitações e facilidade de uso. Quando viável, as soluções identificadas foram testadas para avaliar a sua aplicabilidade às variáveis selecionadas. As instâncias específicas em que testes de aplicabilidade foram realizados ou não puderam ser conduzidos, mas foram consideradas relevantes para exploração adicional, estão delineadas na seção 3.

Os resultados destacam que grelhas de estimativas populacionais de alta resolução podem oferecer estimativas precisas de densidade populacional, embora estas possam apresentar melhor calibração para áreas urbanas. A revisão identificou que estas estimativas de população em grelha se mostravam adequadas em medições indiretas de migração, como sejam densidade populacional, densidade populacional rural e variação de densidade populacional rural. Estas estimativas mostraram-se especialmente precisas quando utilizadas grelhas de maior resolução (100m). Estimativas de resolução mais grosseira (1000m) demonstraram pequenas inconsistências ao nível administrativo de freguesia, mantendo discrepâncias relativamente pequenas para cálculos comparativos entre vários anos como sejam numa variação da densidade populacional.

Medições de migração direta obtidas através de grelhas de estimativas de migração líquida mostraram discrepâncias significativas em relação a estatísticas oficiais. Alternativas para medidas diretas de migração, como migração líquida, apresentaram diferenças notáveis, mas a publicação de duas grelhas de estimativa globais no ano corrente talvez possa indicar melhorias potenciais nessa área. Medidas de mobilidade pendular e recreativa são abordadas por diferentes modelos, mas sua disponibilidade é limitada a regiões geográficas específicas ou difícil de ajustar, exigindo esforço significativo de codificação. Tal pode não ser viável como desenvolvimento secundário enquanto conduzindo outro tipo de estudo rural.

Os resultados revelam ainda que o campo de investigação dedicado ao desenvolvimento de modelos de migração e mobilidade está bem estabelecido, particularmente em contextos urbanos. Grelhas de estimativas populacionais com cobertura global mostraram-se adequados para medições de densidade populacional, densidade populacional rural e variação da densidade populacional rural, mas a resolução das grelhas populacionais tem impacto nos resultados. Enquanto uma mais alta resolução apresenta valores muito similares a valores dos censos, estimativas a mais baixa resolução espacial demonstram algumas inconsistências em níveis administrativos mais detalhados (freguesia) para a população total, enquanto mantêm discrepâncias relativamente pequenas na variação da densidade populacional.

Em suma, este estudo concluiu que a análise de abordagens emergentes na modelação de migração e mobilidade humana aplicáveis a estudos rurais revelou limitações relacionadas a cobertura geográfica, particularmente em áreas rurais de Portugal, e confirmou um foco urbano nos modelos emergentes, onde a mobilidade rural é, na melhor das hipóteses, uma componente de entrada em mobilidade urbana. Além da revisão de modelos alternativos, este estudo destacou as preocupações éticas transmitidas pelos estudos revisados. Embora apontem para um consenso sobre preservação de privacidade, discussões sobre consentimento, anonimização de dados ou a natureza ambivalente usos da tecnologia estiveram ausentes em muitos casos.

Em termos de futuros horizontes na investigação de mobilidade rural, este estudo permite antecipar uma potencial evolução de modelos que potencialmente seriam capazes de capturar um espectro mais variado de mobilidade rural. Por exemplo, a publicação de duas novas grelhas de estimativas globais de migração líquida, apesar de suas limitações atuais, sugere que em breve poderá ser possível analisar migração em maiores resoluções. Tal possibilidade poderia melhorar significativamente estimativas populacionais utilizadas em estudos rurais. Além disso, os modelos revistos têm melhorado com base em processos de treino utilizando dados em tempo real, maior resolução espacial e rápidos avanços tecnológicos. É possível que algumas soluções de modelos ultrapassem as fronteiras geográficas para que foram criados ou trainados, com uma potencial transferibilidade de modelos baseados no uso de fontes de dados mais facilmente acessíveis. Tal desenvolvimento será potencialmente mais relevante em termos da quantificação de mobilidade pendular e recreativa em ambientes rurais.

Por fim, vários modelos, com enfase nos que integram a previsão ou geração de fluxos de mobilidade, dependem de dados de movimento que apresentam novas questões. Além das preocupações mais comuns de privacidade relacionadas com a obtenção e uso de dados em si, tanto a fase de treino como a de geração de um modelo – especialmente quanto utilizando técnicas menos ou não explicáveis como "*deep learning*" – geram novos problemas de consentimento. Até que ponto a explicação detalhada de aspetos tão inerentes à natureza humana como mobilidade e migração foi ou pode ser consentida pelas pessoas cujo comportamento assim fica detalhado. As discussões atuais sobre ética em inteligência artificial merecem atenção tanto por aqueles que programam novos modelos, como por investigadores que procuram soluções para problemas específicos.

Table of Contents

List of Tables and Figures

1. Introduction

The interest to explore human mobility through emerging approaches, methods, and applications is informed by two distinct professional engagements. First, originating from the background on agricultural studies and the promotion of rural activities in Portugal. Second, through the contributions to international humanitarian assistance in cases of internal displacement. While these two areas are distinct, they provide useful contrasting views as parts of them mutually rely on detailed human mobility data. An understanding of how people move underlies initiatives related to development investment and wildfire mitigation in stable rural areas as it does allow effective delivery of humanitarian assistance in regions of displacement.

Both domains of activity present instances of dependency on mobility variables for capturing shifts in population numbers and geographic distribution. For instance, this data is used to target humanitarian aid delivery and in creating models for rural development or wildfire mitigation. In common, these cases tend to face challenges of timeliness, completeness, and relevance of official statistics for the proposed intents.

However, a meaningful distinction emerges from the urgency of accurate information gathering being particularly acute in humanitarian scenarios. There, action is contingent upon quick and accurate estimates of population displacement. On the other hand, in studies related to rural Portugal, the use of official statistics appears to be prevalent even towards the end of census cycles. In those cases, official census figures of population density, net migration, and commuting mobility collected in periods of ten years are likely to present some significant differences with reality.

In this context, the primary research question driving this study is whether emerging methods could be a valid replacement alternative for official statistics when obtaining migration and mobility data for rural models. The introductory sections briefly explore the significance of migration and mobility, their measurement challenges, data sources, variables, and tabulations. The study is based in the findings from a systematic literature review of potential methods for acquiring these variables, and of assessing the practicality of their use through applicability tests. Two specific studies, covering rural development models and collective wildfire mitigation and adaptation have been used as examples for testing applicability. Although this study has used concepts and frameworks acquired from professional experience in displacement scenarios, it is centered exclusively on a stable rural context in Portugal.

1.1. Migration and Mobility

1.1.1. Significance

Studies on internal migration and mobility as drivers of rural transformation have underscored the role of population mobility in shaping rural landscapes and societies (e.g., M. Bell & Muhidin, 2009; Bosworth et al., 2020; Greenwood, 1997; Nori & Farinella, 2020; J. L. Zhang & Bryant, 2020). Some studies, such as the research by M. M. Bell and Osti (2010) discussed the impact of mobility on the character and functionality of rural areas, highlighting how the movement of people within and into rural regions has redefined their socio-economic dynamics, land use patterns, and overall resilience. Moreover, Baptista (2010) while discussing Portugal and Nori & Farinella (2020) for the broader context of the Mediterranean, have provided insights into the transformations in rural areas and their links with migration and mobility.

Either explicitly, under the broader themes of migration (e.g., Lucas, 2007; Rye & O'Reilly, 2021) and mobility (Bosworth et al., 2020; Klous et al., 2020; Meredith et al., 2021), or implicitly, while focusing on specific issues such as context dependent depopulation implications (e.g., Syssner, 2020), abandonment of agricultural land (e.g., Rey Benayas et al., 2007), changes from space of production to space of consumption (e.g., Baptista, 2010; Eusébio et al., 2017), return to rural areas (e.g., Santos, 2023) , or tourism (e.g., López-Sanz et al., 2021), several works in contemporary rural studies discuss how components of migration and mobility influence demographic trends and shape rural communities.

Grasping migration and mobility dynamics is therefore important for managing and planning in rural areas. For instance, Nori & Farinella (2020) partially attributed the growing presence of immigrants in European rural areas to agricultural work and rural settings being decreasingly attractive to local populations, opening a more favorable environment for international immigrants. Others have discussed opportunities left by depopulation and abandonment in terms of attractiveness for international workers, tourists and other settlers (Eusébio et al., 2017). As the rural socio-cultural fabric and land use transform, evidenced by a shift from agricultural livelihoods to commutes to urban employment or international migration with only seasonal returns to the countryside, planning requires adaption to these evolving patterns (Baptista, 2010).

Addressing opportunities and challenges arising from these shifts in rural settings requires a nuanced understanding of migration and mobility. In rural studies, they have been analyzed alongside other demographic variables, such as family composition, aging, rates of fertility and mortality (Curtis & Kulcsár, 2019). Migration and mobility also impact core population metrics, including total population, population density, births, deaths, in-migration, and out-migration (Smith, 1992).

Similarly to humanitarian contexts, the significance of migration and mobility data lies in its ability to enable inference of personal or group presence at a specified location, at a defined moment in time (European Commission & United Nations, 2020). Such ability creates possibilities for applications across different domains, while it also generates different types of concerns. This research specifically looks at the possibility of deriving migration and mobility from output variables in emerging models.

1.1.2. Definitions

In general, migration and mobility represent distinct phenomena in literature, but the boundaries between both can be fluid. Broadly, the term 'mobility' is associated to the movement of individuals (Cresswell, 2006; Salah et al., 2022; Schwerdtle et al., 2020; Wesolowski et al., 2013), although, for some authors, the term is restricted to movement within a specific area or region (Barbosa-Filho et al., 2018; European Commission. Joint Research Centre., 2022; Klous et al., 2020; Kraemer et al., 2020; Ruktanonchai et al., 2021; Sorichetta et al., 2015). Some perspectives define mobility as an ability to move or travel (Greenwood, 1997; Kälin, 2022).

On the other hand, migration typically refers to more permanent or long-term movements of individuals from one place to another (Barbosa-Filho et al., 2018; European Commission. Joint Research Centre., 2022; Greenwood, 1985; Kälin, 2022; Klous et al., 2020; Kraemer et al., 2020; Ruktanonchai et al., 2021; Salah et al., 2022; Schwerdtle et al., 2020; Sorichetta et al., 2015; Wesolowski et al., 2013). While there is a general trend towards these definitions, the distinction is not always clear-cut. In several of these studies, factors such as the duration of stay, intentions, distance traveled, and frequency of movement can blur lines between migration and mobility.

However, for the purposes of this study, mobility is understood as all forms of movement regardless of distance, while migration is viewed as movements with an intention of long term or permanent relocation. This distinction, albeit artificial for many nuanced cases, provides important structure for the categorization of variables and analysis of the studies retrieved during literature review.

1.1.3. Measurement challenges

Challenges in measuring migration and mobility have been underscored by different authors. The relationship between these challenges and the dynamic, diverse, and unpredictable nature of human behavior have been extensively noted (Alessandretti et al., 2020; Cresswell, 2006; Hasan et al., 2013; Kälin, 2022; Solmaz & Turgut, 2019; Sorichetta et al., 2015). Mobility behaviors, shaped by multiple dimensions of reasons and motivations (Hasan et al., 2013; Toch et al., 2019), may be difficult to analyze in domains with fluid definitions and standards (Barbosa-Filho et al., 2018; European Commission. Joint Research Centre., 2022; IOM, 2019; Kälin, 2022). Adding to these issues, data availability limitations (Barbosa-Filho et al., 2018; Greenwood, 1985; Wesolowski et al., 2013), privacy concerns (e.g., X. Li et al., 2021; Sorichetta et al., 2015) leading to individual choices for not producing location data (Luca et al., 2020), and broader methodological challenges in data collection (Klous et al., 2020; Solmaz & Turgut, 2019) further complicate objectivity in measurements.

In addition, there are disparities in the geographical scope, volume, and quality of data, as highlighted by Kraemer et al. (2020), and problems specific to tracking temporal changes and seasonal movements (Ruktanonchai et al., 2021). The challenge extends to the desired scale – human and geographic – of a study (e.g., Barbosa-Filho et al., 2018), and the knowledge that previous research has shown significant variability in predictive performance (Cuttone et al., 2018). The combination of these different problems has impacts on temporal and spatial accuracies expected from mobility and migration estimates (Cuttone et al., 2018; Greenwood, 1985).

If these difficulties verified the intangibility of human mobility (Creswell, 1994), contemporary research is creating a new evidence space. Results from studies by Ruktanonchai et al. (2021), focusing on practical geospacial and sociodemographic predictors of human mobility, and Toch et al. (2019), surveying machine learning methods and applications for analyzing large-scale human mobility data, partially contradict notions of intangibility. These studies are examples of how research resorting to modern positioning technologies - such as GPS, cellular triangulation, and WiFi – along with machine learning techniques, is making human movement patterns increasingly discernible and analyzable. Others, such as the use of mobile phone data for passive collection, have also yielded detailed seasonal movement patterns across various regions (Ruktanonchai et al., 2021).

It is in this evolving context, that the current research aims to explore the recent landscapes of migration and mobility studies as they have been influenced by technological advacement. It specifically looks at the potential and practicality of using emerging methods to quantify variables of migration and mobility in geographically localized models.

1.1.4. Data sources

The emergence of data sources has contributed to recent improvement in the measurements of mobility. The International Recommendations on Internally Displaced

Persons Statistics (IRIS) (European Commission & United Nations, 2020) provide a useful framework for understanding these sources in the context of mobility data. This study has adopted several taxonomies from IRIS.

IRIS proposes three main categories of data sources. The first encompasses official sources, which are produced by national statistical offices and meet specific statistical quality standards. These sources include population and housing census, sample household surveys, and administrative data and registers. The second category, operational data sources, comprises of data collected by governments or humanitarian agencies for specific purposes like humanitarian assistance. This type of data is likely to present variability in terms of statistical quality. The third category encompasses other data sources, including alternative such as big data from mobile network operators. These can be combined with official and operational data or used independently in mobility analysis (European Commission & United Nations, 2020).

Recent advances in mobility modeling have particularly benefited from the latter two categories: operational and alternative data sources. Alternative data sources, such as internet traffic, satellite imagery, social media locations, or mobile phone data, provide precious information for validation and inference processes (European Commission & United Nations, 2020, p. 73). Operational data sources are designed for addressing specific mobility-related issues, such as internal displacement as outlined by IRIS. In situations where official statistics are impractical, operational data sources that follow specific quality standards can potentially transition into official statistics (European Commission & United Nations, 2020, p. 74). Operational data sources are likely to provide accuracy and detail for specific types of mobility and their associated challenges. As explored in results and discussion, the combination of these two categories with official data sources has significantly increased readability of some types of human movement.

In this research, the IRIS framework was employed to classify data sources during the systemic review.

1.1.5. Variables and tabulations

Another section from IRIS with influence on this study covers variables and tabulations. IRIS defines basic classificatory variables, or variables at the basis for collecting and compiling statistics in the type of internal mobility it covers. Although several variables are introduced, those specifically related to migration and mobility are relatively simple: district or administrative area of habitual and current residence (European Commission & United Nations, 2020, p. 58). Another concept of interest is the distinction that IRIS proposes between stock and flow measurements: stock as the total number of people under a certain category of mobility (e.g., in IRIS that category is internal displacement) in a specified location, at a defined moment in time, and flow as a dynamic measure of how many people entered or exited from a certain mobility category (e.g., internal displacement) within a particular time-period (European Commission & United Nations, 2020, p. 33).

This study uses the IRIS framework for variables and tabulations to inform the extraction of variables from two case studies and to analyze models identified in the literature review.

1.2. Emerging Methods and Ethical debates

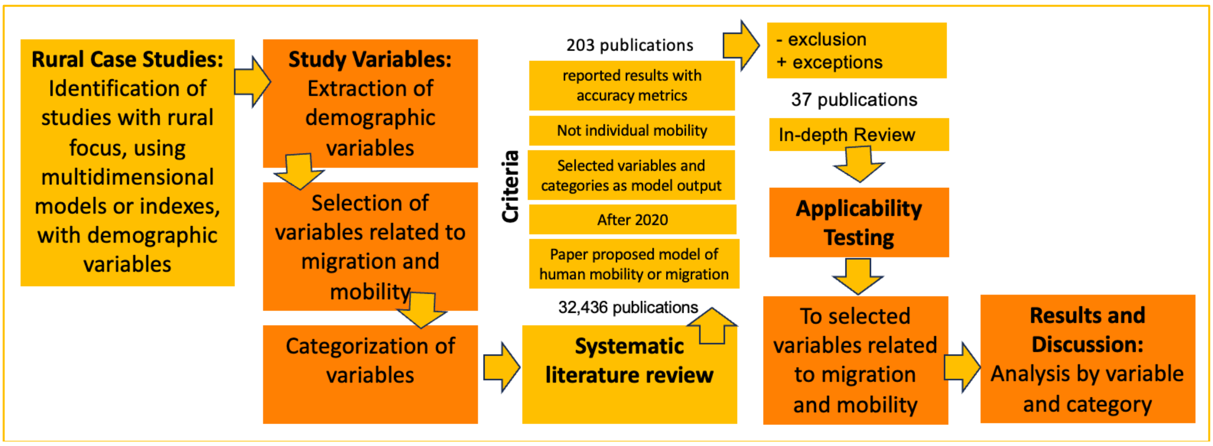
This study is based on the assumptions that emerging models are increasingly accurate and such accuracy is closely tied to the availability of data and the robustness of machine learning techniques. The research design included notions of gridded population estimations with increasing readability of potential location as well as distinctions between predictive and generative models. The awareness of flow generation models, or the synthetic creation of realistic mobility datasets among locations without specific knowledge of real flows (Luca et al., 2020) and of deep neural networks in a deep gravity model for explaining mobility (Simini et al., 2022) underpinned the motivations for the current research. Furthermore, there was an initial assumption that emerging methods tended to respond more to urban modeling needs than to needs in rural settings. Reviewed methods were therefore scrutinized for their original development context – whether rural or urban.

Advances in methods are accompanied by many ethical problems. This research is informed by a broader ethical discourse shaped by recent developments in legal frameworks and guidelines for responsible AI implementation. It is also informed by diverse ethics debates which have gained prominence in recent years, leading to different sets of principles and guidelines for ethical artificial intelligence (e.g., see Fjeld et al., 2020; Floridi et al., 2018; Floridi & Cowls, 2021; Jobin et al., 2019; Mittelstadt et al., 2016). The ethical dimension, although only briefly touched upon in this study, is extensively explored in a simultaneous PhD project around data and agency of displaced persons. Mainly retained for this study, the collection and utilization of data, particularly in rural areas where anonymity may potentially be compromised more easily, needs careful attention to informed consent, deidentification processing, and prevention of re-identification. Ongoing discussions on model transparency and the dual-use nature of technology for both beneficial and potentially detrimental purposes, such as surveillance, remain central to ethical concern.

2. Research Methods

The research methodology employed in this study was a systematic approach to literature review, aimed at isolating and analyzing demographic variables within rural-focused studies that utilize modeling techniques. As a first step, the process involved identifying case studies that employ demographic variables in the context of rural development. From these, variables specifically related to migration and mobility were extracted and subsequently categorized. A second phase entailed a systematic literature review of publications post-2020 that proposed models of human mobility or migration, resulting in a pool of 32,436 publications. A set of inclusion and exclusion criteria, including the relevance to individual mobility and the presence of accuracy metrics in the results, narrowed down the number to 203 publications. After further refinement and the application of exception criteria, an in-depth review of 37 publications was conducted. In a third phase, the selected works were submitted to applicability testing for assessing the relevance and practicality of the variables and categories identified. The final phase of the methodology was an analytical assessment of the variables related to migration and mobility by category, resulting in an overview and discussion of current modeling approaches with relevance in rural mobility. The different methodological steps are illustrated in Figure 1.

Figure 1. *Methodological Flowchart and Literature Review Process for* "*Measuring Human Mobility in Rural Areas: Emerging Approaches, Methods, and Applications*".



2.1. Research Question

The research question underpinning this study was "Is it possible to retrieve migration and mobility variables from emerging methods instead of official statistics, and can those methods be used in practical rural modelling applications?".

2.2. Rural Case Studies

The research incorporates a systematic literature review to identify and test emerging models and applications relevant to rural settings. In selecting case studies for methodological guidance and to test the applicability of variables, the scope was more targeted. The studies chosen, specifically those by Abreu et al. (2019) and Canadas et al. (2023), fulfilled specific criteria: they were recent publications that employed modeling or indexing approaches suitable for rural contexts in Portugal and utilized a multidimensional array of variables. These two studies were influential as foundation for the selection of variables and testing of a real-world application of alternative measurements.

Abreu et al. (2019) proposes the use of a rural development index to support planning and decision making in public investment. Although a detailed validation of the index is not within this study's remit, the work of Abreu et al. (2019) demonstrates how migration and mobility are integral to rural planning. Within the population dimension, the index includes factors such as population density, rate of natural increase, demographic dependency, and net migration. Here, net migration is conceptualized as a measure of a region's "attractiveness capacity" – the net balance between immigration *into* and emigration *from* a specific geographic region during a given period. This approach to attractiveness is based on the author's assumption that a higher index of rural development may lead to a higher capacity on attracting population (Abreu et al., 2019). While different perspectives and statistical definitions of attractiveness may arise, the quantification of population movements into and out of a geographic area remains relevant in planning of rural public investment and development initiatives.

Canadas et al. (2023) propose an approach described as wildfire mitigation and adaptation (WM&A), which encompasses prevention measures and the creation of defensible or survivable spaces as an alternative to fire suppression (Canadas et al., 2023). They have studied five key dimensions of factors potentially influencing collective WM&A actions, namely land use and land coverage, institutions, external resources, wildfires, and similar to the previous example from Abreu (2019), a population dimension. Only in this case, population is described by distinct variables, such as population density and its variation in rural areas, structure, education, occupation, followed by measurements of migration and mobility, such as residence outside the same parish 5 years before, mobility to work or study in another parish in the same municipality, or proxies such as seasonal/secondary dwellings, tourism and leisure indicators (Canadas et al., 2023, p. 3).

The use of migration and mobility dimensions in these two studies illustrates their significance to rural issues. The use of related indicators demonstrates the application of migration and mobility measurements for issues as diverse as wildfire management and rural development planning. Furthermore, of interest for the current research, these studies reveal a common reliance on official census data to quantify those variables (Tables 1 and 2 provide list of variables and sources used in the two studies).

2.3. Study Variables

This section outlines the process of selection and categorization of study variables drawing from the works of Abreu et al. (2019) and Canadas et al. (2023). Variables have been selected based on their relevance to the research question and their ability to represent migration and mobility. The process entailed an understanding of the studies, the extraction of variables, and their categorization.

Abreu et al. (2019) Rural Development Index (RDI), was formulated with the intention of enabling comparative analysis of rural development dimensions across different territorial units and the identification of geographical areas where public development investment could produce better results. The study acknowledged the complexity of targeting policy interventions in rural areas based on requirements for local-regional specificity. In this context the RDI was constructed based on the four dimensions, each composed of different variables, compiled for 15 municipalities in the North Alentejo Region of Portugal. Table 1 extracts the variables that compose the population dimension (Abreu et al., 2019). A complete list of variables composing the population, social, economic, and environmental dimensions has been included in annex 1.

Table 1. *Variables in the Population Dimension of the Rural Development Index (RDI) Proposed by Abreu et al. (2019).*

Acronym	Variable name	Year	Unit	Source
PopDens	Population density	2011	Nr/km2	(1)
NatInc	Rate of natural increase	2011	%	(1)
NetMig	Net migration	2011	Nr	(1)
DmgDep	Demographic dependency index	2011	%	(1)

Note: (1) Population and building Census, Statistics Portugal. (Adapted from Abreu et al., 2019, pp. 1112–1113).

Abreu et al. (2019) acknowledged methodological challenges in the selection of variables, weighting, and aggregation techniques. Of relevance for this research, the study revealed that the most developed municipalities according to RDI values were those with highest population density. Furthermore, findings also suggested that positive natural increases in local population and a decrease in aging populations were positively related to socio-economic development of rural areas (Abreu et al., 2019).

The definition proposed for each population variable by Abreu et al. (2019) was examined to understand their relevance to migration and mobility. The 4 variables (PopDens, NatInc, NetMig, DmgDep), were categorized in relation to how they directly or indirectly related to migration and mobility patterns. For this study Net Migration (NetMig) was considered under Direct Migration (DM) as a direct measurement of migration in the population balance equation. Population Density (PopDens) was considered under Indirect Migration (IM), for the relations suggested by the authors with the potential attractiveness of a region (Abreu et al., 2019, p. 1112), but also from the dependence of the total population used in the calculation of density, from a migration component of the population balance equation. The Rate of Natural Increase (NatInc) and the Demographic Dependency Index (DmgDep) were considered as Other Demographic or Socioeconomic variables (ODS), rather than direct or indirect migration and mobility measurements based on their higher dependency on natural variables (births and deaths) in the population balance equation (Brettell & Hollifield, 2023; McFalls, 2007).

In the second study, Canadas et al. (2023) explored collective wildfire mitigation and adaptation efforts across 116 parishes within a wildfire-prone region of Portugal. The research aimed to uncover local factors and external resources influencing collective critical actions. The study employed primary and secondary data, and techniques such as principal component analysis (PCA) and random forest modeling for the exploration of relationships between variables. The authors suggested a new approach by identifying structural factors underpinning collective wildfire mitigation and adaptation in Mediterranean Europe. The study's variables were grouped into five dimensions: land use/land coverage (LULC), population, institutions, external resources, and wildfire experience. Table 2 extracts the variables within the population dimension for the model proposed by Canadas et al. (2023). A complete list of variables included in the land use/land coverage (LULC), population, institutions, external resources, and wildfire experience has been included in annex 2.

Canadas et al. (2023) worked with independent variables mostly available with values at parish level, and estimated night stays by factoring in accommodation capacity and municipality-level data. The authors have suggested that a given community may be characterized by a particular combination of factors leading to a unique action mix, contending that effects of these factors and their interactions on the action mix can be modelled to produce more generalized knowledge of the effect of each factor across contexts.The relationships between these variables and collective wildfire mitigation and adaptation actions, were studied with a random forest modeling approach. The authors used mean squared error (MSE) to measure the impact of each variable on model accuracy. Furthermore, partial dependence plots were used for visual analysis of the marginal effects of each variable. The entire study area was rural. In terms of results, parishes with larger proportions of built-up areas and higher population densities demonstrated more substantial efforts in wildfire mitigation under the Forest Intervention Zone (FIZ) framework. Conversely, lower social vitality, characterized by an older population, older buildings, lower population density, and less tourists, appeared to pose challenges to mitigation. Social vitality, as identified, seemed to foster trust, knowledge sharing, and reduce transaction costs in collective action. (Canadas et al., 2023).

Table 2. *Variables in the Population Dimension of the Model Proposed by Canadas et al. (2023)*

Acronym	Variable name	Year	Unit	Source
PopDens	Population density	2011	Nr/km2	(1)
RPopVar	Variation of rural population density	1981–2011	%	(1)
Young	Proportion of young population (< 15 years)	2011	%	(1)
Elderly	Proportion of elderly population (≥ 65 years)	2011	%	(1)
Aging	Aging index	2011	%	(1)
EldAlone	Proportion of elderly living alone or with others in the same age group	2011	%	(1)
Illitera	Illiteracy rate	2011	%	(1)
Educat	Proportion of the population with secondary or higher education	2011	%	(1)
PrimSect	Proportion of the population working in the primary sector	2011	%	(1)
LiveOutP	Proportion of the population that lived outside the parish 5 years ago	2011	%	(1)
WorkOutP	Proportion of the population that works or studies in another parish in the same municipality	2011	%	(1)
Seasonal	Proportion of seasonal/secondary dwellings	2011	%	(1)
BuildAge	Average age of buildings	2011	Nr. yea	(1)
FamLabor	Proportion of farm family labor - annual work unit (AWU)	2009	%	(2)
TourisAg	Tourist entertainment companies	2021	Nr/100 km2	(3)
AccomCap	Accommodation capacity (number of beds/people)	2021	Nr/1000 inhabitants	(3)
NightSty	Night stays	2020	Nr/km2	(3)
RivBeach	River beaches and bathing areas	2020	Nr/1000 inhabitants	(4)

Sources: (1) Population and building Census, Statistics Portugal; (2) General Agricultural Census, Statistics Portugal; (3) Portuguese National Tourism Institute; (4) River beaches of

Portugal. Note. The complete list of variables across land use/land coverage (LULC), population, institutions, external resources, and wildfire experience dimensions has been included in annex 2 (Adapted from Canadas et al., 2023, p. 11).

As for the previous case, the variables proposed by Canadas et al. (2023) within the population dimension were examined to categorize their relation to migration and mobility patterns. The variable Population that Lived Outside the Parish 5 Years Ago (LiveOutP), was considered as a direct migration measurement contributing to incoming migration in the population balance equation, thus classified as DM. Variables indirectly related to migration patterns, such as PopDens and Variation of Rural Population Density (RPopVar), are influenced by population movements but do not serve as direct measurements of migration. PopDens and RPopVar were thus considered as IM. The study from Canadas et al. (2023) prompted the introduction of additional categories. These include Recreational Mobility (RM), which encompass the variable Night Stays (NightSty), primarily tied to leisure and tourism-related mobility. Furthermore, the category of Commuting Mobility (CMM), featuring the variable Proportion of the Population that Works or Studies in Another Parish in the Same Municipality (WorkOutP), signifying routine mobility for work or educational purposes within the same municipality.

The remaining variables within the Population dimension proposed by Canadas et al. (2023), such as Proportion of young population of less than 15 years old (Young), Proportion of elderly population above 65 years old (Elderly), Aging index (Aging), Proportion of elderly living alone or with others in the same age group (EldAlone), Illiteracy rate (Illitera), Proportion of the population with secondary or higher education (Educat), Proportion of the population working in the primary sector (PrimSect), Proportion of seasonal/secondary dwellings (Seasonal), Average age of buildings (BuildAge), Proportion of farm family labor - annual work unit (FamLabor), Tourist entertainment companies (TourisAg), Accommodation capacity (AccomCap), and River beaches and bathing areas (RivBeach) were considered as Other Demographic or Socioeconomic variables (ODSM), rather than direct or indirect migration and mobility measurements. The exclusion of Proportion of seasonal/secondary dwellings (Seasonal), Tourist entertainment companies (TourisAg), Accommodation capacity (AccomCap), and River beaches and bathing areas (RivBeach) was motivated by the presence of a more direct variable of recreational mobility in the measurement of night stays. However, these 4 variables could be further explored in future research.

The categorization of the variables extracted from both works and included in this study, based on the type of relationships with migration and mobility, is illustrated in table 3.

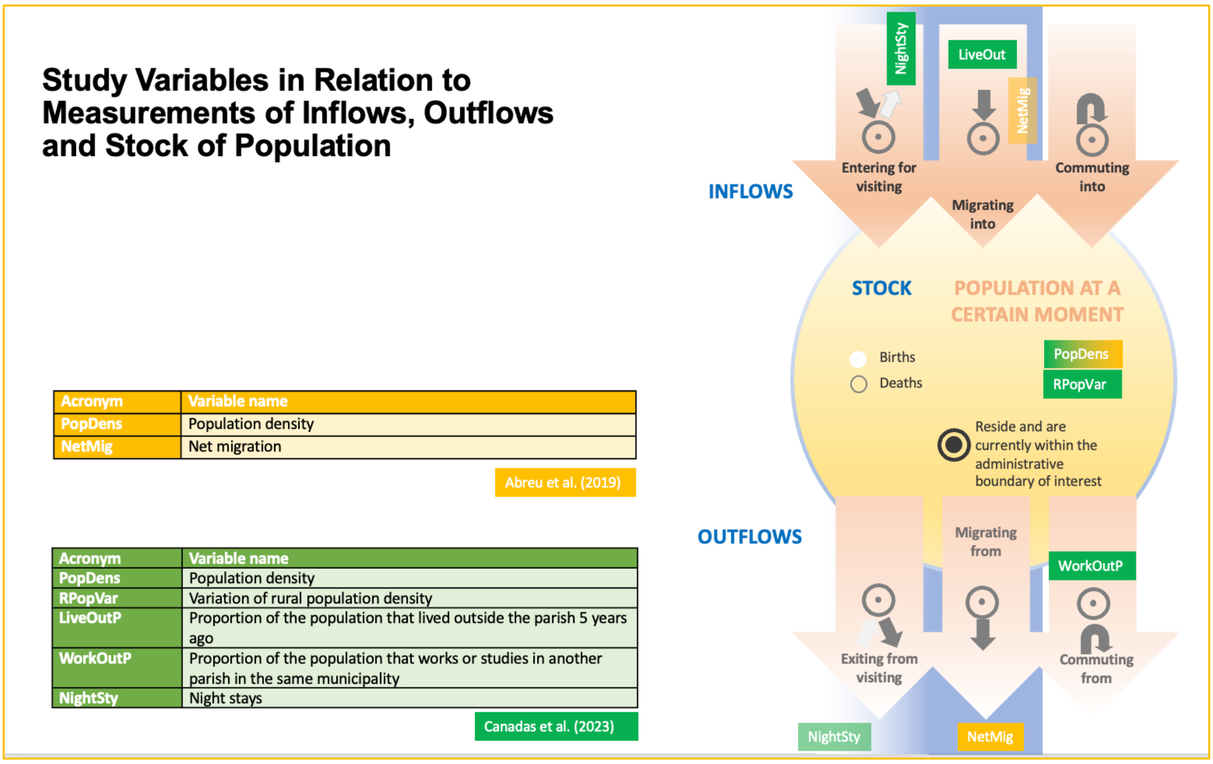
Table 3. *Population Variables from Abreu et al. (2019) and Canadas et al. (2023) in terms of Categorization and Inclusion as Study Variables.*

Variable	In Abreu et al. (2019)	In Canadas et al. (2023)	Category	Inclusion
PopDens	**Yes**	**Yes**	**IM**	**Yes**
NatInc	Yes	No	ODS	No
NetMig	**Yes**	**No**	**DM**	**Yes**
DmgDep	Yes	No	ODS	No
RPopVar	**No**	**Yes**	**IM**	**Yes**
Young	No	Yes	ODS	No
Elderly	No	Yes	ODS	No
Aging	No	Yes	ODS	No
EldAlone	No	Yes	ODS	No
Illitera	No	Yes	ODS	No
Educat	No	Yes	ODS	No
PrimSect	No	Yes	ODS	No
LiveOutP	**No**	**Yes**	**DM**	**Yes**
WorkOutP	**No**	**Yes**	**CM**	**Yes**
Seasonal	No	Yes	ODS	No
BuildAge	No	Yes	ODS	No
FamLabor	No	Yes	ODS	No
TourisAg	No	Yes	ODS	No
AccomCap	No	Yes	ODS	No
NightSty	**No**	**Yes**	**RM**	**Yes**
RivBeach	No	Yes	ODS	No

(IM) Indirect Migration, (DM) Direct Migration, (CM) Commuting Mobility, (RM) Recreational Mobility measurements. Marked in green the variables selected for this study.

Figure 2 seeks to contextualize how the selected study variables from Abreu et al. (2019) and Canadas et al. (2023) relate to demographic inflows, outflows, and the stock of population within rural areas. It should be noted that the placement of these variables within a framework of demographic flows and stock does not intend to evaluate their application in the two reference studies nor provide commentary on their relevance to them. The visual representation provides important tabulations for the analysis undertaken in the current research.

Figure 2. *Representation of selected study variables in relation to demographic inflows, outflows, and the stock of population within rural areas.*



Acronym	Variable name
PopDens	Population density
NetMig	Net migration

Acronym	Variable name
PopDens	Population density
RPopVar	Variation of rural population density
LiveOutP	Proportion of the population that lived outside the parish 5 years ago
WorkOutP	Proportion of the population that works or studies in another parish in the same municipality
NightSty	Night stays

The inflows at the top of Figure 2 depict three main types of movements: visiting, commuting into, and migrating into the area. This representation illustrates how the variables selected do not directly measure all expected flows but, in some cases, provide proxy information to part of them. For example, NightSty only offers partial data about the visiting inflow as it does not account for all visitors outside tourism-related facilities. LiveOutP indicates the proportion of the current population that lived outside the parish five years before but provides limited information of a total migration inflow compounded by recent migrations. In contrast, NetMig, positioned between inflows and outflows, is a net measure that accounts for the balance between migration movements into and from the area. For the outflows, at the bottom, WorkOutP, relates to the portion of the population that commutes out of the area. A broader understanding of total commuting flows would require a counterpart variable for incoming movements.

At the center, the stock represents the total population at a given moment. The stock is primarily influenced by the balance of inflows and outflows, and natural changes in resident population such as births and deaths. The current research has centered on mobility and not on natural demographic change. Such limitation needs to be considered in the analysis of variables like Population Density (PopDens) and Variation of Rural Population Density (RPopVar), as they are not necessarily linked to this conception of total stock and transient flows (commuting and visiting). Instead, their calculations are based on the resident population within rural classifications, already affected by births and deaths, and not on the total stock that includes commuters and visitors. However, migration flows, as they affect place of residence, can be partially captured in PopDens and RPopDens if considering properly measurements of births and deaths. Furthermore, RPopVar positioning in the diagram is contingent on the entire stock being within a rural context.

The use of this framework served to align selected variables within a conceptual framework of this research, assisting in the identification of potential limitations of analysis.

2.4. Systematic Literature Review

To identify relevant alternative measurement models for the selected variables, a systematic literature review was performed based on a search in PubMed, Scopus and Google Scholar, for studies published since 2020, using the following search terms in different combinations between title/abstract and keywords: “rural mobility”, “crowd flow”, “human mobility”, “flow generation”, “deep learning”, “net migration”, “population density”, “population distribution”, or “population model”. This search resulted in 32,436 publications. Those were restricted to studies on models of human mobility, and selected for inclusion if the study proposed a model related to study variables, provided outputs not on individual level (e.g., excluded individual trajectory prediction) and reported results with accuracy metrics, resulting in 203 publications. Articles related to traffic management, traffic predictions, airline predictions, covid-19 mobility measures or impacts of mobility in disease spread were not included. Articles not proposing possible solutions for the variables in study were subject to light review. When identified methods required literature from years prior to 2020, or recent surveys before that year were systematically retrieved in review, they were included in the study. After reviewing article abstracts and methods a total of 37 articles were considered suitable to include in the literature review (annex 3). Each model was reviewed for aspects such as name, data sources, access to data sources, output variables, spatial resolution, temporal resolution, access to output, accuracy metrics, methodology, geographic coverage, use cases, advantages, limitations, and ease of use. When feasible, proposed solutions from recent literature were tested to evaluate their applicability to the selected variables. The specific instances where these tests were conducted or not possible to conduct but considered relevant to further explore is delineated in the results and discussion section.

Additionally, this research contributes to a broader body of work conducted as part of a simultaneous PhD project titled “Understanding Agency and Autonomy of Displaced Persons in Humanitarian Contexts: Exploring the Interplay between Data, Technology Use, and Conceptual Frameworks”, which expands on current study by adding models specifically related or potentially impacting internal displacement, forced migration, and forced mobility, both in rural and in urban settings.

2.5. Applicability Tests

Applicability tests were conducted in cases garnering the following criteria: 1) model output variables coincided with study variables or provided the basis for their retrieval, 2) spatial resolution coincided with requirements for study variables, 3) model output covered the study area, and there was 4) open access to model output. When these tests were conducted the model output datasets used are identified under the relevant section with information on their open access source in footnote and summarized in annex 4.

For exploring the feasibility of georeferenced datasets, QGIS an open-source geographic information system (GIS) software application was used. Reference datasets that were frequently include the Official Administrative Map of Portugal – CAOP 2022, at municipal and parish levels, 2022 version (DGT, 2023) for administrative limits, and census data for Portugal from 2001 (INE, 2002) and 2011 (INE, 2012b).

A common procedure involved the use of the administrative limits in CAOP 2022 to extract the geographical area of interest at parish and municipality levels, from the two case studies. As output products from the models reviewed covered areas broader than the study area (and then Portugal in some cases), these administrative boundaries were first used as a mask to clip raster or basis for the selection of vector layers in extraction processes.

Results from testing estimates provided by models were evaluated for accuracy by comparison with the 2011 census (INE, 2012b), using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Percentage Error (MPE), and Mean Absolute Percentage Error (MAPE).

3. Results and Discussion

3.1. Indirect Migration Measurement

3.1.1. Population Density

Methods to obtain population density include gridded density estimations and indirect calculation based on gridded population count estimations. Leyk et al. (2019) reviewed large-scale gridded population data products in the year just before the time frame established for this study. The review is widely cited and included the Center for International Earth Science Information Network's (CIESIN) Gridded Population of the World (GPWv4.11) and Global Rural-Urban Mapping Project (GRUMPv1), the European Commission Joint Research Centre (JRC) and CIESIN's Global Human Settlement Population Layer (GHS-POP), Oak Ridge National Laboratory's LandScan; ESRI's World Population Estimate (WPE), WorldPop datasets, Facebook and CIESIN's High Resolution Settlement Layer (HRSL), JRC's European GHS Population Grid and the U.S. Census Bureau's country grids (Demobase). It covered modeling approaches, input data, and outputs, providing useful comparison and use cases for different population grids.

Leyk et al. (2019) highlighted lower accuracy of built-up land layers in rural settings, emphasizing the need for future work to improve population estimates in rural areas, where the reliability of existing data was limited (Leyk et al., 2019). The gridded products reviewed by Leyk et al. (2019) have remained the most used estimates for population distribution with broad geographical coverage found throughout this study, although their origins all predate the time frame established for the review (from 2020).

Among those, WorldPop datasets warrant special attention in this section, due to the publication in 2020 of 1km resolution population density grids for Portugal (WorldPop, 2020), following their 2018 publication of the Portugal unconstrained population grid at 100m resolution (WorldPop, 2018). Both are on yearly grids for the period 2000-2020. WorldPop is one of the highest modelled gridded products, based on population figures and ancillary data layers, such as roads, land cover, built structures, urban areas, nighttime lights, infrastructure, environmental data, protected areas, and water bodies (Leyk et al., 2019). It is based on dasymetric mapping for redistributing official population data to regular grids, where weight layers are established based on ancillary data (Leyk et al., 2019; Lloyd et al., 2017; Tatem, 2017),

Abreu et al. (2019) and Canadas et al. (2023) have both integrated population density from the Census 2011 (INE, 2012b) within their respective models.

Given that the WorldPop population density grids were based on the previous population count grids, the applicability test for the two case studies started by exploring the previous population count grids. When using yearly Unconstrained Population Counts at 100 m resolution (UPC-100m) and yearly unconstrained population density (UPD-1km) at 1 km resolution sourced from WorldPop gridded data, led to the conclusion that they present minimal differences with Census figures, both at parish and municipality level. Table 4 illustrates differences for the parishes in the municipality of Arganil.

Table 4. *Comparison of total population from WorldPop UPC-100m and 2011 Census Data for Arganil Municipality*

Parish	Difference (Nr)	Population WorldPop 2011 (Nr)	Population Census 2011 (Nr)	Accuracy
Arganil	0,63	4001	4002	MAE 0,642969704 RMSE 0,86313669 MPE - 0,009598712 MAPE 0,09314822
Benfeita	-0,08	394	394	
Celavisa	0,00	182	182	
Folques	0,85	355	356	
Piódão	0,01	178	178	
Pomares	0,00	513	513	
Pombeiro da Beira	-1,87	1012	1010	
São Martinho da Cortiça	-0,82	1320	1319	
Sarzedo	0,52	684	685	
Secarias	0,54	429	430	
Teixeira e Cepos	-0,01	270	270	
Cerdeira e Moura da Serra	1,38	438	439	
Coja e Barril de Alva	-1,00	1709	1708	
Vila Cova de Alva e Anceriz	-1,31	660	659	

The comparative performance of WorldPop estimates with Census data at parish (table 4) provided MAE 0,642969704, RMSE 0,86313669, MPE -0,009598712 and MAPE 0,09314822. At the parish level, WorldPop UPC-100 m resolution demonstrates potential as acceptable alternative for population estimates. For the census year, this outcome was somewhat expected, considering the use of census data in census downscaling used in WorldPop estimates, as elaborated by Lloyd et al. (Lloyd et al., 2017). Results from the more recent census round will allow to evaluate the performance for the most recent period.

Population density calculations for the geographical scope of Canadas et al. (2023) based on population counts from WorldPop UPC-100m for the area of the parishes delimited by CAOP-2022, aggregated by their respective municipality, were depicted in table 5 for the years of 2001 and 2011. Accuracy was evaluated for 2011 by comparison with Census figures for population density.

Table 5. *Population Density Based on WorldPop UPC-100 m Calculated for Parishes and Aggregated at Municipality Level for 2001 and 2011, in the geographical area covered by Canadas et al. (2023)*

Municipality	Population WorldPop UPC-100m 2001 (Nr)	Population WorldPop UPC-100m 2011 (Nr)	Calculated PopDens WorldPop UPC-100m 2001 (Nr/km2)	Calculated PopDens WorldPop UPC-100m 2011 (Nr/km2)	PopDens Census 2011 (Nr/km2)	Accuracy (2011)
Alvaiázere	8437	7288	52,6	45,4	45,4	MAE 0,031578947 RMSE 0,079471941, MPE 0,043792424 MAPE 0,04619312
Ansião	13671	13083	77,6	74,3	74,6	
Arganil	13624	12146	40,9	36,5	36,5	
Castanheira de Pêra	3733	3191	55,9	47,8	47,8	
Figueiró dos Vinhos	7348	6165	42,4	35,5	35,6	
Góis	4861	4259	18,5	16,2	16,2	
Lousã	15760	17612	113,9	127,3	127,2	
Mação	8441	7337	21,1	18,3	18,3	
Miranda do Corvo	13062	13091	103,4	103,6	103,6	
Oleiros	6677	5721	14,2	12,1	12,1	
Oliveira do Hospital	22112	20855	94,3	88,9	88,9	
Pampilhosa da Serra	5220	4482	13,2	11,3	11,3	
Pedrógão Grande	4397	3914	34,2	30,4	30,4	
Penela	6590	5980	48,9	44,4	44,4	
Proença-a-Nova	9611	8314	24,3	21	21	
Sertã	16720	15880	37,4	35,5	35,5	
Tábua	12601	12070	63,1	60,4	60,4	
Vila de Rei	3100	3452	16,2	18	18	
Vila Nova de Poiares	6804	7274	80,6	86,1	86,2	

Accuracy evaluations for the use of the WorlPop unconstrained population grid at 100m resolution (WorldPop, 2018) showed a possible acceptable alternative for population count and for population density calculations both at parish and municipality levels.

The alternative WorldPop UPD-1km resolution (WorldPop, 2020) could in principle be more straightforward to use. However, the results are slightly different. When using the average density provided by zonal calculation at parish level aggregated by municipality there is an accumulation of errors. The same parishes in the municipality of Arganil are used in table 6, and the aggregation by municipality for the area studied by Canadas et al. (2023) in table 7, with accuracy evaluations for the census year (2011).

Table 6. *Comparison of population density from WorldPop UPC-100m, UPD-1km and 2011 Census Data for Arganil Municipality*

Parish	**PopDens WorldPop UPC-100m 2011 (Nr/km2)**	**PopDens Census 2011 (Nr/km2)**	**PopDens WorldPop UPD-1km 2011 (Nr/km2)**	**Accuracy WorldPop UPD-1km 2011**
Arganil	117,3	117,3	113,2	MAE 2,696151868 RMSE 3,893381964 MPE 1,551358936 MAPE 6,967031304
Benfeita	18,1	18,1	18,2	
Celavisa	11,9	11,9	11,6	
Folques	19,3	19,4	18,5	
Piódão	4,9	4,9	4,8	
Pomares	16,3	16,3	15,4	
Pombeiro da Beira	31,0	30,9	29,2	
São Martinho da Cortiça	41,8	41,8	41,5	
Sarzedo	59,2	59,3	65,5	
Secarias	61,9	62,0	56,8	
Teixeira e Cepos	8,3	8,3	8,0	
Cerdeira e Moura da Serra	23,8	23,8	29,6	
Coja e Barril de Alva	70,3	70,3	73,1	
Vila Cova de Alva e Anceriz	38,6	38,5	29,3	

Table 7. *Comparison of population density from WorldPop UPD-1km and 2011 Census Data from parish aggregation at Municipality-level*

Municipality	**Aggregated PopDens WorldPop UPD-1km 2001 (Nr/km2)**	**Aggregated PopDens WorldPop UPD-1km 2011 (Nr/km2)**	**PopDens Census 2011 (Nr/km2)**	**Parishes Aggregated PopDens WorldPopPD 2020 (Nr/km2)**	**Accuracy (2011)**
Alvaiázere	53,6	46,5	45,4	40,9	MAE 5,421052632 RMSE
Ansião	98,5	94,1	74,6	90,1	
Arganil	41,2	36,8	36,5	33,2	
Castanheira de Pêra	55,1	47,1	47,8	40,9	
Figueiró dos Vinhos	40,0	33,6	35,6	28,7	
Góis	27,0	23,7	16,2	21,1	
Lousã	96,8	107,9	127,2	119,2	

Mação	23,1	20,1	18,3	17,9	8,20391434
Miranda do Corvo	90,8	91,0	103,6	91,0	
Oleiros	12,5	10,7	12,1	9,3	MPE
Oliveira do Hospital	101,9	96,1	88,9	91,2	2,145828772
Pampilhosa da Serra	16,3	14,1	11,3	12,3	MAPE 11,5178744
Pedrógão Grande	34,3	30,6	30,4	27,5	
Penela	41,0	37,3	44,4	34,4	
Proença-a-Nova	21,6	18,7	21,0	16,4	
Sertã	38,6	36,8	35,5	35,1	
Tábua	61,4	58,8	60,4	56,5	
Vila de Rei	15,4	17,1	18,0	18,9	
Vila Nova de Poiares	68,2	72,8	86,2	77,1	

From this analysis emerges that utilizing WorldPop UPC-100m at the finer 100 m resolution, results in remarkably low error rates and a high level of accuracy when compared to census data. Accuracy diminishes when shifting to yearly WorldPop UPD-1km at a coarser 1km resolution, particularly when aggregated at municipality level, where evaluation metrics reveal a more significant divergence between WorldPop population density estimates and census figures. Use cases using parish level data, as the two examples used in this study, can benefit from the additional computation of population density from the yearly WorldPop fine-resolution estimates of population count (100 m), as they provide the most accurate representation of population density at a detailed spatial scale, and allow for yearly estimates. In contrast, coarser-resolution estimates (1km) of population density may cover larger areas without the need for additional density calculations, and as such less demanding in computational needs.

3.1.2. Variation of Rural Population Density

Variation of rural population density (RPopVar) builds upon the sources and methods discussed in section 3.1.1., with the added requirement of differentiating between population in rural and non-rural areas. The simplest approach, in the case where the entire study area was classified as rural, would be to use either the WorldPop yearly UPC-100m (WorldPop, 2018) or the yearly UPD-1kkm (WorldPop, 2020), available from 2000 to 2020, to calculate variations. Variations in population density, not accounting for rural and non-rural areas, can be computed both for a specific period and as annual variation by: VPopDens (%) = [(Population Density in the Most Recent Year of the Period - Population Density in the Earliest Year of Comparison) / Population Density in the Earliest Year of Comparison] x 100. The results obtained for VPopDens, before accounting for areas classified as rural, urban or peri urban, between 2001 and 2020, using both the yearly WorldPop UPC-100m (VPopDens A) and the yearly UPD-1km (VPopDens B), are illustrated in table 8.

Table 8. *Comparison of the Variation of Population Density at Municipality-level between 2001 and 2020, using WorldPop UPC-100m (VPopDens A) and UPD-1km (VPopDens B).*

Municipality	**(VPopDens A) WorldPop VPopDens UPC-100m 2001-2020 (%)**	**(VPopDens B) WorldPop VPopDens UPD-1km 2001-2020 (%)**	**Difference VPopDens A VPopDens B**	**Difference Evaluation**
Alvaiázere	-24,3%	-23,7%	-0,6%	Mean difference -0,7% STDEVP 0,00380329
Ansião	-8,0%	-8,5%	0,5%	
Arganil	-19,6%	-19,4%	-0,1%	
Castanheira de Pêra	-25,8%	-25,8%	0,0%	
Figueiró dos Vinhos	-28,3%	-28,3%	-0,1%	
Góis	-22,2%	-21,9%	-0,3%	
Lousã	23,4%	23,1%	0,3%	
Mação	-23,2%	-22,5%	-0,7%	
Miranda do Corvo	0,4%	0,2%	0,2%	
Oleiros	-25,4%	-25,6%	0,2%	
Oliveira do Hospital	-10,5%	-10,5%	0,0%	
Pampilhosa da Serra	-25,0%	-24,5%	-0,5%	
Pedrógão Grande	-19,9%	-19,8%	-0,1%	
Penela	-17,0%	-16,1%	-0,9%	
Proença-a-Nova	-23,9%	-24,1%	0,2%	
Sertã	-9,4%	-9,1%	-0,3%	
Tábua	-7,9%	-8,0%	0,1%	
Vila de Rei	22,8%	22,7%	0,1%	
Vila Nova de Poiares	13,5%	13,0%	0,5%	

Although the two methods (WorldPop VPopDens UPC-100m and WorldPop VPopDens UPD-1km) present differences for obtaining population density estimates, as discussed in 3.1.1, the calculation of variation at municipality level and over a long period of time results in a relatively small average discrepancy between the two methods (0.7%). For a long period, they seem to generally provide similar estimates for the variation of population density, with minor variations.

3.1.2.1. Identification of rural areas

The method above addressed variation of population density in all areas, including urban. The variable RPopVar in Canadas et al. (2023) requires to estimate variations in population density specifically in rural areas. Different authors (Moreno-Monroy et al., 2021; Ruktanonchai et al., 2021; Thomson et al., 2022) referred to the Settlement Model (SMOD) of the Global Human Settlement Layer (GHSL) from the European Union's Joint Research Centre (JRC) for classification of urban, peri-urban and rural areas. The gridded GHS-SMOD classification is based on a combination of population size, population density, and built up area density (European Commission. Statistical Office of the European Union., 2021). GHS-SMOD derives from two other GHSL datasets: GHS-Built which maps the extent of built-up areas and is composed of built-up surface, building height, built-up volume, and settlement characteristics, and GHS-Pop, a spatial raster dataset that captures population dynamics over time (1975-2030) at 5-year intervals. Deriving from these two datasets and the 'degree of urbanization method from EUROSTAT, GHS-SMOD delineates settlement typologies (European Commission. Joint Research Centre., 2023a).

GHS-SMOD classifies grid cells according to the following classes: 30: "Urban Centre"; 23: "Dense Urban Cluster", 22: "Semi-dense Urban Cluster", 21: "Suburban or per-urban", 13: "Rural cluster", 12: "Low Density Rural", 11: "Very low density rural", and 10: "Water". In this classification, the aggregation of the classes 30 – 23 – 22 – 21 produces the "urban domain", while the combined 13 – 12 – 11 – 10 classes form the "rural domain" (European Commission. Joint Research Centre., 2023a). Table 9 illustrates GHS SMOD and WorldPop UPC-100m combination for GHS SMOD cells classified as 10, 11, 12, 13, in the calculation of RPopVar for one of the municipalities (Arganil) in Canadas et al. (2023). It first established the rural population density (RPopDens) to calculate RPopVar between 2001 and 2011.

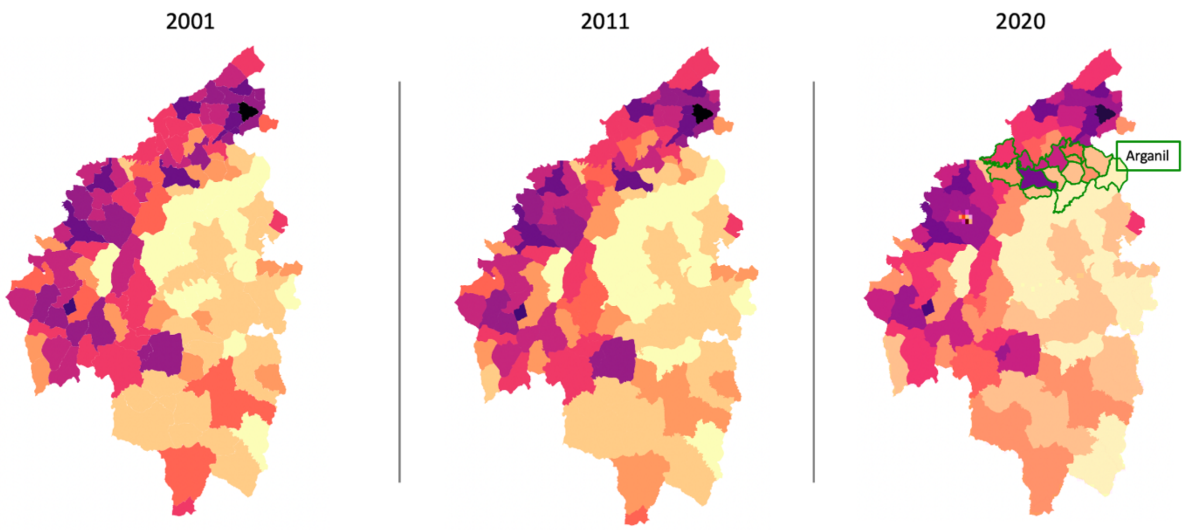
Table 9. *Variation in Rural Population Density between 2001, 2011, 2020) using WorldPop UPC-100m in combined GHS SMOD classes 11, 12 and 13 for calculation of Rural Population Density in the rural domain for the municipality of Arganil.*

Parish	**RPopDens 2001 (UPC-100m in GHS SMOD 11, 12, 13) (Nr/km2)**	**RPopDens 2011 (UPC-100m in GHS SMOD 11, 12, 13) (Nr/km2)**	**RPopDens 2020 (UPC-100m in GHS SMOD 11, 12, 13) (Nr/km2)**	**RPopVar 2001-2011 (%)**	**RPopVar 2011-2020** (%)	**RPopVar 2001-2020 (%)**
Arganil	131,775	117,308	105,811	-11,0%	-9,8%	-19,7%
Benfeita	20,29	18,098	16,325	-10,8%	-9,8%	-19,5%

Celavisa	13,319	11,917	10,747	-10,5%	-9,8%	-19,3%
Folques	21,762	19,348	17,461	-11,1%	-9,8%	-19,8%
Piódão	5,413	4,826	4,356	-10,8%	-9,7%	-19,5%
Pomares	18,123	16,137	14,541	-11,0%	-9,9%	-19,8%
Pombeiro da Beira	34,67	30,992	27,931	-10,6%	-9,9%	-19,4%
São Martinho da Cortiça	44,841	39,894	36,022	-11,0%	-9,7%	-19,7%
Sarzedo	66,28	59,234	53,404	-10,6%	-9,8%	-19,4%
Secarias	69,091	61,876	55,802	-10,4%	-9,8%	-19,2%
Cepos e Teixeira	9,187	8,254	7,444	-10,2%	-9,8%	-19,0%
Cerdeira e Moura da Serra	26,802	23,758	21,426	-11,4%	-9,8%	-20,1%
Côja e Barril de Alva	78,836	70,34	63,422	-10,8%	-9,8%	-19,6%
Vila Cova de Alva e Anseriz	43,187	38,556	34,763	-10,7%	-9,8%	-19,5%

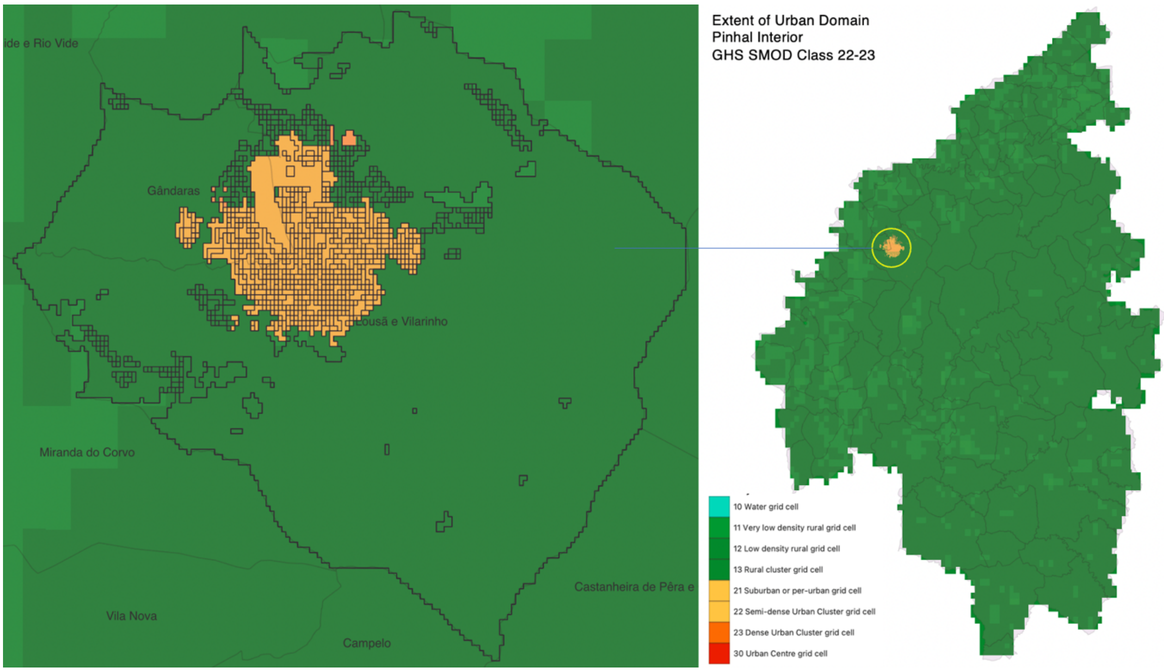
Results present similar values to those obtained by direct use of WorldPop UPC-100m to the parishes, as they all fall within areas classified as rural. Figure 1 shows a comparison of rural population density (lighter colors for lower values) for all parishes in Pinhal Interior for the years 2001, 2011 and 2020, using WorldPop UPC-100m and GHS SMOD classes 11, 12, 13.

Figure 3. *Visual Representation of RPopDens for 2001, 2011 and 2020 for the parishes in Pinhal Interior, using WorldPop UPC-100m and GHS SMOD rural domain.*



For the purposes of testing applicability, the relatively small area of Pinhal Interior classified by GHS SMOD as belonging to the urban domain, illustrated in figure 2, has been used for testing applicability. The urban domain is limited to parts of two parishes, Gândaras and Lousã e Vilarinho.

Figure 4. *Extent of urban domain according to GHS SMOD classification (classes 21-22-23) in Pinhal Interior.*



In this area, the overlay of GHS SMOD was used to calculate the values for population and area in both rural and urban domains. The resulting population densities are illustrated in table 10.

Table 10. *Rural Population Density for parishes with urban extents according to GHS SMOD classification.*

Parish	Urban Population 2020 (UPC-100m and GHS SMOD 21,22,23,30) (Nr)	Total Population 2020 (UPC-100m) (Nr)	Area (CAOP 2022) (ha)	PopDens 2020 (Nr/km2)	Area Urban Domain (GHS SMOD 21,22,23,30) (km2)	Urban PopDens 2020 (Nr/km2)	Rural PopDens 2020 (Nr/km2)
Gandaras	707	2218	1004	220,9	1,9	363,1	186,7
Lousa e Vilarinho	8130	22097	7240	305,2	6,3	1283,9	211,4

Yearly Rural PopDens is possible to calculate with the combined use of WorldPop UPC-100m and GHS-SMOD classification of rural and urban domains. From such values, following the methods in table 9 arises the possibility of RPopVar calculations between yearly estimates.

3.1.3. Ongoing advances in population density models

Several studies that continue to advance the understanding of population density, population distributions and disaggregation at finer scales and with higher accuracy, are useful to mention for the novelty. For instance, C. Li & Managi (2023) proposed an approach based on a random forest method to generate gridded population data for Japan, with more localized ancillary data, to retrieve higher accuracy in total population and male and female disaggregation (C. Li & Managi, 2023). Bao et al. (2022) proposed an alternative machine learning strategy, using an ensemble learning algorithm stacking of gradient boosting decision tree (GBDT), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM) and support vector regression (SVR) base models. The stacking yielded an better overall performance than the four individual models, and outperformed the commonly used random forest model for gridded population products (Bao et al., 2022). Other studies have also aimed at developing models for estimating population within buildings, either by using easier to access data (M. Wang et al., 2022), or much more difficult and expensive sources, such as LiDAR (H. Chen et al., 2021). Additionally, the use of proprietary 50cm resolution imagery and microcensus data, combined with a convolutional neural network (Res-Net 50) has been used to automatically extract features (such as buildings) from imagery to produce more census-independent estimates. However, this method comes with significant cost of acquiring the necessary imagery (Neal et al., 2022). In summary, recent advances in population density models involve model alternatives and the exploration of different ancillary data, often combining both for results tailored for a specific spatial scale (e.g., country), while continuing to evolve around the concept of gridded estimates. These findings hold relevance in terms of modelling possibilities but are not directly usable for the variables in this study, mostly because they area localized solutions.

Furthermore, publications on population distribution and density show a notable urban focus. Examples include, the availability of open datasets of ancillary data for population distribution research in European cities (Doda et al., 2022) and integration of social media and nighttime light data to enhance estimates in Zhejiang, China (L. Wang et al., 2020). This urban trend extends to studies of the accuracy of gridded estimates. Thomson et al. (2022) reported a lower accuracy when estimating population density in slum areas within urban Namibia (Thomson et al., 2022). Furthermore, assessments of gridded population estimates in cities across different income brackets highlighted less accurate estimations for low- and middle-income countries (Kuffer et al., 2022).

3.1.4. Ongoing advances in variation of population density models

Models for measuring variations in population density have witnessed significant developments, primarily focused on achieving higher spatiotemporal resolution. Bao et al. (2023) integrated ambient population data and building volume data to generate population distribution maps with hourly granularity and a spatial resolution of 100 meters in Beijing. The resulting high temporal resolution unveiled unknown patterns in variations of population distribution (Bao et al., 2023). Such high spatiotemporal resolution is relevant for sections 3.3 and 3.4. Other studies have explored different spatiotemporal frames. For instance, Teng et al. (2022) focused on evolving patterns of population distribution over a two-decade span, analyzing shifts in population agglomeration and urbanization trends, through spatial distribution maps of population density at a resolution of 1 km x 1 km from 2000 to 2020 in China (Teng et al., 2022). On a different temporal scale, Cheng et al. (2022) have mapped monthly population distribution and variation at a 1 km resolution in China, using time-series data from mobile phone positioning and employing a hybrid downscaling model incorporating random forest and area-to-point kriging. Such approach enhanced the resolution and accuracy of monthly estimates for population distribution and variation, especially during notable events such as festivals, holidays, and short-term labor flow periods. These findings hold relevance in terms of modelling possibilities but are also not directly usable for the variables in this study.

3.2. Direct Migration Measurement

3.2.1. Net migration

Abreu et al. (2019) defined Net Migration (NetMig) as the difference between immigration into and emigration from the region during a given period, reflecting according to the authors the attractiveness capacity of the territory (Abreu et al., 2019, p. 1112). The Portuguese Statistical Office (INE) included a time frame and only incoming movements to define attractiveness in 2011, as the ratio between resident population which 5 years before resided in another territorial unit or in another country and the total population resident in the territorial unit (INE, 2012a, p. 32). Along such lines, Canadas et al. (Canadas et al., 2023) used LiveOut as the proportion of the population that lived outside the parish 5 years before, with data from 2011 Portuguese Census.

Open access gridded datasets have also been proposed as solutions for net migration estimates. Recent research has attempted high-resolution gridded datasets for net migration. Namely, the study conducted by Niva et al. (2023), currently in preprint and not yet peer reviewed, proposes a 10 km resolution global gridded dataset of annual net migration covering the years between 2000 and 2019 (Niva, Horton, Virkki, Heino, Kallio, et al., 2023). The authors have used population counts from WorldPop combined with downscaled gridded births and deaths data for estimating net migration at sub-national level (admin level 2). This annual net migration dataset resulted from the collection and harmonization of national and sub-national-level birth and death rate datasets for 216 countries. The authors have then applied multiple linear regression, adjusted according to country income bands, and area-to-point kriging (ATPK) to downscale the subnational birth and death rates to 5 arcmin (~10 km at the equator) resolution. Downscaled births and deaths were used to calculate natural population change for a gridded cell, and then compared with the corresponding annual gridded population count (WorldPop), to calculate net migration for each year and grid cell (Niva, Horton, Virkki, Heino, Kosonen, et al., 2023). For this dataset the authors note that subnational validation data for population, births and deaths were collected at administrative level 2 (municipality) whereas production and harmonization of gridded datasets were based level 1 (province or district) for the global grid, and only this level 1 data (province or district) was used for validating the net migration estimates. The authors also recommend using the dataset for multiple years, with 3, 5 and 20 year net-migration sums provided at gridded level, or by aggregated over larger area due to some noise in gridded annual data (Niva, Horton, Virkki, Heino, Kosonen, et al., 2023). Results from using this dataset to the study area in one of the case studies were inconclusive, as shown in table 11.

Table 11. *Discrepancies between net migration figures in 2011 for the area studied by Abreu et al. (2019) according to the authors, the Census 2011 and the estimates retrieved from Niva et al. (2023) gridded dataset.*

Parishes	**Net migration 2011 Census 2011**	**Net migration 2011 Niva et al. 2023**	**Net migration 2011 Abreu et al. 2019**	**Evaluation Net migration 2011 Census 2011 vs Niva et al., 2023**
Alter do Chão	38	-10	11	MAE
Arronches	28	-6	33	61
Avis	59	-12	35	
Campo Maior	-1	-1	-18	RMSE
Castelo de Vide	85	-12	-7	70,546
Crato	59	-16	0	MPE
Elvas	55	-2	-173	182,9
Fronteira	30	-10	-27	
Gavião	60	-16	-13	

Marvão	59	-14	-8	MAPE 182,9
Monforte	50	-3	-12	
Nisa	1	-11	-22	
Ponte de Sor	137	-9	-61	
Portalegre	-31	-5	-214	
Sousel	88	-14	-15	

For this study the origin of the net migration values used in Abreu et al. (2019) was difficult to retrieve. In addition, the values for net migration from the census (2011) at municipality level and the estimates for the same year retrieved from processing the net migration grid from Niva et al. (2023) present significant inconsistencies. The comparison indicates notable problems of accuracy, with MAE 61, RMSE 70,5, MPE and MAPE values of 182.9, suggesting that Niva's estimates for the study area deviate substantially from the census figures. While the approach is novel in proposing gridded net migration values, indicating a new branch for gridded estimates, there is still a need for improvement in the estimation models, or the ancillary data used to downscale births and deaths.

Another recent model from the European Commission Joint Research Centre (European Commission. Joint Research Centre., 2023b) has also proposed gridded global net migration estimates in five-year intervals from 1975 to 2020 at a spatial resolution of 1 km, based on natural change across the rural-urban continuum with as an open dataset (Alessandrini et al., 2023). The authors also resort to indirect methods for estimating net migration pointing at limitations of data availability preventing more precise estimation of net migration at high spatial resolution. The estimates rely on strong assumptions, a common characteristic of other attempts, suggesting that estimates do have not the objective to produce statistics of mobility in absolute terms to be aggregated at higher geographical level, but on showing differences in spatial patterns of mobility in relative terms across different geographical areas and over relatively long intervals of five years. The study points at mobility patterns revealed by the estimates to be consistent with net migration data at sub-national level (European Commission. Joint Research Centre., 2020, 2023b).

Comparison of the EC JRC dataset with net migration figures from the Census of 2011 proves difficult, given the availability of raster datasets at 5-year intervals, from 1980 to 2020. The closest in the available datasets to the census year (2011) are the estimates for the year 2010. Testing the dataset, with the 5-year values averaged for one year period, comparing net migration estimates for the year 2010 with the census figures from 2011, resulted in a very inaccurate estimation, with MAE 144, RMSE 159, MPE -91,899 and MAPE 3381,601.

Of note, Fielding refers to Portugal as an exception country in Western Europe, alongside Switzerland, Austria and Ireland, where migration flow data can't be obtained as easily as in other countries, suggesting that migration flow data has brought significant improvement compared to birthplace statistics (Fielding, 2023).

The two studies (European Commission. Joint Research Centre., 2023b; Niva, Horton, Virkki, Heino, Kosonen, et al., 2023) seem to still present accuracy challenges in direct applicability for yearly net migration estimates in replacement of census data.

3.2.2. Proportion of the population that lived outside the parish 5 years before

A replacement estimate for the proportion of the population that lived outside the parish 5 years before (LiveOut) indicator used by Canadas et al. (2023) could not be retrieved in this study. While net migration provides the balance between incoming and outgoing residents, the LiveOut indicator could only be estimated based on different models and calculations, such as downscaling of previous incoming and outgoing migrations over a specific time interval. We saw in 3.2.1 that net migration estimates present significant limitations as direct replacement for census data. It is thus assumed that LiveOut indicator finds increased obstacles. However, if to measure attractiveness, studies on predictors of migration have gathered significant attention. For instance, X. Chen (2020) explored the potential of using nighttime lights data for predicting population migration in small areas across European Union (EU) countries, demonstrating that, in comparison to factors like population size and gross domestic product (GDP), nighttime lights data served as a valuable predictor. Nighttime lights showed a prediction power similar to that of population size and significantly stronger than GDP per capita (X. Chen, 2020). A different study of geospatial and sociodemographic predictors of human mobility suggested that socioeconomic variables including urbanicity, poverty, and female education strongly explained mobility patterns, in addition to geospatial covariates such as accessibility to major population centers and temperature (Ruktanonchai et al., 2021). In a way, each of the studies reviewed for this dissertation has measured the efficiency of different predictors of migration and mobility. While LiveOut could not be retrieved, it’s meaning as attractiveness may potentially find future solutions in a more systematic study of predictors.

3.3. Commuting Mobility Measurement

3.3.1. Proportion of population that works or studies in another parish in the same municipality.

The measurement of commuting mobility plays an important role in understanding human mobility patterns and their predictability. In a study by Song et al. (2010), the predictability of human mobility patterns was examined using anonymized mobile phone user data. Findings from measuring entropy in each individual trajectory indicated 93% potential predictability in user mobility (Song et al., 2010, p. 1018).

To situate commuting mobility within wider developments of mobility models, the taxonomy proposed by Luca et al. (2020) proved useful. The authors categorized mobility models into two main objectives: predictive and generative. Predictive models aim to forecast future mobility, either at the individual or collective level, encompassing next-location predictive models, which forecast an individual's future locations based on their historical mobility data, and crowd flow predictive models, which focus on predicting the flow of people in geographic regions.

Generative models, on the other hand, intend to create realistic mobility data either as trajectory generation or flow generation. Trajectory generation models, concern the generation of synthetic trajectories that can reproduce, realistically, individual statistical patterns of human mobility, while flow generation models, aim to generate realistic flows among locations, considering their characteristics and the distance among them, but without any knowledge about the real flows (Luca et al., 2020). Luca et al. exclusively reviewed deep learning models, and no other techniques, highlighting a limited literature available on flow generation models.

Estimates for proportion of the population that works or studies in another parish in the same municipality (WorkOutP), relate to flow generation models. Luca et al., who have exclusively reviewed deep learning models, and no other techniques, highlighted limited literature available on flow generation models (Luca et al., 2020).

Taking a closer look at them, one notable approach is Deep Gravity, proposed by Simini et al. (2022). Deep Gravity generates flows probabilities based on features that include land use, road networks, transportation, and the availability of services such as food and healthcare, using deep neural networks to identify non-linear relationships between those features and mobility flows. The model showed good geographic generalization capability, with significantly higher performance than traditional gravity models (based on Zipf, 1946), especially in densely populated regions (Simini et al., 2022). While commuting data is accessible in a 25km gridded format for England, Italy and New York State, and the authors have provided the Python code for Deep Gravity to open experiments, the different geographical coverage and the complexity of the code make it, for now, challenging to obtain WorkOutP values for the specific case studies through this method. The authors, however, conclude their study with relevant questions: "*Can we use rural areas flows to generate flows in cities? On the other hand, can we use cities' flows to generate flows in rural areas? And* *can we use a model trained on an entire country to generate flows on a different one?*" (Simini et al., 2022, p. 21).

Other studies examined in this review, such as those on geo-contextual embeddings for commuting flow prediction (Liu et al., 2020) and origin-destination flow generation with graph convolutional networks (Yao et al., 2021), offer useful understanding into model approaches and predictor datasets. However, those studies do not align geographically, datasets are not publicly available, or modeling effort is too high. Similarly, other models reviewed presented similar challenges. Some also relied on difficult to access data sources, such as mobile phone user data (M. Li et al., 2022). An alternative deep learning framework to generate synthetic mobility data based on travel surveys instead of relying on census data (Arkangil et al., 2023) presented interesting concepts and advanced deep learning models, while remaining unable to provide an alternative for WorkOutP.

The variable WorkOutP seems thus to be challenging to retrieve using current methods. However, research initiatives such as Deep Gravity are concerned with the concept of transferability, potentially pacing the way for future developments with wider geographical coverage.

3.4. Recreational Mobility Measurement

3.4.1. Night Stays

Recreational mobility measurements share some modeling similarities with commuting mobility, as they could potentially be retrieved from flow generation models. The most significant difference is in the availability of annual ancillary data, such as night stays but also tourist entertainment companies, accommodation capacity (number of beds/people) and river beaches and bathing areas, complementing other related census data such as proportion of seasonal or secondary dwellings. In a way, recreational mobility appears to be more easily predictable than the complexity of commuting mobility.

Furthermore, some studies have proposed models for mapping residential vacancies with multisource spatiotemporal data (X. Li & Gong, 2022) or for understanding seasonal movements (Ruktanonchai et al., 2021). However, since the number of night stays is retrievable on for early values from official sources, there seems to be no need for alternative estimates.

Nevertheless, findings from Y. Zhang et al (2022) regarding finely-grained spatiotemporal predictions of population distribution, where the time of day, housing-related variables, and precipitation types are important predictors, imply the importance of temporal resolution in recreational mobility. Additionally, research by Minora et al. (2023) on the nowcasting of tourist nights spent, has demonstrated the potential for even more frequent and timely statistics for recreational mobility. The authors have explored digital traces and web searches for nowcasting the number of monthly nights spent at sub-national scale across 11 European countries, while noting that privacy and surveillance concerns limit the utility of some of these tools. Portugal was not part of the 11 countries studied by Minora et al. (2023). Moreover, most models reviewed in this section relied on mobile phone data (e.g., Meredith et al., 2021; Minora et al., 2023; Ruktanonchai et al., 2021; Y. Zhang et al., 2022).

Finally, also for night stays no alternative measurement model was identified during the review. However, the necessity for alternative in this case is relatively minor, considering the regular release of official figures. The study of recreational and seasonal mobility seems to present faster development than commuting mobility.

4. Conclusions

This study critically examined the potential of emerging methods to quantify human mobility within rural landscapes, compared against official statistics. With the primary objective of assessing the possibility of retrieving migration and mobility variables from emerging methods instead of official statistics, it researched applications in measuring direct and indirect migration, commuting, and recreational mobility. The review process has revealed a landscape rich in innovation. The findings suggest that alternative approaches have been sought after and supported by technological advancements with modeling migration and mobility as an apparently a well-established field, particularly in urban contexts. The findings also suggest that despite presenting some useful alternatives, emerging models in these domains further underscore the complexity of human mobility.

Variation of Emerging Methods: The systematic literature review and applicability tests undertaken in this research underscore a growing precision in some of the emerging methods. For instance, gridded datasets of population and population density with global outputs, remained especially accurate when utilizing fine-resolution grids. Coarser grided resolution revealed inconsistencies at lower administrative levels for population density but maintained relatively small discrepancies in longer term calculations of population density variations. However, the review and testing also highlighted limitations and ongoing challenges with difficulties in retrieving practical alternatives for direct migration, commuting, and recreational mobility estimates. Methods proposing gridded net migration estimates resulted in a notable variance with comparable official data. In addition, commuting and recreational mobility measurements have been addressed by different models, but their availability remains limited to specific geographical regions or difficult transfer without significant coding and data acquisition effort.

Practical Implications for Rural Development: the review of emerging approaches on human mobility applicable to rural studies unveiled limitations related to geographical coverage, in this case of rural areas of Portugal. It also confirmed an urban focus in emerging models, where rural mobility is – at the most – a incoming component into the larger city. With of grids with global coverage, among reviewed methods there were limited examples with specified rural application or a rural focus.

Ethical Considerations and Data Privacy: alongside reviewing methods, this study has noted carefully on the ethical concerns conveyed by the reviewed studies. Although they point at some broad consensus on privacy preservation, discussions around informed consent, data anonymization, or dual-use nature of technology were absent in many cases.

Future Horizons in Rural Mobility Research: the review allows to anticipate an evolution of models that could capture a more varied spectrum of rural mobility. For instance, the publication of two novel global grids of net migration in the present year, despite their current limitations, consents the idea that soon could be possible to estimate net migration at very localized level. Such possibility could greatly improve population estimates in rural studies. Furthermore, models have improved based on real-time data, higher spatial resolution, and fast technological advances available in certain contexts. It is possible that some model solutions may transcend geographical boundaries, with transferability of training for the use of more common data sources. This is particularly relevant in terms of commuting and recreational mobility in rural environments.

5. References

Abreu, I., Nunes, J. M., & Mesias, F. J. (2019). Can Rural Development Be Measured? Design and Application of a Synthetic Index to Portuguese Municipalities. *Social Indicators Research*, *145*(3), 1107–1123. https://doi.org/10.1007/s11205-019-02124-w

Alessandretti, L., Aslak, U., & Lehmann, S. (2020). The Scales of Human Mobility. *Nature*, *587*(7834), 402–407. https://doi.org/10.1038/s41586-020-2909-1

Alessandrini, A., Deuster, C., & Natale, F. (2023). *Global estimates of net migration at high spatial resolution 1975-2020 [Dataset]* (JRC132876) [dataset]. Publications Office of the European Union. https://doi.org/10.2760/06672

Arkangil, E., Yildirimoglu, M., Kim, J., & Prato, C. (2023). A deep learning framework to generate synthetic mobility data. *2023 8th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 1–6. https://doi.org/10.1109/MT-ITS56129.2023.10241677

Bao, W., Gong, A., Zhang, T., Zhao, Y., Li, B., & Chen, S. (2023). Mapping Population Distribution with High Spatiotemporal Resolution in Beijing Using Baidu Heat Map Data. *Remote Sensing*, *15*(2), 458. https://doi.org/10.3390/rs15020458

Bao, W., Gong, A., Zhao, Y., Chen, S., Ba, W., & He, Y. (2022). High-Precision Population Spatialization in Metropolises Based on Ensemble Learning: A Case Study of Beijing, China. *Remote Sensing*, *14*(15), 3654. https://doi.org/10.3390/rs14153654

Baptista, F. O. (2010). A Transicao Rural. In *Espaço Rural: Declínio da Agricultura* (pp. 125–160). Celta Editora.

Barbosa-Filho, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., & Tomasini, M. (2018). Human Mobility: Models and Applications. *Physics Reports*, *734*, 1–74. https://doi.org/10.1016/j.physrep.2018.01.001

Bell, M. M., & Osti, G. (2010). Mobilities and Ruralities: An Introduction. *Sociologia Ruralis*, *50*(3), 199–204. https://doi.org/10.1111/j.1467-9523.2010.00518.x

Bell, M., & Muhidin, S. (2009). *Cross-National Comparison of Internal Migration* (Human Development Research Paper, p. 66) [2009/30]. United Nations Development Program.

Bosworth, G., Price, L., Collison, M., & Fox, C. (2020). Unequal futures of rural mobility: Challenges for a “Smart Countryside.” *Local Economy: The Journal of the Local Economy Policy Unit*, *35*(6), 586–608. https://doi.org/10.1177/0269094220968231

Brettell, C., & Hollifield, J. F. (Eds.). (2023). *Migration theory: Talking across disciplines* (Fourth edition). Routledge.

Canadas, M. J., Leal, M., Soares, F., Novais, A., Ribeiro, P. F., Schmidt, L., Delicado, A., Moreira, F., Bergonse, R., Oliveira, S., Madeira, P. M., & Santos, J. L. (2023). Wildfire mitigation and adaptation: Two locally independent actions supported by different policy domains. *Land Use Policy*, *124*, 106444. https://doi.org/10.1016/j.landusepol.2022.106444

Canadas, M. J., & Novais, A. (2019). Forest owners and fuels management coordination. When neighbours’ actions matter. *Scandinavian Journal of Forest Research*, *34*(1), 67–77. https://doi.org/10.1080/02827581.2018.1530369

Chen, H., Wu, B., Yu, B., Chen, Z., Wu, Q., Lian, T., Wang, C., Li, Q., & Wu, J. (2021). A New Method for Building-Level Population Estimation by Integrating LiDAR, Nighttime Light, and POI Data. *Journal of Remote Sensing*, *2021*, 2021/9803796. https://doi.org/10.34133/2021/9803796

Chen, X. (2020). Nighttime Lights and Population Migration: Revisiting Classic Demographic Perspectives with an Analysis of Recent European Data. *Remote Sensing*, *12*(1), 169. https://doi.org/10.3390/rs12010169

Cheng, Z., Wang, J., & Ge, Y. (2022). Mapping monthly population distribution and variation at 1-km resolution across China. *International Journal of Geographical Information Science*, *36*(6), 1166–1184. https://doi.org/10.1080/13658816.2020.1854767

Cresswell, T. (2006). *On the move: Mobility in the modern Western world*. Routledge.

Creswell, J. W. (1994). *Research design: Qualitative, quantitative and mixed method approaches*. Sage Publications.

Curtis, K. J., & Kulcsár, L. J. (2019). 22 Rural Demography. In D. L. Poston (Ed.), *Handbook of Population* (pp. 599–617). Springer International Publishing. https://doi.org/10.1007/978-3-030-10910-3_23

Cuttone, A., Lehmann, S., & González, M. C. (2018). Understanding predictability and exploration in human mobility. *EPJ Data Science*, *7*(1), 2. https://doi.org/10.1140/epjds/s13688-017-0129-1

DGT, D.-G. do T. (2023). *Carta Administrativa Oficial de Portugal—CAOP2022 (Continente)* (Version 2022) [dataset]. https://geo2.dgterritorio.gov.pt/caop/CAOP_Continente_2022-gpkg.zip

Doda, S., Wang, Y., Kahl, M., Hoffmann, E. J., Ouan, K., Taubenböck, H., & Zhu, X. X. (2022). So2Sat POP - A Curated Benchmark Data Set for Population Estimation from Space on a Continental Scale. *Scientific Data*, *9*(1), 715. https://doi.org/10.1038/s41597-022-01780-x

European Commission. Joint Research Centre. (2020). *Estimating net migration at high spatial resolution.* Publications Office. https://data.europa.eu/doi/10.2760/383386

European Commission. Joint Research Centre. (2022). *Data innovation in demography, migration and human mobility.* Publications Office. https://data.europa.eu/doi/10.2760/027157

European Commission. Joint Research Centre. (2023a). *GHSL data package 2023.* Publications Office. https://data.europa.eu/doi/10.2760/098587

European Commission. Joint Research Centre. (2023b). *Global estimates of net migration at high spatial resolution 1975-2020.* Publications Office. https://data.europa.eu/doi/10.2760/06672

European Commission. Statistical Office of the European Union. (2021). *Applying the degree of urbanisation: A methodological manual to define cities, towns and rural areas for international comparisons : 2021 edition.* Publications Office. https://data.europa.eu/doi/10.2785/706535

European Commission & United Nations. (2020). *International Recommendations on Internally Displaced Persons Statistics (IRIS): March 2020.* Publications Office. https://data.europa.eu/doi/10.2785/18809

Eusébio, C., Carneiro, M. J., Kastenholz, E., Figueiredo, E., & Soares Da Silva, D. (2017). Who is consuming the countryside? An activity-based segmentation analysis of the domestic rural tourism market in Portugal. *Journal of Hospitality and Tourism Management*, *31*, 197–210. https://doi.org/10.1016/j.jhtm.2016.12.006

Fielding, A. J. (2023). Internal Migration in Western Europe. In L. A. Kosiński & R. M. Prothero (Eds.), *PEOPLE ON THE MOVE studies on internal migration*. ROUTLEDGE.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3518482

Floridi, L., & Cowls, J. (2021). A Unified Framework of Five Principles for AI in Society. In L. Floridi (Ed.), *Ethics, Governance, and Policies in Artificial Intelligence* (Vol. 144, pp. 5–17). Springer International Publishing. https://doi.org/10.1007/978-3-030-81907-1_2

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Greenwood, M. J. (1985). HUMAN MIGRATION: THEORY, MODELS, AND EMPIRICAL STUDIES*. *Journal of Regional Science*, *25*(4), 521–544. https://doi.org/10.1111/j.1467-9787.1985.tb00321.x

Greenwood, M. J. (1997). Chapter 12 Internal migration in developed countries. In *Handbook of Population and Family Economics* (Vol. 1, pp. 647–720). Elsevier. https://doi.org/10.1016/S1574-003X(97)80004-9

Hasan, S., Schneider, C. M., Ukkusuri, S. V., & González, M. C. (2013). Spatiotemporal Patterns of Urban Human Mobility. *Journal of Statistical Physics*, *151*(1), 304–318. https://doi.org/10.1007/s10955-012-0645-0

INE, P. (2002). *Censos 2001. XIV Recenseamento Geral da População: IV Recenseamento Geral da Habitação. Resultados definitivos: Portugal.* [dataset]. Instituto Nacional de Estatística. https://www.ine.pt/xurl/pub/133411

INE, P. (2012a). *Censos 2011. 8: Alentejo*.

INE, P. (2012b). *Censos 2011. XV Recenseamento Geral da População: V Recenseamento Geral da Habitação. Resultados definitivos: Portugal.* [dataset]. Instituto Nacional de Estatística. https://www.ine.pt/xurl/pub/73212469

IOM. (2019). *Workshop Report on Forecasting Human Mobility in Contexts of Crises*. International Organization for Migration. https://displacement.iom.int/reports/workshop-report-forecasting-human-mobility-contexts-crises?close=true

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kälin, W. (2022). Locating International Law on Human Mobility in the Context of Climate Change. *Proceedings of the ASIL Annual Meeting*, *116*, 160–162. https://doi.org/10.1017/amp.2023.50

Klous, G., Kretzschmar, M. E. E., Coutinho, R. A., Heederik, D. J. J., & Huss, A. (2020). Prediction of human active mobility in rural areas: Development and validity tests of three different approaches. *Journal of Exposure Science & Environmental Epidemiology*, *30*(6), 1023–1031. https://doi.org/10.1038/s41370-019-0194-6

Kraemer, M. U. G., Sadilek, A., Zhang, Q., Marchal, N. A., Tuli, G., Cohn, E. L., Hswen, Y., Perkins, T. A., Smith, D. L., Reiner, R. C., & Brownstein, J. S. (2020). Mapping global variation in human mobility. *Nature Human Behaviour*, *4*(8), 800–810. https://doi.org/10.1038/s41562-020-0875-0

Kuffer, M., Owusu, M., Oliveira, L., Sliuzas, R., & Van Rijn, F. (2022). The Missing Millions in Maps: Exploring Causes of Uncertainties in Global Gridded Population Datasets. *ISPRS International Journal of Geo-Information*, *11*(7), 403. https://doi.org/10.3390/ijgi11070403

Leyk, S., Gaughan, A. E., Adamo, S. B., De Sherbinin, A., Balk, D., Freire, S., Rose, A., Stevens, F. R., Blankespoor, B., Frye, C., Comenetz, J., Sorichetta, A., MacManus, K., Pistolesi, L., Levy, M., Tatem, A. J., & Pesaresi, M. (2019). The spatial allocation of population: A review of large-scale gridded population data products and their fitness for use. *Earth System Science Data*, *11*(3), 1385–1409. https://doi.org/10.5194/essd-11-1385-2019

Li, C., & Managi, S. (2023). Gridded Datasets for Japan: Total, Male, and Female Populations from 2001–2020. *Scientific Data*, *10*(1), 81. https://doi.org/10.1038/s41597-023-01989-4

Li, M., Gao, S., Qiu, P., Tu, W., Lu, F., Zhao, T., & Li, Q. (2022). Fine-grained crowd distribution forecasting with multi-order spatial interactions using mobile phone data. In *Transportation Research Part …*. Elsevier. https://doi.org/10.1016/j.trc.2022.103908

Li, X., & Gong, P. (2022). Mapping Residential Vacancies with Multisource Spatiotemporal Data: A Case Study in Beijing. *Remote Sensing*, *14*(2), 376. https://doi.org/10.3390/rs14020376

Li, X., Xu, H., Huang, X., Guo, C. (Atlas), Kang, Y., & Ye, X. (2021). Emerging geo-data sources to reveal human mobility dynamics during COVID-19 pandemic: Opportunities and challenges. *Computational Urban Science*, *1*(1), 22. https://doi.org/10.1007/s43762-021-00022-x

Liu, Z., Miranda, F., Xiong, W., Yang, J., Wang, Q., & Silva, C. (2020). Learning Geo-Contextual Embeddings for Commuting Flow Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(01), 808–816. https://doi.org/10.1609/aaai.v34i01.5425

Lloyd, C. T., Sorichetta, A., & Tatem, A. J. (2017). High resolution global gridded data for use in population studies. *Scientific Data*, *4*(1), 170001. https://doi.org/10.1038/sdata.2017.1

López-Sanz, J. M., Penelas-Leguía, A., Gutiérrez-Rodríguez, P., & Cuesta-Valiño, P. (2021). Sustainable Development and Rural Tourism in Depopulated Areas. *Land*, *10*(9), 985. https://doi.org/10.3390/land10090985

Luca, M., Barlacchi, G., Lepri, B., & Pappalardo, L. (2020). A Survey on Deep Learning for Human Mobility. *ACM Computing Surveys*, *55*(1), 1–44. https://doi.org/10.1145/3485125

Lucas, R. E. B. (2007). Migration and rural development. *Journal of Agricultural and Development Economics*, *4*(1), 99/122.

McCaffrey, S., Wilson, R., & Konar, A. (2018). Should I Stay or Should I Go Now? Or Should I Wait and See? Influences on Wildfire Evacuation Decisions. *Risk Analysis*, *38*(7), 1390–1404. https://doi.org/10.1111/risa.12944

McFalls, J. A. (2007). *Population: A Lively Introduction*. *62*(1), 36.

McWethy, D. B., Schoennagel, T., Higuera, P. E., Krawchuk, M., Harvey, B. J., Metcalf, E. C., Schultz, C., Miller, C., Metcalf, A. L., Buma, B., Virapongse, A., Kulig, J. C., Stedman, R. C., Ratajczak, Z., Nelson, C. R., & Kolden, C. (2019). Rethinking resilience to wildfire. *Nature Sustainability*, *2*(9), 797–804. https://doi.org/10.1038/s41893-019-0353-8

Meredith, H. R., Giles, J. R., Perez-Saez, J., Mande, T., Rinaldo, A., Mutembo, S., Kabalo, E. N., Makungo, K., Buckee, C. O., Tatem, A. J., Metcalf, C. J. E., & Wesolowski, A. (2021). Characterizing human mobility patterns in rural settings of sub-Saharan Africa. *eLife*, *10*, e68441. https://doi.org/10.7554/eLife.68441

Minora, U., Iacus, S. M., Silva, F. B. E., Sermi, F., & Spyratos, S. (2023). Nowcasting tourist nights spent using innovative human mobility data. In *PloS one* (Vol. 18, Issue 10). https://doi.org/10.1371/journal.pone.0287063

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 205395171667967. https://doi.org/10.1177/2053951716679679

Moreno-Monroy, A. I., Schiavina, M., & Veneri, P. (2021). Metropolitan areas in the world. Delineation and population trends. *Journal of Urban Economics*, *125*, 103242. https://doi.org/10.1016/j.jue.2020.103242

Neal, I., Seth, S., Watmough, G., & Diallo, M. S. (2022). Census-independent population estimation using representation learning. *Scientific Reports*, *12*(1), 5185. https://doi.org/10.1038/s41598-022-08935-1

Niva, V., Horton, A., Virkki, V., Heino, M., Kallio, M., Kinnunen, P., Abel, G. J., Muttarak, R., Taka, M., Varis, O., & Kummu, M. (2023). *Data for: World's human migration patterns in 2000-2019 unveiled by high-resolution data [Data set]* [dataset]. https://zenodo.org/records/7997134. https://doi.org/10.5281/zenodo.7997133

Niva, V., Horton, A., Virkki, V., Heino, M., Kosonen, M., Kallio, M., Kinnunen, P., Abel, G. J., Muttarak, R., Taka, M., Varis, O., & Kummu, M. (2023). World's human migration patterns in 2000–2019 unveiled by high-resolution data. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-023-01689-4

Nori, M., & Farinella, D. (2020). Mobility and Migrations in the Rural Areas of Mediterranean EU Countries. In M. Nori & D. Farinella, *Migration, Agriculture and Rural Development* (pp. 39–67). Springer International Publishing. https://doi.org/10.1007/978-3-030-42863-1_3

Oliveira, S., Zêzere, J. L., Queirós, M., & Pereira, J. M. (2017). Assessing the social context of wildfire-affected areas. The case of mainland Portugal. *Applied Geography*, *88*, 104–117. https://doi.org/10.1016/j.apgeog.2017.09.004

Rey Benayas, J. M., Martins, A., Nicolau, J. M., & Schulz, J. J. (2007). Abandonment of agricultural land: An overview of drivers and consequences. *CABI Reviews*, *2007*. https://doi.org/10.1079/PAVSNNR20072057

Ruktanonchai, C. W., Lai, S., Utazi, C. E., Cunningham, A. D., Koper, P., Rogers, G. E., Ruktanonchai, N. W., Sadilek, A., Woods, D., Tatem, A. J., Steele, J. E., & Sorichetta, A. (2021). Practical geospatial and sociodemographic predictors of human mobility. *Scientific Reports*, *11*(1), 15389. https://doi.org/10.1038/s41598-021-94683-7

Rye, J. F., & O'Reilly, K. (2021). *International labour migration to Europe's rural regions*. Taylor & Francis.

Salah, A. A., Canca, C., & Erman, B. (2022). *Ethical and legal concerns on data science for large scale human mobility*. 26.

Santos, R. (2023). Return Migration and Rural Tourism Development in Portugal. *Tourism Planning & Development*, *20*(4), 636–659. https://doi.org/10.1080/21568316.2021.1953121

Schwerdtle, P. N., Stockemer, J., Bowen, K. J., Sauerborn, R., McMichael, C., & Danquah, I. (2020). A Meta-Synthesis of Policy Recommendations Regarding Human Mobility in the Context of Climate Change. *International Journal of Environmental Research and Public Health*, *17*(24), 9342. https://doi.org/10.3390/ijerph17249342

Simini, F., Barlacchi, G., Luca, M., & Pappalardo, L. (2022). *Deep Gravity: Enhancing mobility flows generation with deep neural networks and geographic information* (arXiv:2012.00489). arXiv. http://arxiv.org/abs/2012.00489

Smith, D. P. (1992). Migration in Population Analysis. In K. C. Land (Ed.), *Formal Demography* (pp. 283–301). Springer US. https://doi.org/10.1007/978-1-4899-2465-0_9

Solmaz, G., & Turgut, D. (2019). A Survey of Human Mobility Models. *IEEE Access*, *7*, 125711–125731. https://doi.org/10.1109/ACCESS.2019.2939203

Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of Predictability in Human Mobility. *Science*, *327*(5968), 1018–1021. https://doi.org/10.1126/science.1177170

Sorichetta, A., Hornby, G. M., Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Scientific Data*, *2*(1), 150045. https://doi.org/10.1038/sdata.2015.45

Syssner, J. (2020). Policy Implications of Rural Depopulation. In J. Syssner, *Pathways to Demographic Adaptation* (pp. 37–52). Springer International Publishing. https://doi.org/10.1007/978-3-030-34046-9_4

Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data*, *4*(1), 170004. https://doi.org/10.1038/sdata.2017.4

Teng, F., Wang, Y., Wang, M., & Wang, L. (2022). Monitoring and Analysis of Population Distribution in China from 2000 to 2020 Based on Remote Sensing Data. *Remote Sensing*, *14*(23), 6019. https://doi.org/10.3390/rs14236019

Thomson, D. R., Leasure, D. R., Bird, T., Tzavidis, N., & Tatem, A. J. (2022). How accurate are WorldPop-Global-Unconstrained gridded population data at the cell-level?: A simulation analysis in urban Namibia. *PLOS ONE*, *17*(7), e0271504. https://doi.org/10.1371/journal.pone.0271504

Toch, E., Lerner, B., Ben-Zion, E., & Ben-Gal, I. (2019). Analyzing large-scale human mobility data: A survey of machine learning methods and applications. *Knowledge and Information Systems*, *58*(3), 501–523. https://doi.org/10.1007/s10115-018-1186-x

Wang, L., Fan, H., & Wang, Y. (2020). Improving population mapping using Luojia 1-01 nighttime light image and location-based social media data. *Science of The Total Environment*, *730*, 139148. https://doi.org/10.1016/j.scitotenv.2020.139148

Wang, M., Wang, Y., Li, B., Cai, Z., & Kang, M. (2022). A Population Spatialization Model at the Building Scale Using Random Forest. *Remote Sensing*, *14*(8), 1811. https://doi.org/10.3390/rs14081811

Wesolowski, A., Buckee, C. O., Pindolia, D. K., Eagle, N., Smith, D. L., Garcia, A. J., & Tatem, A. J. (2013). The Use of Census Migration Data to Approximate Human Movement Patterns across Temporal Scales. *PLoS ONE*, *8*(1), e52971. https://doi.org/10.1371/journal.pone.0052971

WorldPop. (2018). *Portugal, 100m. Population. Unconstrained individual countries 2000-2020 (100m resolution)* [dataset]. School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). https://dx.doi.org/10.5258/SOTON/WP00645

WorldPop. (2020). *Portugal, 1000m. Population density. Unconstrained individual countries 2000-2020 (1km resolution)* [dataset]. School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia

University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). https://dx.doi.org/10.5258/SOTON/WP00645

Yao, X., Gao, Y., Zhu, D., Manley, E., Wang, J., & Liu, Y. (2021). Spatial Origin-Destination Flow Imputation Using Graph Convolutional Networks. *IEEE Transactions on Intelligent Transportation Systems*, *22*(12), 7474–7484. https://doi.org/10.1109/TITS.2020.3003310

Zhang, J. L., & Bryant, J. (2020). Bayesian Disaggregated Forecasts: Internal Migration in Iceland. In S. Mazzuco & N. Keilman (Eds.), *Developments in Demographic Forecasting* (Vol. 49). Springer International Publishing. https://doi.org/10.1007/978-3-030-42472-5

Zhang, Y., Zhang, Y., Huang, B., & Liu, X. (2022). A hybrid model for high spatial and temporal resolution population distribution prediction. *International Journal of Digital …*. https://doi.org/10.1080/17538947.2022.2155718

Zipf, G. K. (1946). The P 1 P 2 D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review*, *11*(6), 677. https://doi.org/10.2307/2087063

Annex 1 – Variables included in the rural development index proposed by Abreu et al. (2019)

	Abbreviation	Weight
Population		
Population density (inhab/km^2)—the intensity of population settlement expressed as the ratio between total population and surface area. A more developed territory is more attractive. The higher is its value, the lower the isolation of territories, therefore increasing the RDI value	PopDens	0.25
Rate of natural increase (%)—the difference between the number of live births and the number of deaths occurring during a given period, usually a calendar year divided by the mid-year population of that period	NatInc	0.25
Net migration (No.)—the difference between immigration into and emigration from the country or region during a given period, reflects the attractiveness capacity of each territory. It is assumed that a higher development will lead to a higher capacity on attracting populations	NetMig	0.25
Demographic Dependency Index (%)—the ratio of the elderly (ages 65 and older) plus the young (under age 15) to the population in the working ages (ages 15–64). Changes in the dependency ratio provide an indication of the potential social support requirements resulting from changes in population age structures	DmgDep	0.25
Social		
Literacy (%)—being education one of the most undisputed variables of development, the literacy rate is its base. Although we are in the twenty-first century, illiteracy rates are still significant in some rural areas	Lit	0.25
Proportion of resident population with at least the lower secondary education 3rd. cycle completed (%)—proportion of resident population with 15 and more years old with at least the lower secondary education 3rd. cycle completed. While literacy can be seen as a condition for minimum integration in society, compulsory education can be seen as a minimum condition to ensure employability. Not only a more developed society has the means for its youth to follow compulsory education, but higher levels of education also lead to higher productivity and living conditions	Educ	0.25
Physicians (No.) per 1000 inhabitants—this variable introduces the dimension of basic health conditions in the territory under study	Phys	0.25
Proportion of conventional dwellings of usual residence (%) with installation existence (electricity, water, toilet, bath/shower and heating)—to include dwellings quality in rural areas	Instal	0.25
Economy		
Proportion of family agricultural population with remunerated activity outside agricultural holding (%)—introducing the concept of activities diversification in the agricultural environment	OthRem	0.25
Average monthly earnings (€)—assessing the differences in labour remuneration between the various municipalities of the territory under study	Earn	0.25

Population		
Population density (inhab/km^2)—the intensity of population settlement expressed as the ratio between total population and surface area. A more developed territory is more attractive. The higher is its value, the lower the isolation of territories, therefore increasing the RDI value	PopDens	0.25
Rate of natural increase (%)—the difference between the number of live births and the number of deaths occurring during a given period, usually a calendar year divided by the mid-year population of that period	NatInc	0.25
Net migration (No.)—the difference between immigration into and emigration from the country or region during a given period, reflects the attractiveness capacity of each territory. It is assumed that a higher development will lead to a higher capacity on attracting populations	NetMig	0.25
Demographic Dependency Index (%)—the ratio of the elderly (ages 65 and older) plus the young (under age 15) to the population in the working ages (ages 15–64). Changes in the dependency ratio provide an indication of the potential social support requirements resulting from changes in population age structures	DmgDep	0.25
Social		
Literacy (%)—being education one of the most undisputed variables of development, the literacy rate is its base. Although we are in the twenty-first century, illiteracy rates are still significant in some rural areas	Lit	0.25
Proportion of resident population with at least the lower secondary education 3rd. cycle completed (%)—proportion of resident population with 15 and more years old with at least the lower secondary education 3rd. cycle completed. While literacy can be seen as a condition for minimum integration in society, compulsory education can be seen as a minimum condition to ensure employability. Not only a more developed society has the means for its youth to follow compulsory education, but higher levels of education also lead to higher productivity and living conditions	Educ	0.25
Physicians (No.) per 1000 inhabitants—this variable introduces the dimension of basic health conditions in the territory under study	Phys	0.25
Proportion of conventional dwellings of usual residence (%) with installation existence (electricity, water, toilet, bath/shower and heating)—to include dwellings quality in rural areas	Instal	0.25
Economy		
Proportion of family agricultural population with remunerated activity outside agricultural holding (%)—introducing the concept of activities diversification in the agricultural environment	OthRem	0.25
Average monthly earnings (€)—assessing the differences in labour remuneration between the various municipalities of the territory under study	Earn	0.25

(adapted from Abreu et al. 2019)

Annex 2 - Variables included in the study conducted by Canadas et al. (2013)

Acronym	Variable name	Year/ period	Unit	Source	Average	Max - min	Std. deviat.
DIMENSION 1 – Land use/land cover (LULC)							
Built_up	Proportion of built-up areas	2015	%	(1)	3.6	16.3 – 0.4	2.9
Agricult	Proportion of agricultural areas	2015	%	(1)	14.9	40.6 – 1.6	9.3
Forest	Proportion of forest/wildland areas	2015	%	(1)	79.9	97.8 – 48.3	11.6
AgricBuff	Farming in a 100 m buffer around built-up areas	2015	%	(1)	54.1	84.0 – 22.2	12.3
ForBuff	Forest/wildland in a 100 m buffer around built-up areas	2015	%	(1)	44.9	76.9 – 16.0	12.2
Euc	Proportion of eucalyptus	2015	%	(1)	21.5	67.0 – 0.4	16.2
EucVar	Variation of eucalyptus area	1995–2015	%	(1)	6.9	39.1 – -6.7	7.7
Pin	Proportion of maritime pine	2015	%	(1)	38.4	72.9 – 1.9	16.7
PinVar	Variation of maritime pine area	1995–2015	Nr.	(1)	-8.2	16.8 – -50.5	12.3
Shannon	Shannon index	2015	%	(1)	1.1	1.5 – 0.6	0.2
EucProd	Net primary productivity of eucalyptus	2000–2014	Kg of carbon /m^2/ year	(1)	1.1	1.4 – 0.8	0.1
PinProd	Net primary productivity of maritime pine	2000–2014	Kg of carbon /m^2/ year	(1)	1.0	1.3 – 0.7	0.1
LivStock	Livestock units (sheep, goats, and cattle)	2009	Nr/km^2	(2)	2.2	11.2 – 0.0	2.1
Traction	Mechanical traction/horse-power availability	2009	Nr/ha of rural area	(2)	1.3	4.3 – 0.0	1.0
DIMENSION 2 – Population							
PopDens	Population density	2011	Nr/km^2	(3)	46.8	255.3 – 4.0	43.2
RPopVar	Variation of rural population density	1981–2011	%	(3)	-33.5	41.9 – -68.6	19.1
Young	Proportion of young population (< 15 years)	2011	%	(3)	9.9	17.4 – 1.9	3.4
Elderly	Proportion of elderly population (≥ 65 years)	2011	%	(3)	34.3	58.8 – 17.1	10.0
Aging	Aging index	2011	%	(3)	455.3	2250.0 – 105.3	374.1
EldAlone	Proportion of elderly living alone or with others in the same age group	2011	%	(3)	23.6	49.6 – 10.0	8.7
Illitera	Illiteracy rate	2011	%	(3)	11.3	27.6 – 2.3	5.0
Educat	Proportion of the population with secondary or higher education	2011	%	(3)	17.9	35.0 – 3.9	5.7
PrimSect	Proportion of the population working in the primary sector	2011	%	(3)	5.8	36.4 – 0.0	5.5
LiveOutP	Proportion of the population that lived outside the parish 5 years ago	2011	%	(3)	9.9	35.3 – 1.3	4.3
WorkOutP	Proportion of the population that works or studies in another parish in the same municipality	2011	%	(3)	16.2	16.2 37.8 – 0.6	8.4
Seasonal	Proportion of seasonal/secondary dwellings	2011	%	(3)	38.6	72.9 – 9.5	14.5
BuildAge	Average age of buildings	2011	Nr. years	(3)	43.3	66.1 – 20.5	8.1
FamLabor	Proportion of farm family labour - annual work unit (AWU)	2009	%	(2)	93.8	99.9 – 50.1	7.3
TourisAg	Tourist entertainment companies	2021	Nr/100 km^2	(4)	1.5	12.4 – 0.0	2.7
AccomCap	Accommodation capacity (number of beds/people)	2021	Nr/1000 inhabitants	(4)	67.8	438.2 – 0.0	80.8
NightSty	Night stays	2020	Nr/km^2	(4)	70.4	1170.9 – 0.0	136.4
RivBeach	River beaches and bathing areas	2020	Nr/1000 inhabitants	(5)	0.9	16.9 – 0.0	2.1
DIMENSION 3 - Institutions							
AvgUAA	Average utilised agricultural area (UAA) - farms	2009	ha	(2)	2.0	6.1 – 0.7	0.8
Commons	Existence of commons	2015–2020	Binary	(6)	0.6	1 – 0	0.5
PulpPape	Existence of areas managed by pulp paper companies	2015–2020	Binary	(6)	0.4	1 – 0	0.5
MunSeat	Parish is coincident or not with the municipality seat	2021	Binary	n.a.	0.1	1 – 0	0.4
HistVill	Number of classified traditional (historical and schist) villages	2021	Nr.	(7)	0.2	5.0 – 0.0	0.7
LocAssoc	Number of local associations with wildfire-related initiatives	2015–2021	Nr.	(6) (8)	0.4	4.0 – 0.0	0.7
LandReg	Existence of simplified land registry	2015–2020	Binary	(6)	0.4	1 – 0	0.5
DIMENSION 4 – External resources							
PubInspe	Number of years that the parish was a priority for inspection of mandatory fuel management around villages	2018–2020	Nr.	(9)	1.5	3.0 – 0.0	1.1
PrioVuln	Combined priority-vulnerability levels	2020	Nr.	(9) (10)	1.6	2.0 – 0.0	0.6
SecFores	Parish included in the "Secure Forests" programme	2015–2020	Binary	(6)	0.6	1 – 0	0.5
SafeVill	Proportion of villages in the parish included in the "Safe Village, Safe People" programme	2017–2020	%	(11)	3.6	100.0 – 0.0	12.0
RDP2020	Value assigned under the Rural Development Programme (RDP2020)	2015–2020	€/km^2	(12)	4842.44	22449 – 0	4742.62
RDPMunic	Value assigned to the municipalities under the RDP2020	2015–2020	€/km^2	(12)	1809.97	10648 – 0	2615.56
RDPAssoc	Value assigned to wildfire-related associations under the RDP2020	2015–2020	€/km^2	(12)	1579.57	17957 – 0	3468.19
ForFund	Value assigned under the Permanent Forest Fund	2013–2019	€/ha of forest area	(13)	31.91	245.50 – 9.60	27.65
POSEUR	Value assigned under the Programme Sustainability and Efficiency in the Use of Resources (POSEUR)	2014–2020	€/km^2	(14)	1362.45	7648.62 – 118.58	1214.70
ForSappr	Existence of forest sappers' teams	2015–2020	Binary	(6)	0.2	1 – 0	0.4
DIMENSION 5 – Wildfires							
Fire0316	Accumulated burnt area during the 2003–2016 period (historic burnt area)	2003–2016	%	(13)	31.7	128.2 – 0.0	28.1
Fire1719	Accumulated burnt area during the 2017–2019 period (recent burnt area)	2017–2019	%	(13)	53.0	100.0 – 0.0	37.8

Annex 3 – Reviewed Studies

Study Variables

1. Abreu, I., Nunes, J. M., & Mesias, F. J. (2019). Measuring Rural Development: Design and Application of a Synthetic Index in Portuguese Municipalities. Soc. Indic. Res., 145(3), 1107–1123. [DOI: 10.1007/s11205-019-02124-w]
2. Canadas, M. J., Leal, M., Soares, F., Novais, A., Ribeiro, P. F., Schmidt, L., ... Santos, J. L. (2023). Wildfire Mitigation and Adaptation: Actions in Different Policy Domains. Land Use Policy, 124, 106444. [DOI: 10.1016/j.landusepol.2022.106444]

Population Density

1. Leyk, S., Gaughan, A. E., Adamo, S. B., et al. (2019). Review of Large-scale Gridded Population Data. Earth Syst. Sci. Data, 11(3), 1385–1409. [DOI: 10.5194/essd-11-1385-2019]
2. Lloyd, C. T., Sorichetta, A., & Tatem, A. J. (2017). High Resolution Global Gridded Population Data. Sci. Data, 4(1), 170001. [DOI: 10.1038/sdata.2017.1]
3. Tatem, A. J. (2017). WorldPop, Open Data for Spatial Demography. Sci. Data, 4(1), 170004. [DOI: 10.1038/sdata.2017.4]
4. WorldPop. (2018). Portugal, 100m. Population. Unconstrained Individual Countries 2000-2020 (100m resolution) [dataset]. [URL: https://dx.doi.org/10.5258/SOTON/WP00645]
5. WorldPop. (2020). Portugal, 1000m. Population Density. Unconstrained Individual Countries 2000-2020 (1km resolution) [dataset]. [URL: https://dx.doi.org/10.5258/SOTON/WP00645]
6. Bao, W., Gong, A., Zhao, Y., et al. (2022). High-Precision Population Spatialization in Metropolises Based on Ensemble Learning: A Case Study of Beijing, China. Remote Sens., 14(15), 3654. [DOI: 10.3390/rs14153654]
7. Chen, H., Wu, B., Yu, B., et al. (2021). Building-Level Population Estimation Using LiDAR, Nighttime Light, and POI Data. J. Remote Sens., 2021, 2021/9803796. [DOI: 10.34133/2021/9803796]
8. Doda, S., Wang, Y., Kahl, M., et al. (2022). So2Sat POP: A Benchmark Data Set for Population Estimation from Space. Sci. Data, 9(1), 715. [DOI: 10.1038/s41597-022-01780-x]
9. Kuffer, M., Owusu, M., Oliveira, L., et al. (2022). Causes of Uncertainties in Global Gridded Population Datasets. ISPRS Int. J. Geo-Inf., 11(7), 403. [DOI: 10.3390/ijgi11070403]
10. Li, C., & Managi, S. (2023). Gridded Datasets for Japan: 2001–2020. Sci. Data, 10(1), 81. [DOI: 10.1038/s41597-023-01989-4]

11. Neal, I., Seth, S., Watmough, G., et al. (2022). Census-Independent Population Estimation. Sci. Rep., 12(1), 5185. [DOI: 10.1038/s41598-022-08935-1]
12. Thomson, D. R., Leasure, D. R., Bird, T., et al. (2022). Accuracy of WorldPop-Global-Unconstrained gridded population data. PLoS ONE, 17(7), e0271504. [DOI: 10.1371/journal.pone.0271504]
13. Wang, L., Fan, H., & Wang, Y. (2020). Improved Population Mapping Using Nighttime Light Image and Social Media Data. Sci. Total Environ., 730, 139148. [DOI: 10.1016/j.scitotenv.2020.139148]
14. Wang, M., Wang, Y., Li, B., et al. (2022). Population Spatialization Model at Building Scale Using Random Forest. Remote Sens., 14(8), 1811. [DOI: 10.3390/rs14081811]

Variation of Rural Population Density

15. Bao, W., Gong, A., Zhang, T., et al. (2023). Mapping Population Distribution in Beijing Using Baidu Heat Map Data. Remote Sens., 15(2), 458. [DOI: 10.3390/rs15020458]
16. Cheng, Z., Wang, J., & Ge, Y. (2022). Mapping Monthly Population Distribution in China. Int. J. Geogr. Inf. Sci., 36(6), 1166–1184. [DOI: 10.1080/13658816.2020.1854767]
17. European Commission. Joint Research Centre. (2023). GHSL Data Package. Publications Office. [URL: https://data.europa.eu/doi/10.2760/098587]
18. European Commission. Statistical Office of the European Union. (2021). Applying the Degree of Urbanisation. Publications Office. [URL: https://data.europa.eu/doi/10.2785/706535]
19. Ruktanonchai, C. W., Lai, S., Utazi, C. E., et al. (2021). Predictors of Human Mobility. Sci. Rep., 11(1), 15389. [DOI: 10.1038/s41598-021-94683-7]

Net Migration

20. Chen, X. (2020). Nighttime Lights and Population Migration: Revisiting Classic Demographic Perspectives. Remote Sens., 12(1), 169. [DOI: 10.3390/rs12010169]

6. European Commission. Joint Research Centre. (2023). Global Estimates of Net Migration at High Spatial Resolution 1975-2020. Publications Office. [DOI: 10.2760/06672]
7. Fielding, A. J. (2023). Internal Migration in Western Europe. In L. A. Kosiński & R. M. Prothero (Eds.), People on the Move: Studies on Internal Migration. Routledge.
8. Niva, V., Horton, A., Virkki, V., et al. (2023). World's Human Migration Patterns in 2000–2019 Unveiled by High-Resolution Data [Dataset]. [DOI: 10.5281/zenodo.7997133]

9. Ruktanonchai, C. W., Lai, S., Utazi, C. E., et al. (2021). Geospatial and Sociodemographic Predictors of Human Mobility. Sci. Rep., 11(1), 15389. [DOI: 10.1038/s41598-021-94683-7]

Commuting mobility

25. Arkangil, E., Yildirimoglu, M., Kim, J., & Prato, C. (2023). A Deep Learning Framework to Generate Synthetic Mobility Data. 2023 8th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), 1–6. [DOI: 10.1109/MT-ITS56129.2023.10241677]
26. Li, M., Gao, S., Qiu, P., et al. (2022). Fine-Grained Crowd Distribution Forecasting with Multi-Order Spatial Interactions Using Mobile Phone Data. In Transportation Research Part .... Elsevier. [DOI: 10.1016/j.trc.2022.103908]
27. Liu, Z., Miranda, F., Xiong, W., et al. (2020). Learning Geo-Contextual Embeddings for Commuting Flow Prediction. Proc. AAAI Conf. Artif. Intell., 34(01), 808–816. [DOI: 10.1609/aaai.v34i01.5425]
28. Luca, M., Barlacchi, G., Lepri, B., & Pappalardo, L. (2020). A Survey on Deep Learning for Human Mobility. ACM Comput. Surv., 55(1), 1–44. [DOI: 10.1145/3485125]
29. Simini, F., Barlacchi, G., Luca, M., & Pappalardo, L. (2022). Deep Gravity: Enhancing Mobility Flows Generation with Deep Neural Networks and Geographic Information. arXiv:2012.00489. [arXiv: http://arxiv.org/abs/2012.00489]
30. Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of Predictability in Human Mobility. Science, 327(5968), 1018–1021. [DOI: 10.1126/science.1177170]
31. Yao, X., Gao, Y., Zhu, D., et al. (2021). Spatial Origin-Destination Flow Imputation Using Graph Convolutional Networks. IEEE Trans. Intell. Transp. Syst., 22(12), 7474–7484. [DOI: 10.1109/TITS.2020.3003310]
32. Zipf, G. K. (1946). The P 1 P 2 D Hypothesis: On the Intercity Movement of Persons. Am. Sociol. Rev., 11(6), 677. [DOI: 10.2307/2087063]

Recreational mobility

33. Li, X., & Gong, P. (2022). Mapping Residential Vacancies with Multisource Spatiotemporal Data: A Case Study in Beijing. Remote Sensing, 14(2), 376. [DOI: 10.3390/rs14020376]
34. Meredith, H. R., Giles, J. R., Perez-Saez, J., et al. (2021). Characterizing Human Mobility Patterns in Rural Settings of Sub-Saharan Africa. eLife, 10, e68441. [DOI: 10.7554/eLife.68441]
35. Minora, U., Iacus, S. M., Silva, F. B. E., et al. (2023). Nowcasting Tourist Nights Spent Using Innovative Human Mobility Data. In PloS One, 18(10). [DOI: 10.1371/journal.pone.0287063]

36. Ruktanonchai, C. W., Lai, S., Utazi, C. E., et al. (2021). Practical Geospatial and Sociodemographic Predictors of Human Mobility. Scientific Reports, 11(1), 15389. [DOI: 10.1038/s41598-021-94683-7]
37. Zhang, Y., Zhang, Y., Huang, B., & Liu, X. (2022). A Hybrid Model for High Spatial and Temporal Resolution Population Distribution Prediction. Int. J. Digital [DOI: 10.1080/17538947.2022.2155718]

Annex 4 – Datasets

Alessandrini, A., Deuster, C., & Natale, F. (2023). *Global estimates of net migration at high spatial resolution 1975-2020 [Dataset]* (JRC132876) [dataset]. Publications Office of the European Union. https://doi.org/10.2760/06672

DGT, D.-G. do T. (2023). *Carta Administrativa Oficial de Portugal—CAOP2022 (Continente)* (Version 2022) [dataset]. https://geo2.dgterritorio.gov.pt/caop/CAOP_Continente_2022-gpkg.zip

INE, P. (2002). *Censos 2001. XIV Recenseamento Geral da População: IV Recenseamento Geral da Habitação. Resultados definitivos: Portugal.* [dataset]. Instituto Nacional de Estatística. https://www.ine.pt/xurl/pub/133411

INE, P. (2012). *Censos 2011. XV Recenseamento Geral da População: V Recenseamento Geral da Habitação. Resultados definitivos: Portugal.* [dataset]. Instituto Nacional de Estatística. https://www.ine.pt/xurl/pub/73212469

Niva, V., Horton, A., Virkki, V., Heino, M., Kallio, M., Kinnunen, P., Abel, G. J., Muttarak, R., Taka, M., Varis, O., & Kummu, M. (2023). *Data for: World's human migration patterns in 2000-2019 unveiled by high-resolution data [Data set]* [dataset]. https://zenodo.org/records/7997134. https://doi.org/10.5281/zenodo.7997133

WorldPop. (2018). *Portugal, 100m. Population. Unconstrained individual countries 2000-2020 (100m resolution)* [dataset]. School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). https://dx.doi.org/10.5258/SOTON/WP00645

WorldPop. (2020). *Portugal, 1000m. Population density. Unconstrained individual countries 2000-2020 (1km resolution)* [dataset]. School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). https://dx.doi.org/10.5258/SOTON/WP00645