



Análise Espacial do Impacto da Gentrificação no Mercado Imobiliário de Lisboa

Cristina Dias Abrunhosa Araújo Morgado

Dissertação

apresentada como requisito parcial para obtenção do grau de Mestre em Ciência e Sistemas de Informação Geográfica

**NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação**

Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**ANÁLISE ESPACIAL DO IMPACTO DA GENTRIFICAÇÃO NO MERCADO IMOBILIÁRIO DE
LISBOA**

por

Cristina Dias Abrunhosa Araújo Morgado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciência e Sistemas de Informação Geográfica

Orientada por:

Ana Cristina Marinho da Costa, PhD, NOVA Information Management School

Abril, 2025

DECLARAÇÃO DE INTEGRIDADE

1. Assinatura digital

Declaro que o trabalho contido neste documento é da minha autoria e não de outra pessoa. Toda a assistência recebida de outras pessoas está devidamente identificada e é efetuada referência a todas as fontes utilizadas (publicadas ou não).

O trabalho não foi anteriormente submetido ou avaliado na NOVA Information Management School ou em qualquer outra instituição. Declaro ainda que tenho conhecimento das Regras de Conduta e do Código de Honra da NOVA Information Management School.

UTILIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL GENERATIVA

Tarefas	NÃO	SIM	Ferramentas de Inteligência Artificial Generativa
Compreender melhor as questões relacionadas com a investigação	X		
Resumir texto de bibliografia / recursos		X	Scispace
Resumo do(s) método(s) utilizado(s)	X		
Tradução de texto		X	DeepL Translate
Verificação gramatical		X	ChatGPT
Parafrasear ou reescrever texto de outras pessoas/recursos		X	ChatGPT; Paperpal
Codificação em R, Python, etc.	X		
Obter ajuda sobre um software	X		
Criação e edição de imagens, mapas, vídeos, etc.	X		
Análise de dados	X		
Especificar abaixo outras tarefas não mencionadas acima:			
Seleção de bibliografia		X	Litmaps

AGRADECIMENTOS

Embora esta dissertação seja, em última instância, um trabalho individual, o seu percurso e a sua concretização só foram possíveis graças ao apoio, incentivo e contributo das pessoas que me rodeiam. Sempre ouvi dizer que somos quem somos e conseguimos o que conseguimos graças às pessoas com quem nos cruzamos. Ao longo desta jornada, tive a sorte de encontrar quem me inspirasse, desafiasse e ajudasse a crescer. A todos, a minha profunda gratidão.

Em primeiro lugar, quero agradecer à minha orientadora, a Professora Doutora Ana Cristina Costa, pelo apoio incansável, pelas sugestões valiosas, pelos ensinamentos e pelas palavras motivadoras. A sua orientação foi essencial para a construção deste trabalho e para o meu crescimento enquanto profissional.

Agradeço também ao Professor Doutor Marco Painho, coordenador do Mestrado em Ciência e Sistemas de Informação Geográfica, pelo acompanhamento atento e pelas reflexões e sugestões que ajudaram a enriquecer esta dissertação.

À minha família pelo suporte incondicional e pelas palavras de incentivo que nunca faltaram. Em especial, ao meu irmão Zé, pelas ideias, pelas críticas construtivas e por me desafiar a olhar para este trabalho com uma perspetiva mais rigorosa. Ao Carlos, pelo amor, paciência e ajuda constante. O teu apoio foi essencial para que esta jornada chegasse a bom porto.

Finalmente, aos meus amigos, pelo companheirismo, pelas discussões enriquecedoras, mas também pelos momentos de descontração que ajudaram a tornar este percurso mais leve e gratificante. Em particular, ao Tiago, à Margarida e à Carolina, com quem me cruzei nesta fase da faculdade, e cujas ideias e sugestões foram especialmente valiosas no decorrer deste trabalho. E às minhas amigas de longa data, a Filipa e a Sara, por estarem sempre presentes, mesmo à distância, e por contribuírem com a sua amizade e apoio emocional.

ANÁLISE ESPACIAL DO IMPACTO DA GENTRIFICAÇÃO NO MERCADO IMOBILIÁRIO DE LISBOA

RESUMO

Nas últimas décadas, as grandes cidades europeias, incluindo Lisboa, têm sido palco de profundas transformações urbanas, frequentemente impulsionadas pelo fenómeno da gentrificação. Esta dinâmica tem contribuído para a deterioração do acesso à habitação, resultado de uma valorização acelerada do mercado imobiliário, desproporcional face aos rendimentos das famílias. Neste contexto, a presente dissertação tem como principal objetivo investigar de que forma os processos de gentrificação influenciam os preços do arrendamento na Área Metropolitana de Lisboa, bem como identificar os fatores socioeconómicos e urbanísticos mais determinantes. Para tal, são desenvolvidos dois modelos de regressão geograficamente ponderada multiescala (MGWR), aplicados aos anos censitários de 2001 e 2021. Os resultados confirmam a existência de autocorrelação e heterogeneidade espacial no mercado de arrendamento lisboeta, e evidenciam a relevância de variáveis como o nível de escolaridade, o regime de ocupação da habitação, o uso sazonal dos alojamentos e a idade dos edifícios, na explicação da variação espacial dos valores das rendas. Em última instância, pretende-se que este estudo contribua para uma compreensão mais aprofundada das dinâmicas territoriais associadas à gentrificação, oferecendo suporte ao planeamento urbano na formulação de políticas públicas mais equitativas, capazes de mitigar os efeitos negativos do aumento exacerbado dos custos habitacionais.

PALAVRAS-CHAVE

Análise Geoespacial; Gentrificação; Preços da Habitação; Preços de Arrendamento; Regressão Geograficamente Ponderada Multiescala

Objetivos do Desenvolvimento Sustentável (ODS):



SPATIAL ANALYSIS ON THE IMPACT OF GENTRIFICATION ON LISBON'S HOUSING MARKET

ABSTRACT

In the last few decades, major European cities, including Lisbon, have been the scene of profound urban transformations, often driven by the phenomenon of gentrification. This dynamic has contributed to a worsening of access to housing, as a result of the accelerated appreciation of the housing market, which has been disproportionate to household incomes. In this context, the main aim of this dissertation is to investigate how gentrification processes influence rental prices in the Lisbon Metropolitan Area, as well as to identify the most determining socio-economic and urban factors. For this purpose, two multiscale geographically weighted regression (MGWR) models were developed and applied to the 2001 and 2021 census years. The results confirm the existence of spatial autocorrelation and heterogeneity in Lisbon's rental market and highlight the relevance of variables such as level of education, housing ownership, seasonal use of dwellings and the age of the buildings, in explaining the spatial variation in rents. Ultimately, this study is intended to contribute to a more in-depth understanding of the territorial dynamics associated with gentrification, providing support for urban planning in the formulation of more equitable public policies capable of mitigating the negative effects of the sharp increase in housing costs.

KEYWORDS

Gentrification; Geospatial Analysis; House Prices; Multiscale Geographically Weighted Regression; Rent Prices

Sustainable Development Goals (SGD):



ÍNDICE DO TEXTO

DECLARAÇÃO DE INTEGRIDADE	ii
UTILIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL GENERATIVA	iii
AGRADECIMENTOS	iv
RESUMO	v
ABSTRACT	vi
ÍNDICE DO TEXTO	vii
ÍNDICE DE FIGURAS	ix
ÍNDICE DE TABELAS	x
LISTA DE SIGLAS E ABREVIATURAS	xi
1. Introdução	1
1.1. Contextualização da Problemática	1
1.2. Questão de Investigação e Objetivos	2
1.3. Abordagem Metodológica	2
1.4. Estrutura do Documento	3
2. Revisão de Literatura	4
2.1. Gentrificação	4
2.2. Modelação Espacial dos Preços da Habitação	6
2.3. Modelação dos Preços da Habitação em Portugal	8
2.4. Fatores que Afetam o Preço da Habitação	10
2.4.1. Características Estruturais	15
2.4.2. Características da Vizinhança	15
2.4.3. Atributos de Localização	16
2.4.4. Gentrificação	16
3. Dados e Metodologia	18
3.1. Área de Estudo	18
3.2. Metodologia	19
3.2.1. Dados	20
3.2.2. Recolha e Pré-processamento dos Dados	22
3.2.3. Análise Exploratória dos Dados Espaciais	22
3.2.4. Multicolinearidade Global	24
3.2.5. Modelo MGWR	25
4. Resultados do estudo de 2001	27
4.1. Análise Exploratória dos Dados Espaciais	27
4.1.1. Estatística Descritiva	28
4.1.2. Visualização dos Dados	29
4.1.3. Análise de Padrões Espaciais	30
4.1.4. Relações Bivariadas	31

4.2. Multicolinearidade Global.....	34
4.3. Modelo MGWR	35
4.3.1. Larguras de Banda e Estatísticas dos Coeficientes	35
4.3.2. Multicolinearidade Local	36
4.3.3. Observações Influentes	37
4.3.4. Adequação do Modelo	38
4.3.5. Análise de Resíduos	39
4.3.6. Coeficientes Locais	40
5. Resultados do estudo de 2021	43
5.1. Análise Exploratória dos Dados Espaciais	43
5.1.1. Estatística Descritiva	44
5.1.2. Visualização dos Dados.....	44
5.1.3. Análise de Padrões Espaciais	45
5.1.4. Relações Bivariadas	46
5.2. Multicolinearidade Global.....	49
5.3. Modelo MGWR	49
5.3.1. Larguras de Banda e Estatísticas dos Coeficientes	50
5.3.2. Multicolinearidade Local	51
5.3.3. Observações Influentes	52
5.3.4. Adequação do Modelo	53
5.3.5. Análise de Resíduos	54
5.3.6. Coeficientes Locais	55
6. Discussão de Resultados.....	58
6.1. Comparação dos Modelos (2001-2021).....	58
6.2. Os efeitos da Gentrificação	60
7. Conclusões e Investigação Futura	63
Referências Bibliográficas	67
Apêndice A: Histogramas da Distribuição dos Potenciais Fatores Explicativos de 2001	73
Apêndice B: Mapas Temáticos dos Potenciais Fatores Explicativos, em 2001	74
Apêndice C: Relações bivariadas locais significantes entre fatores explicativos e o valor médio mensal das rendas, em 2001	76
Apêndice D: Histogramas da Distribuição dos Potenciais Fatores Explicativos de 2021	77
Apêndice E: Mapas Temáticos dos Potenciais Fatores Explicativos, em 2021.....	78
Apêndice F: Relações bivariadas locais significantes entre fatores explicativos e o valor médio mensal das rendas, em 2021	80

ÍNDICE DE FIGURAS

Figura 3.1.1 Enquadramento geográfico da Área Metropolitana de Lisboa	19
Figura 3.2.1 Esquema do processo metodológico	20
Figura 4.1.1 Histograma da distribuição do valor médio mensal das rendas dos alojamentos familiares clássicos arrendados, em 2001	28
Figura 4.1.2 Valor médio mensal das rendas nos alojamentos familiares clássicos arrendados, em 2001	30
Figura 4.1.3 Local Moran's I e Getis-Ord GI da variável 2001_RENDAS	31
Figura 4.1.4 Gráficos de dispersão e matriz de correlação de Pearson do valor médio mensal das rendas e 18 potenciais variáveis explicativas, dados de 2001	32
Figura 4.3.1 Multicolinearidade local no modelo de 2001	37
Figura 4.3.2 Distância de Cook no modelo de 2001	38
Figura 4.3.3 Local R-squared no modelo de 2001	39
Figura 4.3.4 Resíduos estandardizados no modelo de 2001	40
Figura 4.3.5 Coeficientes locais do modelo de 2001	42
Figura 5.1.1 Histograma da distribuição do valor médio mensal das rendas dos alojamentos familiares clássicos arrendados, em 2021	44
Figura 5.1.2 Valor médio mensal das rendas nos alojamentos familiares clássicos arrendados, em 2021	45
Figura 5.1.3 Local Moran's I e Getis-Ord GI da variável 2021_RENDAS	46
Figura 5.1.4 Gráficos de dispersão e matriz de correlação de Pearson do valor médio mensal das rendas e 18 potenciais variáveis explicativas, dados de 2021	47
Figura 5.3.1 Multicolinearidade local no modelo de 2021	52
Figura 5.3.2 Distância de Cook no modelo de 2021	53
Figura 5.3.3 Local R-squared no modelo de 2021	54
Figura 5.3.4 Resíduos estandardizados no modelo de 2021	55
Figura 5.3.5 Coeficientes locais do modelo de 2021	57

ÍNDICE DE TABELAS

Tabela 2.4.1 Síntese de estudos relevantes sobre os preços habitacionais	11
Tabela 3.2.1 Descrição das potenciais variáveis explicativas	21
Tabela 4.1.1 Descrição das potenciais variáveis explicativas consideradas para 2001.....	28
Tabela 4.1.2 Classificação das relações bivariadas locais entre variáveis explicativas e a variável dependente (2001_RENDAS).....	33
Tabela 4.2.1 Resultados da multicolinearidade global para os dados de 2001	34
Tabela 4.3.1 Larguras de banda e estatísticas dos coeficientes para o modelo de 2001	36
Tabela 4.3.2 Medidas de adequação do modelo de 2001	38
Tabela 5.1.1 Descrição das potenciais variáveis explicativas consideradas para 2021.....	43
Tabela 5.1.2 Classificação das relações bivariadas locais entre variáveis explicativas e a variável dependente (2021_RENDAS).....	48
Tabela 5.2.1 Resultados da multicolinearidade global para os dados de 2021	49
Tabela 5.3.1 Larguras de banda e estatísticas dos coeficientes para o modelo de 2021	51
Tabela 5.3.2 Medidas de adequação do modelo de 2021	53
Tabela 6.1.1 Comparação da adequação dos modelos de 2001 e 2021	58
Tabela 6.1.2 Variáveis explicativas dos modelos de 2001 e 2021, por categoria de determinantes dos preços das habitações	59

LISTA DE SIGLAS E ABREVIATURAS

AICc – Critério de Informação de Akaike Corrigido (*Corrected Akaike's Information Criterion*)

AML – Área Metropolitana de Lisboa

CAOP – Carta Administrativa Oficial de Portugal

CN – Número de Condição (*Condition Number*)

COS – Carta de Uso e Ocupação do Solo

DGT – Direção Geral do Território

ENP – Número Efetivo de Parâmetros (*Effective Number of Parameters*)

ESDA – Análise Exploratória de Dados Espaciais (*Exploratory Spatial Data Analysis*)

GWR – Regressão Geograficamente Ponderada (*Geographically Weighted Regression*)

GTWR – Regressão Geograficamente e Temporalmente Ponderada (*Geographically and Temporally Weighted Regression*)

INE – Instituto Nacional de Estatística

MAUP – Problema da Unidade de Área Modificável (*Modifiable Areal Unit Problem*)

MGWR – Regressão Geograficamente Ponderada Multiescala (*Multiscale Geographically Weighted Regression*)

NOVA IMS – Nova Information Management School

OLS – Mínimos Quadrados Ordinários (*Ordinary Least Squares*)

PIB – Produto Interno Bruto

SIG – Sistemas de Informação Geográfica

VIF – Fator de Inflação de Variância (*Variance Inflation Factor*)

1. INTRODUÇÃO

1.1. CONTEXTUALIZAÇÃO DA PROBLEMÁTICA

Nos últimos anos, o acesso à habitação em Portugal tem-se tornado um dos desafios mais prementes em termos de coesão social e desenvolvimento urbano sustentável. De acordo com os dados do Eurostat¹, entre 2015 e 2023, os preços das casas em Portugal mais do que duplicaram, registando um aumento de 105,8%. Este crescimento contrasta fortemente com a média dos países da União Europeia (48,1%) e supera a taxa de inflação total acumulada no mesmo período (26,4%). O mercado de arrendamento também não ficou imune a esta tendência. Embora o aumento dos preços no arrendamento (21%) tenha sido mais moderado em comparação com a compra de habitação, ainda assim, seguiu um ritmo superior ao da média europeia (13,3%). Este cenário reflete uma crescente pressão sobre o acesso à habitação em Portugal, especialmente nas áreas urbanas, onde a procura é elevada e os preços continuam a subir de forma exacerbada. Consequentemente, denota-se uma pressão cada vez maior sobre as famílias, à medida que o desfasamento entre o custo da habitação e o rendimento disponível se acentua (Ciocchetta et al., 2024; Travasso et al., 2020).

Este fenómeno de valorização acelerada do mercado habitacional não pode ser dissociado dos processos de gentrificação que têm moldado profundamente a estrutura social e funcional das cidades portuguesas, particularmente a capital lisboeta. A gentrificação é definida por vários autores (an de Meulen & Mitze, 2014; Finio, 2022) como o processo de revitalização de bairros tradicionalmente habitados por residentes de baixos rendimentos que, face ao influxo de novos investimentos e à fixação de habitantes com maior capacidade económica e financeira, contribui diretamente para a escalada dos preços da habitação e, consequentemente, para a exclusão residencial das populações mais vulneráveis. Segundo Wilhelmsson et al. (2021), a gentrificação apresenta um duplo impacto sobre os centros urbanos. Por um lado, impulsiona a revitalização dos bairros, ao promover um ciclo de investimentos que aumenta as receitas fiscais, decorrentes da valorização imobiliária, e que, por sua vez, permite a expansão dos serviços locais e a melhoria das infraestruturas urbanas. Por outro lado, este processo acarreta, também, efeitos negativos, particularmente para populações de rendimentos mais reduzidos e idosos, que deixam de ter acesso a habitação a preços acessíveis e são pressionados a deslocar-se, levando consigo a identidade cultural e histórica que caracterizava estes bairros.

Em Lisboa, os efeitos da gentrificação e do aumento desproporcional dos preços da habitação têm-se tornado particularmente visíveis, potenciados por fatores como a pressão turística, o crescimento do alojamento local, os investimentos estrangeiros no setor imobiliário e a escassez de oferta habitacional a preços acessíveis. O conjunto destes processos tem contribuído para a deslocação progressiva da população residente para as periferias, aprofundando as desigualdades socioeconómicas e alastrando as dificuldades no acesso à habitação por todo o território.

¹ Consultado no dia 14 de abril de 2025, no *website* oficial do Serviço de Estatística da União Europeia, Eurostat: <https://ec.europa.eu/eurostat/web/interactive-publications/housing-2023>.

Apesar da relevância crescente do tema, a relação entre gentrificação e preços da habitação permanece pouco explorada através de abordagens quantitativas, sobretudo nos contextos urbanos portugueses. A maioria dos estudos existentes tende a abordar a gentrificação com base em análises qualitativas ou descritivas, raramente integrando métodos espaciais que permitam captar a complexidade territorial deste fenómeno.

1.2. QUESTÃO DE INVESTIGAÇÃO E OBJETIVOS

Perante a atualidade e importância da temática, surge a necessidade de investigar os diversos fatores que influenciam os preços da habitação, bem como as transformações sociais e urbanas que têm vindo a impulsionar o agravamento das dificuldades no acesso à habitação. Como tal, foi estabelecida a seguinte questão de investigação:

- De que forma os processos de gentrificação influenciam os preços da habitação em Lisboa, e quais são os fatores socioeconómicos e urbanísticos determinantes deste fenómeno?

Neste sentido, foram definidos cinco objetivos específicos, considerados pertinentes para auxiliar na estruturação do presente trabalho, sendo estes:

- Distinguir as variáveis quantitativas que caracterizam a gentrificação;
- Analisar os padrões espaciais dos preços da habitação em Lisboa – testar a veracidade do pressuposto de que existe autocorrelação espacial e heterogeneidade espacial no mercado imobiliário de Lisboa;
- Identificar os principais fatores socioeconómicos, ambientais e urbanos que afetam o preço das habitações em Lisboa;
- Construir um modelo hedónico adequado para avaliar a influência destes fatores no mercado de arrendamento da Área Metropolitana de Lisboa;
- Analisar a relação entre o processo de gentrificação e a variação espacial dos preços de arrendamento.

1.3. ABORDAGEM METODOLÓGICA

Para atingir os objetivos propostos, a abordagem metodológica adotada baseou-se na estatística espacial, tendo sido desenvolvidos dois modelos hedónicos de regressão geograficamente ponderada multiescala (MGWR), aplicados às freguesias da Área Metropolitana de Lisboa (AML), para os anos de 2001 e 2021. Este estudo teve como base dados censitários e cartográficos que, numa primeira instância foram submetidos a uma análise exploratória, através de ferramentas de estatística espacial, nomeadamente os testes de *Local e Global Moran's I* e *Getis-Ord Gi**, numa segunda fase, permitiram o desenvolvimento e interpretação dos modelos MGWR.

Este tipo de análise, possibilitada pelo desenvolvimento dos Sistemas de Informação Geográfica (SIG) e da geoestatística, permite incorporar a estrutura espacial do mercado de habitação com maior precisão, atendendo a fenómenos como a autocorrelação e a heterogeneidade espacial, aspetos frequentemente presentes no mercado imobiliário urbano, onde variáveis como a localização, a acessibilidade ou as infraestruturas próximas influenciam fortemente o valor das habitações (Anselin, 2002; Bitter, 2008).

Em última instância, pretende-se que esta dissertação contribua para o avanço do conhecimento na interseção entre a geografia urbana e a modelação espacial, ao fornecer uma abordagem empírica e quantitativa que permita explorar as complexas relações entre gentrificação e preços de arrendamento.

1.4. ESTRUTURA DO DOCUMENTO

Tendo a questão de investigação e os objetivos estabelecidos em mente, a estrutura desta dissertação segue um fluxo sequencial, que se pode dividir em sete capítulos fundamentais.

Primeiramente, este capítulo inicial (Introdução) introduz a temática em causa, contextualiza o fenómeno da gentrificação e da atual crise no acesso à habitação e estabelece os objetivos que se pretende atingir com a realização deste trabalho. De seguida, o capítulo 2 (Revisão de Literatura) discute, com base em bibliografia, os conceitos fundamentais associados à modelação espacial dos preços da habitação, incluindo os fatores que influenciam os valores das rendas e a relação com a gentrificação. O terceiro capítulo (Dados e Metodologia) descreve a área de estudo, os dados utilizados e os métodos aplicados. Os capítulos 4 (Resultados do Estudo de 2001) e 5 (Resultados do estudo de 2021) apresentam todo o processo metodológico inerente à aplicação dos respetivos modelos MGWR de 2001 e 2021, desde a análise exploratória dos dados, até à avaliação da adequabilidade dos modelos. No sexto capítulo (Discussão de Resultados) é realizada uma comparação entre os dois modelos desenvolvidos, bem como uma análise dos efeitos da gentrificação no mercado de arrendamento de Lisboa. Finalmente, o último capítulo (Conclusões e Investigação Futura) sintetiza as principais conclusões deste trabalho, destacando as suas implicações teóricas e práticas, e propondo possíveis linhas de investigação futura.

2. REVISÃO DE LITERATURA

A presente revisão de literatura foi elaborada com o objetivo de compilar e analisar as principais contribuições científicas relevantes para o estudo da modelação espacial dos preços da habitação e a sua relação com o fenómeno da gentrificação. O processo de recolha de artigos foi estruturado em diferentes fases, seguindo um método maioritariamente crítico, de seleção com base na adequação aos objetivos estabelecidos.

Primeiramente, a pesquisa foi conduzida na plataforma *Scopus*, utilizando palavras-chave como *hedonic model*, *regression*, *housing market*, *house price* e *gentrification*. Estes termos permitiram identificar estudos que discutem a modelação dos preços da habitação através de métodos de regressão, com ênfase na aplicação de modelos hedónicos. Com o auxílio da ferramenta *VOSviewer*, foi realizada uma breve análise bibliométrica dos artigos encontrados que, não só elucidou sobre a atualidade e relevância do tema da gentrificação, mas também revelou que esta palavra-chave surge frequentemente associada a fatores sociais, mas pouco relacionada com análises quantitativas e espaciais. Esta constatação reforçou a importância de explorar a interação entre os processos de gentrificação e os preços da habitação numa perspetiva espacial, mote desta dissertação.

Numa fase complementar, foi utilizada a inteligência artificial generativa *Litmaps*, que permitiu identificar outros artigos não indexados no *Scopus*, mas comumente citados por vários autores ou associados aos primeiros estudos sobre os temas abordados.

Finalmente, para o subcapítulo que aborda a modelação dos preços da habitação em Portugal, foi utilizado o agregador de recursos NOVA *Discovery Service*, para mais facilmente pesquisar por artigos de autoria e contexto nacional.

2.1. GENTRIFICAÇÃO

O termo *gentrificação* foi introduzido pela socióloga Ruth Glass (1964), para descrever o processo de transformação urbana observado em bairros operários de Londres, onde populações de classe média-alta começavam a substituir os residentes de rendimentos mais baixos. Desde então, o conceito foi debatido em diversas investigações, tendo-se disseminado por todo o mundo. Atualmente, a gentrificação pode ser definida como um processo de transformação urbana, caracterizado pelo influxo de investimento e pela substituição gradual da população residente por grupos com rendimentos e níveis de escolaridade mais elevados. Este fenómeno é frequentemente evidenciado pela reabilitação do edificado, bem como pela valorização dos imóveis, infraestruturas e serviços locais (Banabak, 2024; Finio, 2022; Wilhelmsson et al., 2021).

No entanto, a gentrificação é consensualmente reconhecida como um fenómeno multifacetado, resultante da interação de fatores económicos, sociais, políticos e espaciais, que se expressa de forma diferenciada nos territórios (Finio, 2022). Na verdade, Gago & Cocola-Gant (2019), defendem que Lisboa terá passado por fases de gentrificação diferentes das propostas no modelo teórico anglo-saxónico, especialmente pelo facto de ainda habitar no centro de Lisboa população idosa e com poucas capacidades económicas. Ao contrário do que terá acontecido nas cidades industrializadas ocidentais, em Lisboa, a gentrificação ocorreu essencialmente por meio

do arrendamento direcionado ao turismo. Isto é, a gentrificação foi impulsionada pelo influxo de investimento na compra e reabilitação de casas para a obtenção de lucros através do mercado de arrendamento de curta duração. Alinhados com esta premissa, vários autores exploram os impactos da turistificação² nos centros urbanos portugueses, sublinhando o papel central das políticas públicas de incentivo ao turismo e da liberalização do mercado imobiliário na intensificação deste processo (Daly et al., 2021; Lestegás et al., 2018; Mendes, 2017).

Estudos recentes têm procurado identificar os principais fatores que contribuem para a gentrificação. De entre estes, destacam-se os fatores económicos, nomeadamente a dinâmica entre oferta e procura, isto é, a crescente procura por habitação nas áreas urbanas leva ao aumento dos preços, especialmente nas áreas onde a elasticidade da oferta é mais reduzida, ou seja, nas áreas onde a urbanização é mais densa e a oferta de terrenos disponíveis é mais limitada (Murphy, 2024). Por outro lado, as decisões políticas surgem frequentemente como um dos principais motores da gentrificação. Finio (2022) refere que, muitas vezes, as políticas que favorecem o desenvolvimento urbano, o investimento em infraestruturas e a revitalização de determinadas áreas podem catalisar este fenómeno. No contexto português, Lestegás et al. (2018) e Mendes (2017) criticam a adoção do turismo como principal estratégia económica, referindo que as políticas do governo local que favorecem o turismo e o desenvolvimento imobiliário podem acelerar a gentrificação, na medida em que provocam uma financeirização do mercado de habitação, observando-se a substituição dos arrendamentos de longo prazo pelos de curta duração. Paralelamente, os atributos de localização são igualmente apontados como fatores determinantes da suscetibilidade dos bairros à gentrificação. Áreas próximas de infraestruturas culturais, centros da cidade e transportes públicos tornam-se mais atrativas a novos residentes e, portanto, têm maior probabilidade de sofrer gentrificação (Finio, 2022).

Para quantificar este fenómeno, os investigadores recorrem a metodologias diversas (Finio, 2022; Hawkins et al., 2022; Murphy, 2024). Segundo Finio (2022), os dados censitários continuam a ser uma das principais fontes, pois permitem acompanhar as alterações socioeconómicas e demográficas a pequenas escalas espaciais. Outro método habitual passa pela criação de índices compostos, a partir de variáveis como o rendimento e a escolaridade da população, permitindo medir as mudanças na composição socioeconómica dos bairros e, consequentemente classificá-los como gentrificados. Para tal, torna-se necessário discriminar as variáveis que poderão indicar a presença deste fenómeno que, segundo Finio (2022), podem ser classificadas de acordo com duas categorias fundamentais: a oferta e a procura. A oferta representa as características do tecido edificado, que pode ser identificado através de variáveis como: quota das rendas da *Airbnb*; idade dos edifícios; número de quartos nos alojamentos; impostos sobre a propriedade; regime de ocupação; e tamanho do alojamento. Por sua vez, a procura ilustra as características socioeconómicas da vizinhança, identificadas a partir de variáveis como: idade da população; nível de escolaridade; taxa de desemprego; rendimento familiar; dimensão dos agregados familiares; e densidade populacional.

² Termo frequentemente associado à gentrificação turística, utilizado para retratar as transformações sociais, culturais, económicas e urbanas decorrentes do aumento do turismo em massa nas cidades (Álvarez-Herranz & Macedo-Ruiz, 2021; Daly et al., 2021; Mendes, 2017).

2.2. MODELAÇÃO ESPACIAL DOS PREÇOS DA HABITAÇÃO

A modelação dos preços da habitação é feita, desde meados da década de 1970 (Rosen, 1974), através de modelos hedónicos. Um modelo hedónico baseia-se numa abordagem estatística que se utiliza para determinar o valor de um bem, com base nos seus atributos específicos e quantificáveis (Droj et al., 2024; Fotheringham et al., 2015). Neste caso, em vez de se considerar um imóvel como um valor único, o modelo hedónico desagrega-o em diversas características físicas, de localização e de vizinhança, assumindo que cada uma delas terá um peso diferente no valor total da habitação (Anselin & Lozano-Gracia, 2009).

O modelo hedónico de preços da habitação é frequentemente utilizado como uma ferramenta de apoio à decisão na avaliação imobiliária, pois permite entender como diferentes fatores influenciam os preços das casas e, subsequentemente, fazer previsões sobre tal valorização (Batista & Marques, 2021). Por outro lado, dada a importância do mercado imobiliário para a economia, estes estudos assumem um papel relevante no apoio à administração local e central, no estabelecimento de medidas que garantam o direito à habitação. Simultaneamente, auxiliam no desenvolvimento de técnicas que apoiem a valorização dos impostos prediais, uma das principais fontes de receitas dos municípios, a par com o aumento dos preços da habitação (Droj et al., 2024; Shi et al., 2022).

Este modelo é tradicionalmente implementado através do modelo de regressão de mínimos quadrados ordinários (OLS), que permite determinar o valor de cada atributo, assumindo uma relação linear entre eles e o preço da habitação (Fotheringham et al., 2015). No entanto, várias críticas têm surgido relativamente à sua incapacidade de lidar com os efeitos espaciais inerentes às dinâmicas do mercado de habitação: a dependência espacial e a heterogeneidade espacial (Chica-Olmo et al., 2013; Lee et al., 2016).

No contexto do mercado habitacional, a dependência espacial, também conhecida como autocorrelação espacial, refere-se ao facto de o preço de um imóvel estar tendencialmente relacionado com os preços das habitações vizinhas e com os atributos de localização compartilhados entre si, isto é, as habitações em áreas próximas influenciam e são influenciadas por fatores comuns (Bitter, 2008). Ora, este princípio viola o pressuposto base dos modelos de regressão tradicionais da independência das observações, pelo que a omissão da autocorrelação espacial nos modelos hedónicos pode resultar na identificação errónea de fatores que influenciam os preços da habitação, na perda de capacidade explicativa do modelo e, consequentemente, pôr em causa a confiabilidade da análise (Basu & Thibodeau, 1998; Bitter et al., 2007).

Por sua vez, a heterogeneidade espacial nos preços da habitação refere-se à variação dos fatores que influenciam o valor das casas em diferentes áreas geográficas. Ocorre porque os processos que determinam os preços, tais como as preferências dos compradores e as condições de oferta e procura, mudam conforme o bairro ou a região. Esta variabilidade espacial reflete a existência de submercados de habitação, onde imóveis localizados em áreas diferentes podem ser valorizados de forma distinta, mesmo que tenham características semelhantes (Marques et al., 2013; Yao & Fotheringham, 2016).

Com o intuito de incluir os efeitos da dependência espacial nos modelos hedônicos, Anselin (1988) desenvolveu os primeiros modelos econométricos espaciais, introduzindo métodos como as matrizes de ponderações espaciais (*spatial weight matrix*), *spatial lag* e expansão espacial (*spatial expansion*). A matriz de ponderação espacial, integrada nos modelos baseados em *spatial lag*, veio permitir a quantificação das interações entre unidades espaciais e especificar a intensidade da influência das vizinhanças na variável dependente, tendo, assim, em conta os efeitos da dependência espacial (Anselin, 2002). Por outro lado, os métodos de expansão espacial tiveram por objetivo incluir os efeitos da heterogeneidade espacial, partindo do princípio de que os coeficientes que representam o impacto das variáveis independentes não são constantes em toda a área de estudo (Bitter et al., 2007).

Ainda no âmbito da heterogeneidade espacial, vários autores (Bhattacharjee et al., 2012; Chen et al., 2023; Y. Wu et al., 2020) exploraram a segmentação do mercado habitacional em submercados, com o intuito de melhorar a confiabilidade dos modelos hedônicos e, simultaneamente, fomentar a compreensão das dinâmicas do mercado de habitação, auxiliando na formulação de políticas de habitação urbana eficazes. Contudo, as metodologias de classificação dos submercados ainda não são consensuais. Segundo Wu et al. (2020), pode-se realizar esta divisão com base em dois métodos: de acordo com as divisões geográficas, administrativas ou censitárias; ou a partir de métodos orientados para os dados, como *k-means clustering* e árvores de decisão. Tal como em qualquer método de partição de dados, o objetivo primordial da segmentação do mercado de habitação passa por criar submercados que sejam substitutivos, similares e espacialmente contínuos, isto é, semelhantes e próximos entre si e, simultaneamente, distintos em relação aos restantes (Kopczewska & Ćwiakowski, 2021). No seu estudo sobre o mercado imobiliário de Aveiro, Portugal, Bhattacharjee et al. (2012) dividem a cidade costeira em vários submercados. Numa primeira análise à escala urbana, esta ação permitiu-lhes compreender as diferentes interações entre submercados, nomeadamente o submercado de Vera Cruz, inserido no *Central Business District* (CBD), que revelou interações negativas com as áreas suburbanas, indicando diferenças significativas na população que procura imóveis nestas áreas. Por outro lado, na análise à escala periurbana, a segmentação do mercado de habitação permitiu confirmar a heterogeneidade espacial dos fatores que influenciam os preços da habitação, verificando-se, por exemplo, a particularidade do submercado “praias” como o único em que as características físicas, como ter garagem ou varanda, não aparentam ter influência significativa nos preços das habitações.

Nesta sequência, Farber & Yeates (2006) surgem como pioneiros na aplicação da regressão geograficamente ponderada (GWR) no contexto da modelação espacial dos preços da habitação, uma técnica de regressão espacial que estima modelos locais para cada ponto de observação, ponderando os dados vizinhos conforme a sua distância ao ponto de regressão (Bottero et al., 2022). Utilizando o caso de estudo de Toronto, Canadá, Farber & Yeates (2006) comparam a adequação de vários modelos de regressão, concluindo que o modelo GWR demonstrou maior precisão global, dada a indubitável existência de autocorrelação espacial nos preços das habitações, conclusão esta corroborada, posteriormente, por outros autores como Bitter et al. (2007) e Yao & Fotheringham (2016).

Mais recentemente, têm surgido várias sugestões de melhoria do modelo GWR. Cada vez mais investigações aludem à importância de incorporar a componente temporal nos modelos hedónicos, dado que os fatores que influenciam os preços da habitação podem mudar ao longo do tempo, por meio das condições do mercado ou outras influências externas (Fotheringham et al., 2015; C. Wu et al., 2019). Ao integrar uma série temporal no modelo GWR, Fotheringham et al. (2015) analisam os padrões espaço-temporais da valorização imobiliária em Londres, sublinhando a necessidade de modelos que tenham em conta a não-estacionariedade espacial e temporal dos dados, para melhorar a compreensão e a previsão dos preços da habitação.

Paralelamente, vários autores (Lee et al., 2016; Yao & Fotheringham, 2016) dedicaram-se às questões da escala espacial, alertando para a sensibilidade dos modelos hedónicos ao problema da unidade de área modificável (MAUP). Segundo Lee et al. (2016), este problema é particularmente relevante no contexto dos modelos hedónicos, na medida em que os preços das habitações, geralmente obtidos a uma escala individual, são relacionados a características de bairros ou zonas, cujos valores são agregados em unidades espaciais. Naturalmente, esta abordagem pode levar a inconsistências, uma vez que os padrões espaciais ou correlações identificadas a uma escala podem não ser replicáveis a outra. Neste sentido, autores como Fotheringham & Sachdeva (2022), Lu et al. (2023) e C. Wu et al. (2019) recorrem à regressão geograficamente ponderada multiescala (MGWR), uma vez que este método permite que as relações entre variáveis sejam analisadas em diferentes escalas espaciais dentro do mesmo modelo, ajustando automaticamente a escala de análise para cada variável explicativa.

Nos últimos anos, a evolução tecnológica tem impulsionado o acesso a grandes volumes de dados, que exigem a utilização de métodos baseados em inteligência artificial e integram, entre outras, técnicas de deteção remota e *machine learning*, para alcançar estimativas e previsões cada vez mais precisas sobre os preços das habitações (Droj et al., 2024). Um exemplo desta abordagem é o trabalho desenvolvido por Soltani et al. (2024), onde é utilizada uma rede neuronal artificial para prever preços da habitação com base em variáveis de qualidade ambiental e acessibilidade a infraestruturas de transporte, na província de Teerão, Irão. Os resultados deste estudo revelam diversas vantagens do uso destas técnicas, nomeadamente a sua maior: precisão comparativamente aos modelos hedónicos tradicionais baseados em OLS; adaptabilidade; capacidade de modelar relações não lineares; e robustez face a ruído e *outliers*. Contudo, uma das principais limitações apontadas refere-se à complexidade da interpretação dos resultados. De acordo com Valier (2020), a principal diferença entre os modelos de regressão e os baseados em *machine learning* consiste no facto de os primeiros terem, geralmente, melhor capacidade de inferência, enquanto que os segundos apresentam melhor capacidade preditiva.

2.3. MODELAÇÃO DOS PREÇOS DA HABITAÇÃO EM PORTUGAL

Nos últimos anos, o estudo dos preços da habitação em Portugal tem despertado interesse. De facto, tem-se denotando uma crescente pressão sobre as famílias, que destinam uma porção cada vez maior dos seus rendimentos para os custos habitacionais, fenómeno que não é exclusivo a Portugal, mas sim comum a toda a Europa (Ciocchetta et al., 2024). Contudo, Portugal distingue-se não só pelo ritmo acelerado de crescimento dos

preços da habitação, como também pela dinâmica subjacente a esta problemática. A atratividade de Portugal como destino turístico, que estimulou o aumento explosivo da oferta de alojamentos de curta duração em detrimento dos de longa duração, aliada a um histórico de políticas de habitação desapropriadas e tardias, culminou na grave crise habitacional que se faz sentir nos dias de hoje (Antunes & Ferreira, 2021; Seixas & Antunes, 2019).

O setor habitacional em Portugal evidencia uma situação paradoxal. Por um lado, verifica-se a existência de um número significativo de fogos vagos, mas, por outro, muitas famílias enfrentam dificuldades em aceder a habitações que sejam compatíveis com os seus recursos financeiros e adequadas às necessidades decorrentes da dimensão dos seus agregados familiares (Santos & Castro, 2022).

Vários autores têm desenvolvido estudos no sentido de compreender as dinâmicas do mercado habitacional português.

Num estudo que abrange todo o território continental, G. Carvalho & Ribeiro (2007) desenvolveram uma análise espaço-temporal do mercado habitacional de Portugal, entre 1995 e 2001, cujo objetivo primordial consistiu em compreender a dicotomia litoral-interior característica das dinâmicas socioeconómicas do país. Na verdade, os autores verificaram a existência de dois clusters de valores elevados de oferta em ambas as áreas metropolitanas do Porto e Lisboa e alguns clusters de valores reduzidos no interior do país. No mesmo seguimento, detetaram um grande cluster de valores elevados de fogos vagos no Alentejo, facto indicativo do êxodo rural que afetou a região, especialmente no final do século XX e início do século XXI.

Por sua vez, Bhattacharjee et al. (2016) definiram submercados de habitação na região urbana de Aveiro e Ílhavo através da aplicação de um modelo hedónico baseado numa matriz de ponderações espaciais e análise de *clusters*, alertando para os vários benefícios da segmentação dos mercados de habitação: auxilia os investigadores na compreensão da variação espacial dos preços da habitação; melhora a capacidade de avaliação dos riscos associados ao financiamento de habitação própria, por parte dos bancos e credores; e facilita a pesquisa dos consumidores.

Já Macdonald & Franco (2016) aliam as técnicas de deteção remota e fotointerpretação aos modelos hedónicos, para analisar o impacto dos espaços verdes urbanos no mercado imobiliário da AML, concluindo que, embora a direta proximidade ao Parque Urbano de Monsanto não aparente ter significância, os imóveis que se encontram a cerca de 750 a 1500 metros deste espaço verde tendem a ser valorizados em cerca de 4,2%. De um modo geral, a proximidade a florestas, parques e jardins urbanos afeta positivamente o mercado imobiliário, já que oferecem instalações de lazer e qualidade de vida à população lisboeta.

Por sua vez, Cunha & Lobão (Cunha & Lobão, 2022) criticam a liberalização do mercado de arrendamento ao concluírem que um aumento de 1% na proporção de arrendamentos de curta duração em relação ao parque habitacional é acompanhado por um aumento dos preços da habitação em 27,4% e 16,1%, nos municípios mais expostos ao arrendamento de curta duração, das áreas metropolitanas de Lisboa e Porto, respetivamente.

Lisboa, enquanto capital europeia com uma localização geoestratégica privilegiada, atrai investimentos internacionais, expatriados e turistas, que têm pressionado o mercado imobiliário. Políticas como os *Vistos Gold* e a isenção fiscal para residentes não habituais contribuíram para a especulação imobiliária, em detrimento das necessidades locais (Antunes & Seixas, 2022). Por outro lado, a AML é composta por municípios com realidades socioeconómicas e níveis de desenvolvimento urbano muito distintos, o que torna esta região num caso de estudo emblemático para compreender os padrões espaciais dos fatores que impulsionam a valorização imobiliária.

2.4. FATORES QUE AFETAM O PREÇO DA HABITAÇÃO

Os fatores que influenciam os preços das habitações podem ser divididos em diversas categorias. Tipicamente, a bibliografia considera três grandes classes de determinantes dos preços das habitações (Bhattacharjee et al., 2012; Fotheringham et al., 2015; Lee et al., 2016):

- Características estruturais, tais como a idade da habitação, a área e o número de quartos;
- Características da vizinhança, como a taxa de desemprego, a densidade populacional e o rendimento;
- Atributos de localização, por exemplo, a distância ao CBD, ou a proximidade às infraestruturas de transportes públicos.

O processo de seleção das variáveis explicativas utilizadas no modelo de regressão difere de acordo com os objetivos específicos de cada investigação. Contudo, importa ter em atenção que o número de variáveis independentes escolhidas não deve ser exageradamente reduzido, na medida em que poderá comprometer a interpretação dos resultados. Um número insuficiente de variáveis pode dificultar a avaliação precisa da variação espacial observada no modelo, ao tornar incerto se esta decorre efetivamente das variações nos parâmetros ou da omissão de fatores explicativos importantes (Bitter, 2008).

Por outro lado, deve-se evitar a inserção de variáveis redundantes, já que estas podem causar multicolinearidade, um problema estatístico que ocorre quando duas ou mais variáveis explicativas estão fortemente correlacionadas entre si. A multicolinearidade impede a estimativa precisa dos coeficientes do modelo, pois torna mais difícil distinguir os efeitos individuais de cada variável sobre a variável dependente, o que pode levar a interpretações erróneas e à redução da confiabilidade das inferências estatísticas (Álvarez-Herranz & Macedo-Ruiz, 2021; Ley, 1986; Macdonald & Franco, 2016).

A Tabela 2.4.1 sintetiza várias investigações, com o objetivo de auxiliar a seleção das variáveis mais relevantes que caracterizam não só os preços habitacionais, como também o fenómeno da gentrificação.

Tabela 2.4.1 Síntese de estudos relevantes sobre os preços habitacionais

Artigo	Área de Estudo	Metodologias	Variável Dependente (Y)	Variável Explicativa / Independente (X)
(Al-Masum & Lee, 2019)	Sydney, Austrália	1) Vector error correction model	Preço mediano trimestral de venda de habitações	Produto interno bruto, rendimento disponível bruto das famílias, nº total de fogos nos sectores privado e público, população, taxa de juro dos empréstimos à habitação, taxa de desemprego
(an de Meulen & Mitze, 2014)	Berlim, Alemanha	1) Regressão linear (OLS) 2) ESDA	Preço de arrendamento (€/m²)	Duração do anúncio (meses); idade do imóvel (anos); área (m²); andar em que se situa o apartamento; nº de divisões; saldo migratório líquido; população com antecedentes migratórios (%); dimensão média dos agregados familiares; agregados familiares isolados (%); edifícios de propriedade coletiva (%); habitações ocupadas pelos proprietários (%)
(Atuesta & Hewings, 2019)	Chicago, USA	1) Estatística descritiva 2) coeficiente de efeitos de vizinhança 3) Regressão linear (OLS)	Preços da habitação (log)	Gentrificação ³ ; rendimento mediano das famílias; total de licenças de construção; população branca (%); população negra (%); população hispânica (%); total de alojamentos ocupados por arrendatários; total de alojamentos ocupados por proprietários; proporção de pessoas com mais de 5 anos que residem na mesma casa há cinco anos; população em dormitórios universitários (%); total de unidades habitacionais vagas; população com mais de 16 anos com bacharelato ou licenciatura (%); distância ao CBD; distância a uma estação L ⁴ ; distância à linha de metro; idade da propriedade; nº de quartos; propriedades com ar condicionado central (%); propriedades com lareira (%); propriedades com construção em tijolo (%); propriedades com cave (%); área do lote
(Banabak, 2024)	Viena, Áustria	1) Regressão linear (OLS) 2) Índice de Fischer	Preço de renda por metro quadrado (log)	Área (m²); índice de tempo em transportes públicos até locais centrais; nº de quartos; nº de pisos; varanda; elevador; cave; garagem; aquecimento

³ Variável artificial criada para distinguir áreas gentrificadas (1) de áreas gentrificáveis (0)

⁴ As estações “L” referem-se às paragens do sistema de trânsito rápido que serve a metrópole de Chicago, conhecido como “Chicago L”.

(Batista & Marques, 2021)	Sintra, Portugal	1) GWR 2) Programação em R	Preço das transações de habitação (€)	Área (m ²); idade (anos); área do quintal/terreno (rácio); tipo de edifício; nº de pisos do edifício; idade da população; escolaridade da população; localização do local de trabalho da população
(Bhattacharjee et al., 2012)	Aveiro, Portugal	1) Matriz de ponderações espaciais 2) Regressão linear (OLS)	Preço das transações de habitação (€/m ²)	Tipologia de alojamento (apartamento/moradia); duplex; varanda; terraço; garagem; TV cabo; gás natural; nº de quartos; idade; nº de pisos; área da cozinha (m ²); área da sala (m ²); área total (m ²)
(Bitter et al., 2007)	Tucson, Arizona, USA	1) Regressão linear (OLS) 2) GWR	Preço de venda (US \$)	Nº de pátios; área da habitação (sqft); dimensão do lote (sqft); nº de quartos/área da habitação; idade; instalações sanitárias/número de divisões; qualidade interior da habitação; qualidade estrutural da habitação
(Bottero et al., 2022)	Brisbane, Austrália	1) Regressão linear (OLS) 2) GWR	Preço do Apartamento (AUS \$)	Nº de quartos; nº de casas de banho; nº de carros; distância ao parque urbano; distância ao parque urbano recreativo; distância ao parque urbano desportivo
(Chica-Olmo et al., 2013)	Granada, Espanha	1) Regressão linear (OLS) 2) Kriging 3) Cokriging	Preço do apartamento (€/m ²)	Idade do apartamento (ajustado para grandes renovações); amplitude (metros quadrados construídos/nº de quartos); nº de casas de banho; aquecimento central; distância ao CBD; elevador; piso em que o apartamento se encontra; índice de qualidade da vizinhança
(Geng et al., 2011)	Shenzhen, China	1) GWR	Preço médio de mercado dos conjuntos habitacionais	rácio de lote; rácio de área verde; área; estrutura do edifício; distância até à escola mais próxima; distância até ao hospital mais próximo; distância até ao parque mais próximo; distância até ao centro comercial mais próximo; distância até à bomba de gasolina mais próxima; distância até à estrada mais próxima; distância até à estação de metro mais próxima
(Kopczewska & Ćwiakowski, 2021)	Varsóvia, Polónia	1) GWR 2) Análise de clusters	Preço da habitação (PLN/m ²)	Garagem; nº de quartos; piso em que o apartamento se encontra; nº de pisos acima do piso térreo; nº de pisos abaixo do piso térreo; cave; área bruta de cobertura do edifício (m ²); renovação; idade do edifício; distância ao centro da cidade de carro (m); distância ao centro da cidade de transportes públicos (m); latitude; longitude; tempo de condução até ao centro da cidade; tempo em

				transportes públicos até ao centro da cidade; tempo de condução até ao centro da cidade em hora de ponta; área de floresta num raio de 1Km (m ²); áreas florestais num raio de 1km (%); áreas verdes num raio de 1km (%); áreas privadas num raio de 1km (%); áreas públicas num raio de 1km (%); área verde num raio de 1km (m ²); nº de edifícios históricos num raio de 1km; áreas históricas num raio de 1km (%)
(Lee et al., 2016)	Seoul, Coreia do Sul	1) Regressão linear (OLS) 2) Spatial Autoregressive Model	Preço de venda da habitação (KRW/m ²)	Idade; nº médio de pisos; área média por agregado familiar; área útil; distância até à escola primária mais próxima; distância até à estação de metro mais próxima; nº de paragens de autocarro num raio de 500 m; distância até à área verde mais próxima; nº de estabelecimentos de ensino privado; nº de creches por criança; nº de centros médicos por pessoa; densidade populacional; nº de licenciados; propriedade de casas.
(Ma & Gopal, 2018)	Massachusetts, USA	1) GWR 2) Mixed GWR	Preço mediano das habitações (\$)	Densidade populacional; floresta desprotegida (%); taxa de desemprego (%); área residencial; nº veículos per capita; população com licenciatura ou superior com mais de 25 anos (%); população sénior (%); distância até à estação ferroviária; imposto sobre imóveis residenciais; desempenho dos alunos em Matemática
(Soltani et al., 2024)	Teerão, Irão	1) Redes neuronais artificiais 2) Regressão linear (OLS)	Preço das habitações (log)	Área (m ²); idade; nº de quartos; edifícios situados no mapa do tecido urbano degradado; distância ao CBD; distância ao aeroporto (m); distância aos parques de bicicletas (m); distância à estação BRT ⁵ (m); distância à estação de autocarro (m); distância às estradas principais (m); distância às linhas de metro (m); distância às estações de metro (m); distância a grandes cruzamentos (m); distância ao parque de estacionamento (m); distância à linha de comboios suburbanos (m); distância às paragens de táxi (m); distância ao terminal rodoviário (m); distância a lares de idosos (m); distância a centro de escritórios (m); distância a usos do solo para serviços de saúde (m); distância a usos do solo para serviços de educação (m)

⁵ Autocarro de trânsito rápido, também conhecido como metro de superfície (*Bus Rapid Transit*, em inglês).

(Teixeira et al., 2011)	Castelo Branco, Portugal	1) Redes neuronais artificiais	Preço de venda da habitação (€)	Estado do imóvel; área útil (m ²); nº de assoalhadas; nº de casas de banho; varanda; lareira; ar condicionado; aquecimento central; janelas; eletrodomésticos; arrecadação; garagem; condomínio; nº de andares do prédio; elevador; videovigilância; localização
(Wilhelmsson et al., 2021)	Stockholm, Suécia	1) Getis-Ord Gi 1) Regressão linear (OLS)	Preço da habitação (SEK)	Área de habitação; nº de divisões; mensalidades; distância até ao CBD; distância até ao metro; número de pisos do edifício; elevador; idade; distância até aos locais gentrificados (Getis-Ord)
(C. Wu et al., 2019)	Shenzhen, China	2) MGTWR/GTWR	Preço da habitação (Yuan/m ²)	Rácio de espaços verdes; rácio de lotes; rácio de lugares de estacionamento; distância ao CBD; distância à estação ferroviária; distância à estação de metro; densidade de estações de autocarros em 500 m; distância ao parque; distância ao hospital; distância à escola; rácio de emprego e de residentes; densidade populacional
(Yao & Fotheringham, 2016)	Fife, Escócia	1) Semiparametric GWR	Preço da habitação (£)	Densidade populacional; população em idade de reforma (%); população em idade ativa (%); alojamentos geminados (%); alojamentos em banda (%); alojamentos em apartamentos (%); alojamentos com uma a três divisões (%); alojamentos com sete a nove divisões (%); famílias proprietárias de habitações (%); nº de crimes por 10 000 habitantes; distância ao CBD; distância à costa
(Zaki et al., 2022)	Boston, USA	1) Regressão linear (OLS) 2) Machine learning (XGBoost)	Preço da habitação (\$)	Taxa de criminalidade per capita; proporção de hectares de comércio não retalhista por cidade; concentração de óxido nítrico; proporção de unidades construídas antes de 1940 e ocupadas pelos proprietários; proporção de terrenos residenciais que foram divididos em zonas para lotes com mais de 25000 sqft; nº médio de quartos por habitação; medida de acessibilidade às autoestradas radiais; rácio aluno-professor em cada cidade; valor médio das casas ocupadas pelos proprietários em \$1000; pessoas com estatuto inferior em relação à população total (%); distâncias ponderadas a 5 centros de emprego de Boston; taxa de imposto sobre a propriedade de valor total

As três categorias determinantes dos preços das habitações, mencionadas anteriormente, bem como o fator Gentrificação, são discutidos com maior detalhe nas secções seguintes com o objetivo de auxiliar a seleção das variáveis para o presente estudo.

2.4.1. CARACTERÍSTICAS ESTRUTURAIS

As características físicas dos imóveis, estão diretamente relacionadas com o seu preço, embora a sua relação possa nem sempre ser totalmente linear. A área útil da habitação geralmente apresenta uma correlação positiva com o preço, isto é, quanto mais amplo for o espaço habitacional, maior será o seu preço. No entanto, a partir de determinada área, o espaço adicional pode deixar de acrescentar valor (Soltani et al., 2024). Similarmente, a idade dos edifícios tende a estar associada à deterioração, desgaste e obsolescência funcional, o que desvaloriza as propriedades. Contudo, alguns edifícios antigos podem ser valorizados dada a sua arquitetura histórica, ou devido à sua localização aprazível (Soltani et al., 2024).

Estas características são frequentemente analisadas como um conjunto, pois são indicativas da qualidade ou habitabilidade global do imóvel. Por exemplo, uma habitação que tenha, cumulativamente, sistema de aquecimento central, jardim, cave e um elevado número de quartos será mais valorizada do que uma que apenas apresente uma destas características (an de Meulen & Mitze, 2014).

Não obstante, embora sejam importantes, as características estruturais não são suficientes para corretamente modelar os preços das habitações, pois não só estão relacionadas umas com as outras, como a sua importância depende, igualmente, de outros fatores como a localização (Bitter et al., 2007).

2.4.2. CARACTERÍSTICAS DA VIZINHANÇA

As características da vizinhança representam a qualidade do meio envolvente, geralmente o bairro ou freguesias, de uma habitação. A taxa de criminalidade, a escolaridade da população, o rendimento familiar e a proporção de áreas verdes são alguns dos fatores que poderão contribuir para a variação dos preços habitacionais (Kopczewska & Cwiakowski, 2021; Ma & Gopal, 2018; Zaki et al., 2022).

No seu estudo, Al-Masum & Lee (2019) concluíram que o rendimento bruto das famílias, a oferta de habitação, a taxa de desemprego, e o Produto Interno Bruto (PIB) são estatisticamente significantes na explicação dos preços da habitação em Sydney, enquanto que a população e as taxas de juro dos empréstimos não revelaram ter importância significativa. Lee et al. (2016), por outro lado, referem que o estatuto educacional da população vizinha relaciona-se positivamente com os preços das habitações e a percentagem de habitações ocupadas pelos proprietários revela uma correlação negativa. Similarmente, Samadani & Costa (2021) concluíram que o poder de compra das famílias, a criminalidade e os impostos aparentam ser fatores que afetam substancialmente os preços habitacionais em Portugal.

Alguns autores como Chica-Olmo et al. (2013), ao invés de estudarem estes fatores individualmente, optaram por combiná-los num único índice de qualidade do bairro, associando várias características como a qualidade das escolas, o *status* socioeconómico e a qualidade ambiental.

2.4.3. ATRIBUTOS DE LOCALIZAÇÃO

Por sua vez, os atributos de localização caracterizam a acessibilidade das habitações às instalações, equipamentos e serviços urbanos, já que este é um parâmetro que em muito compromete a escolha dos consumidores na compra ou arrendamento de uma habitação.

Uma das variáveis recorrentemente utilizada nas investigações, consiste na distância ao CBD. O CBD, que frequentemente coincide com o centro da cidade, é onde se concentram grande parte dos serviços, instalações comerciais e polos de desenvolvimento científico e tecnológico, pelo que a proximidade e acessibilidade a este centro tende a aumentar os preços das habitações (Wilhelmsson et al., 2021; Yao & Fotheringham, 2016).

A distância a infraestruturas de transportes é por muitos considerado um fator fulcral para explicar os preços habitacionais (Lee et al., 2016; Soltani et al., 2024; C. Wu et al., 2019). Contudo, a relação entre estas variáveis pode ser complexa e variar não só espacial e temporalmente, como também consoante o meio de transporte analisado. Por um lado, a proximidade a infraestruturas de transporte pode ser valorizadora, ao significar maior acessibilidade a serviços e outras localizações. Por outro lado, esta característica pode influenciar negativamente os preços das habitações, na medida em que estas infraestruturas podem causar ruído, poluição e congestionamento (Soltani et al., 2024). Esta complexidade leva a que sejam obtidos resultados ambíguos. C. Wu et al. (2019) concluíram que a proximidade das estações de metro e a densidade de paragens de autocarro influenciam positivamente a variável dependente. Contrariamente, Wilhelmsson et al. (2021) verificaram que, em Estocolmo, quanto maior distância à estação de metro, maior será o valor da propriedade. Por sua vez, Soltani et al. (2024) identificaram que, enquanto que a proximidade ao aeroporto, ao terminal de autocarros e a cruzamentos de estradas principais relaciona-se negativamente com os preços das habitações, a proximidade às paragens do metro de superfície (BRT) e aos parques de bicicletas partilhadas tem um efeito positivo no valor dos imóveis.

Outros atributos de localização importantes passam pela proximidade a outros serviços e equipamentos urbanos utilizados no dia-a-dia das pessoas. Bottero et al. (2022), por exemplo, concluíram que a proximidade a parques urbanos recreativos valoriza em cerca de 7% os preços das habitações em Brisbane, Austrália. Paralelamente, Yao & Fotheringham (2016) atestam a importância da inclusão destas variáveis nos modelos hedónicos, ao verificar que as variáveis “distância ao centro da cidade” e “distância à costa” são as que apresentam maior variação espacial sobre o seu impacto nos preços dos imóveis. Num estudo diretamente focado na valorização imobiliária provocada pela gentrificação, Wilhelmsson et al. (2021) aferem que a proximidade a bairros gentrificados pode aumentar o valor das habitações em cerca de 6% a 8%.

2.4.4. GENTRIFICAÇÃO

Finalmente, uma vez que esta dissertação se foca nos impactos da gentrificação, torna-se essencial distinguir as variáveis quantitativas que melhor caracterizam este fenómeno.

Segundo Carvalho et al. (2019) um conjunto de fatores que ocorreram gradualmente durante décadas na cidade do Porto, Portugal, foram motores fundamentais da gentrificação. Primeiramente, deu-se um influxo de residentes com rendimentos mais elevados, seguido de um decréscimo populacional nas freguesias do centro, e acompanhado por alterações no cenário de propriedade, isto é, um aumento da proporção de habitações arrendadas em detrimento das ocupadas pelo proprietário. Por outro lado, dada a ligação intrínseca do processo de gentrificação em Portugal ao aumento do turismo urbano, outros indicadores deste fenómeno apontados pelos autores são o aumento dos estabelecimentos hoteleiros e o número de estudantes internacionais inscritos nas universidades da cidade.

De um modo semelhante, Atuesta & Hewings (2019) identificam as áreas gentrificadas de Chicago, ao analisar variáveis como os rendimentos, as licenças de construção, a etnia da população residente, o total de alojamentos ocupados pelo arrendatário, o total de alojamentos familiares vagos e a percentagem de população residente, maior de 16 anos, com licenciatura ou curso profissional.

Neste sentido, variáveis como os rendimentos familiares, a densidade populacional, a proporção de habitações ocupadas pelo proprietário e o número de residentes estrangeiros poderão ser pertinentes para o desenvolvimento do modelo hedónico no âmbito da gentrificação.

3. DADOS E METODOLOGIA

Com base nos conceitos e abordagens discutidos na revisão de literatura, este capítulo apresenta a metodologia adotada para analisar a distribuição espacial dos preços do arrendamento de habitação na Área Metropolitana de Lisboa e os fatores que influenciam a sua variação. Para tal, primeiramente, é apresentada a área de estudo (3.1), através de uma breve contextualização das suas características geográficas e socioeconómicas. De seguida, é explicado todo o processo metodológico (3.2), onde são apresentados os dados utilizados e descritas as técnicas de análise espacial empregues ao longo do estudo, inclusive os métodos usados para calibrar, diagnosticar e validar o modelo MGWR.

3.1. ÁREA DE ESTUDO

A Área Metropolitana de Lisboa (AML), representada na Figura 3.1.1, constitui a principal região urbana de Portugal, desempenhando, portanto, um papel central na economia, e na dinâmica social do país. Apesar de abranger apenas cerca de 3% do território nacional, a AML alberga, segundo os dados dos censos de 2021⁶, cerca de 2 870 208 habitantes, o que equivale a aproximadamente 28% da população residente no país. Esta concentração populacional reflete, igualmente, a sua importância estratégica enquanto motor económico e principal polo de emprego do país, tendo sido responsável por 36% do PIB nacional em 2016 (CCDRLVT & AML, 2020).

Administrativamente, esta região é composta por 18 municípios, apresentando uma redução no número de freguesias de 207 em 2001, para 118 em 2021, facto que evidencia, desde logo, as mudanças na sua organização territorial ao longo do tempo.

Geograficamente, a AML é atravessada pelo rio Tejo, que a divide em duas sub-regiões principais – Grande Lisboa, a Norte, e Península de Setúbal, a Sul. Estas duas áreas apresentam realidades distintas, onde se evidenciam desigualdades sociais, económicas e demográficas que, adicionadas à estrutura urbana historicamente centralizada em torno da cidade de Lisboa, moldam o território e as suas dinâmicas.

Estas assimetrias são também visíveis internamente nos padrões de consumo e despesas dos agregados familiares. De acordo com o Inquérito às Despesas das Famílias de 2022/2023 (Instituto Nacional de Estatística, 2024), a habitação, água, eletricidade, gás e outros combustíveis representam 39,3% da despesa total anual média por agregado familiar em Portugal, valor que aumenta para 43,8% quando se considera apenas a AML. Estes dados sublinham o impacto desproporcional do mercado habitacional na região, tornando-a um caso de

⁶ Consultado no dia 27 de março de 2025, no *website* oficial de resultados dos censos de 2021, do Instituto Nacional de Estatística: <https://tabulador.ine.pt/indicador/?id=0011609>.

estudo essencial para analisar a relação entre valorização imobiliária, acesso à habitação e desigualdades territoriais.

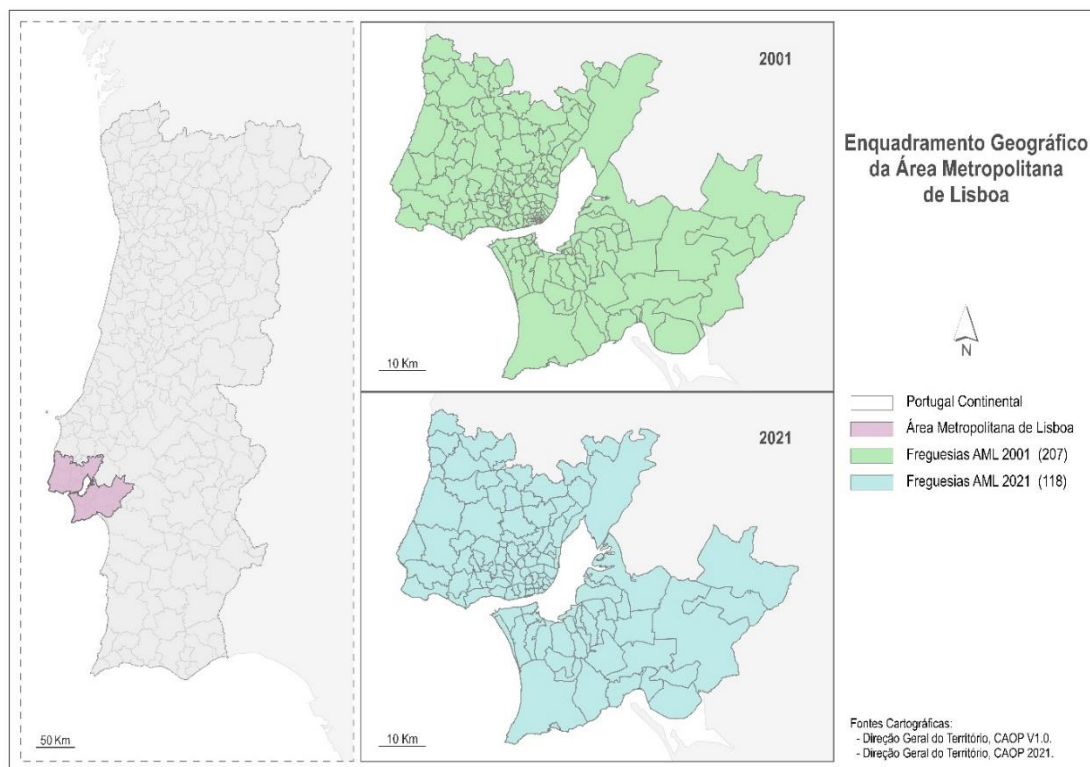


Figura 3.1.1 Enquadramento geográfico da Área Metropolitana de Lisboa

3.2. METODOLOGIA

Tal como já manifestado, a presente investigação tem como objetivo analisar a distribuição espacial dos preços do arrendamento de habitação na AML e identificar os fatores que influenciam a sua variação, com especial atenção ao fenómeno da gentrificação. Para tal, a metodologia adotada segue um processo estruturado, esquematizado na Figura 3.2.1, que será relatado nos próximos subcapítulos.

Importa, antes de mais, destacar que, com vista aos objetivos estipulados, foram realizados dois estudos diferentes, um para o ano de 2001 e outro para 2021. Embora estes dois estudos usem algumas variáveis distintas, seguem exatamente o mesmo processo metodológico, sendo, numa fase final deste relatório, devidamente analisados e contrapostos.

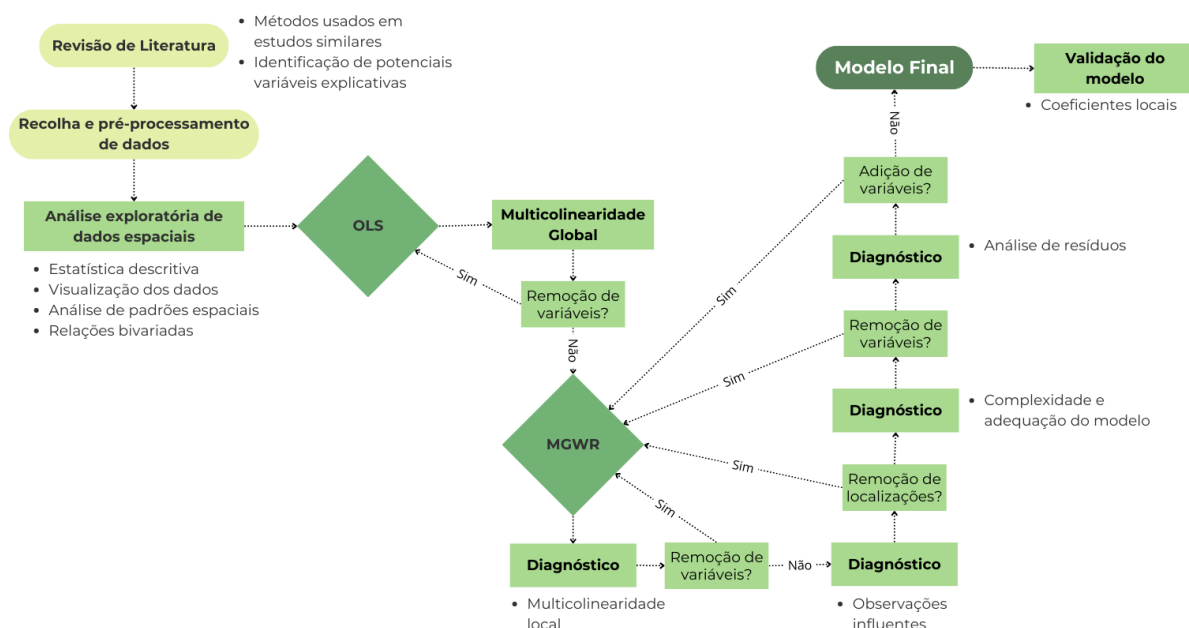


Figura 3.2.1 Esquema do processo metodológico

3.2.1. DADOS

Para este estudo, foram utilizados dados referentes aos anos de 2001 e 2021, provenientes essencialmente dos censos da população e habitação dos respetivos anos, disponíveis de forma aberta no *website* do Instituto Nacional de Estatística (INE).

De modo a permitir a comparação entre os modelos dos dois anos em estudo, foi usada a mesma variável dependente para ambos - valor médio mensal das rendas nos alojamentos familiares arrendados - uma variável quantitativa contínua, cuja unidade de medida é o euro (€).

Tal como descrito na Tabela 3.2.1, a recolha das potenciais variáveis explicativas apoiou-se na revisão de literatura, tendo estas sido distinguidas consoante a tipologia de fatores que influenciam os preços da habitação. Tal como Finio (2022) sugere, a gentrificação pode ser identificada a partir de um conjunto de variáveis que se incluem nas três categorias: características estruturais, como a idade dos edifícios ou o número de divisões; características da vizinhança, tais como a idade e escolaridade da população residente, a taxa de desemprego, ou o regime de propriedade dos alojamentos; e atributos de localização, nomeadamente a distância às infraestruturas de transporte. Por este motivo, as variáveis selecionadas para este estudo visaram não somente retratar os fatores determinantes dos preços das rendas, como também as transformações socioeconómicas e urbanas decorrentes da gentrificação. Não obstante, importa fisar que esta classificação não é estática, isto é, uma mesma variável poderia ser classificada em diferentes categorias, de acordo com interpretações pessoais ou objetivos específicos. Esta divisão visa simplesmente facilitar a interpretação dos resultados, sendo necessário considerar as variáveis conjuntamente, por forma a obter conhecimento sobre os fenómenos que afetam os preços das rendas.

Tabela 3.2.1 Descrição das potenciais variáveis explicativas

Categoria	Variável	Unidade de medida	Fonte(s)	Ano(s)
Características Estruturais	Proporção de alojamentos familiares com aquecimento central	%	INE, Censos	2001, 2021
	Divisões por alojamento familiar clássico de residência habitual	N.º	INE, Censos	2001, 2021
	Proporção de alojamentos familiares clássicos de residência habitual construídos antes de 1960	%	INE, Censos	2001, 2021
	Proporção de edifícios construídos nos últimos 10 anos	%	INE, Censos	2001, 2021
Características de Vizinhança	Densidade de alojamentos	N.º/ km ²	INE, Censos	2001, 2021
	Densidade populacional	N.º/ km ²	INE, Censos	2001, 2021
	Taxa de desemprego	%	INE, Censos	2001, 2021
	Proporção da população residente com ensino superior completo	%	INE, Censos	2001, 2021
	Proporção de alojamentos superlotados	%	INE, Censos	2001
	Percentagem de alojamentos familiares clássicos de residência habitual propriedade dos ocupantes	%	INE, Censos	2001, 2021
	Proporção de alojamentos familiares clássicos de uso sazonal	%	INE, Censos	2001, 2021
	Proporção da população residente de nacionalidade estrangeira	%	INE, Censos	2001, 2021
	Proporção da população residente com 65 ou mais anos	%	INE, Censos	2001, 2021
	Proporção de alojamentos familiares clássicos vagos	%	INE, Censos	2001, 2021
	Rendimento bruto declarado por agregado fiscal	€	INE, Censos	2021
Atributos de Localização	Duração média dos movimentos pendulares da população residente empregada ou estudante	minutos	INE, Censos	2001, 2021
	Distância à classe da COS "Parques e jardins" mais próxima	metros	DGT, COS	2007, 2018
	Distância à classe da COS "Praias, dunas e areais costeiros" mais próxima	metros	DGT, COS	2007, 2018
	Distância à classe da COS "Rede viária e espaços associados" mais próxima	metros	DGT, COS	2007, 2018

Importa referir, ainda, que foram criadas três variáveis de localização, em ambiente SIG, a partir do cálculo da distância euclidiana entre os centroides de cada freguesia e o polígono mais próximo da classe da Carta de Uso e Ocupação do Solo (COS):

- Distância à classe da COS "Parques e jardins" mais próxima;
- Distância à classe da COS "Praias, dunas e areais costeiros" mais próxima;
- Distância à classe da COS "Rede viária e espaços associados" mais próxima.

A COS é um produto cartográfico de referência nacional disponibilizado, em formato vetorial, pela Direção-Geral do Território (DGT), através do Sistema Nacional de Informação Geográfica. Dada a impossibilidade de obter a COS para os anos exatos em estudo, foram descarregadas as duas versões temporalmente mais próximas dos anos pretendidos (a COS 2007 para o ano de 2001 e a COS 2018 para o ano de 2021), partindo do pressuposto que, embora algumas áreas sofram mudanças rápidas na ocupação do solo, grande parte do território não

apresenta alterações significativas do uso do solo em curtos períodos. Adicionalmente, de acordo com as especificações técnicas (Direção-Geral do Território, 2019), estas duas versões da COS baseiam-se na mesma nomenclatura, o que permite uma correta comparação entre os dois anos em estudo, mantendo-se a consistência temática das classes de uso do solo.

3.2.2. RECOLHA E PRÉ-PROCESSAMENTO DOS DADOS

Após a revisão de literatura, que fundamentou a seleção das variáveis a considerar e os métodos a utilizar, a primeira etapa deste processo metodológico consistiu na recolha dos dados acima descritos e o respetivo pré-processamento. Esta etapa, realizada em ambiente *Microsoft Excel*, pressupôs não só a organização e compilação de todos os dados, como também a sua transformação para unidades de medida comparáveis (percentagem, sempre que possível).

Por outro lado, em ambiente *ArcGIS Pro*, procedeu-se à projeção de todas as *layers* para o sistema de coordenadas ETRS 1989 Portugal TM06, garantindo a coerência espacial entre os diferentes conjuntos de dados utilizados.

Como base cartográfica, foi utilizada a Carta Administrativa Oficial de Portugal (CAOP), recorrendo-se à versão v1.0 para 2001 e à versão CAOP 2021 para o ano de 2021, disponibilizadas abertamente pela DGT. Neste sentido, de modo a associar as bases cartográficas (CAOP) às respetivas tabelas de dados, foi usada a ferramenta *join* do *ArcGIS Pro*, fazendo-se a conexão a partir das colunas que continham os códigos identificativos das freguesias (DICOFRE).

3.2.3. ANÁLISE EXPLORATÓRIA DOS DADOS ESPACIAIS

Tal como Haining (2003) refere, os objetivos primordiais da análise exploratória de dados espaciais (ESDA) consistem em reconhecer as características dos dados, detetar padrões espaciais, identificar anomalias e distinguir elementos acidentais de propriedades importantes no conjunto de dados. Para tal, a ESDA faz uso de técnicas e ferramentas usadas em qualquer análise exploratória de dados (EDA), adicionando-lhes outros métodos adaptados às particularidades dos dados espaciais.

Neste estudo, a ESDA teve início com a caracterização das variáveis em estudo, através do cálculo de estatísticas descritivas, como a média, a mediana e o desvio padrão. Para complementar esta análise, procedeu-se à representação gráfica da distribuição dos dados por meio de histogramas, o que permitiu identificar assimetrias e possíveis *outliers*.

De seguida, a visualização dos dados foi realizada através da criação de cartografia temática. Este processo de visualização dos dados por meio de mapas tem-se revelado uma técnica cada vez mais relevante na análise espacial, beneficiando dos avanços tecnológicos, particularmente no domínio dos SIG (Bao & Anselin, 1999). A crescente sofisticação destas ferramentas permite não apenas uma representação rápida e versátil dos dados, mas também a exploração de múltiplas abordagens de visualização, promovendo a interpretação mais intuitiva dos padrões espaciais e das estatísticas associadas.

No presente estudo, o mapeamento dos dados foi efetuado com recurso ao método de classificação por quantis, mais especificamente quintis, assegurando uma distribuição equilibrada das observações em cinco classes. Este método revelou-se particularmente adequado aos objetivos da ESDA, pois permitiu destacar as variações espaciais dos dados, facilitando a identificação de áreas com valores mais elevados ou mais reduzidos. Além disso, a segmentação equitativa dos dados viabilizou comparações diretas entre diferentes regiões, tornando mais evidentes os contrastes espaciais.

De modo a consolidar esta ESDA, recorreu-se, numa etapa seguinte, a ferramentas de análise de padrões espaciais globais (*Global Moran's I*) e locais (*Local Moran's I* e *Getis-Ord Gi statistic*) para testar a existência de autocorrelação espacial e dependência espacial no conjunto de dados.

O teste de *Global Moran's I* tem como objetivo identificar a existência de padrões de concentração ou dispersão dos dados, ao testar a hipótese nula da sua aleatoriedade espacial. Este índice geralmente varia de -1 a 1, sendo que: valores positivos indicam que valores similares tendem a agrupar-se espacialmente (*clusters*); valores negativos sugerem um padrão de dispersão, isto é, valores dissimilares tendem a localizar-se próximos entre si; e um índice de aproximadamente zero indica uma possível não rejeição da hipótese nula, ou seja, os valores tendem a distribuir-se aleatoriamente no espaço (Longley et al., 2015).

Enquanto o teste de *Global Moran's I* apresenta uma medida geral de autocorrelação espacial para toda a área de estudo, a estatística de *Local Moran's I* identifica padrões locais de associação espacial, evidenciando a presença de *clusters* e *outliers*. O mapa resultante exhibe as freguesias que têm uma relação estatisticamente significativa com as vizinhas e o tipo de relação existente. Um padrão local de autocorrelação espacial positiva pode surgir sob a forma de dois tipos de *clusters* locais: *High-high cluster*, quando valores elevados se encontram próximos de valores elevados; ou *Low-low cluster*, quando valores reduzidos se encontram agrupados. Adicionalmente, os padrões locais de autocorrelação espacial negativa evidenciam-se sob a forma de *outliers*: *High-low outlier*, quando um valor elevado se apresenta próximo de valores reduzidos; e *Low-high outlier*, quando um valor reduzido se encontra próximo de valores elevados.

De forma semelhante, a estatística *Getis-Ord Gi* permite testar a não-estacionaridade dos dados, ao identificar padrões espaciais de concentração de valores elevados e reduzidos. Para tal, o mapa obtido representa a localização de agrupamentos de valores elevados (*hot-spots*) e de valores reduzidos (*cold-spots*) estatisticamente significantes, distinguindo-os conforme o nível de confiança (90%, 95%, ou 99%).

Analizados os padrões espaciais, segue-se o estudo das relações bivariadas. Esta subsecção pretende averiguar quais as variáveis que melhor poderão explicar a distribuição dos valores médios das rendas nas freguesias da AML e, simultaneamente, evitar a seleção de variáveis redundantes passíveis de provocar multicolinearidade no modelo. Deste modo, recorreu-se, inicialmente, a gráficos de dispersão e matrizes de correlação de Pearson para quantificar a intensidade e sentido das relações lineares globais entre pares de variáveis, em todo o conjunto de dados. O coeficiente de correlação de Pearson, varia entre -1 e 1, sendo que valores positivos indicam a existência

de uma relação linear direta, valores negativos apontam para uma relação inversa e valores próximos de 0 revelam ausência de associação linear (Sampaio et al., 2010).

Sincronicamente, foram analisadas as relações bivariadas locais entre a variável dependente e cada um dos possíveis fatores explicativos, recorrendo à ferramenta do *ArcGIS Pro*, *Local Bivariate Relationships*. Esta abordagem complementa a análise das relações bivariadas globais, na medida em que mostra como tais relações se podem manifestar de forma heterogênea em diferentes regiões e, paralelamente, mapeá-los com base nos tipos de relações evidentes localmente, podendo estes ser:

- Não-significante (*Not significant*): Indica que não há uma relação estatisticamente significativa entre as duas variáveis na área analisada;
- Linear positiva (*Positive linear*): Representa uma relação direta, isto é, quando a variável explicativa aumenta, a variável dependente também tende a aumentar;
- Linear negativa (*Negative linear*): Revela uma relação inversa entre as variáveis, ou seja, quando uma aumenta a outra tende a diminuir;
- Côncava (*Concave*): Indica uma relação não linear, em que à medida que a variável explicativa aumenta, a variável dependente varia através de uma curva côncava (em formato “U”);
- Convexa (*Convex*): Indica uma relação não linear, em que à medida que a variável explicativa aumenta, a variável dependente varia através de uma curva convexa (em formato “∩”);
- Complexa indefinida (*Undefined complex*): Indica que as variáveis têm uma relação estatisticamente significativa, mas esta não pode ser descrita, de forma fiável, por nenhuma das outras tipologias.

3.2.4. MULTICOLINEARIDADE GLOBAL

Selecionadas as variáveis a usar em cada modelo, a etapa seguinte passa por averiguar se existirá multicolinearidade global, já que o modelo MGWR poderá não prosseguir caso a multicolinearidade entre as variáveis seja demasiado elevada.

Para este efeito, foi executada uma regressão linear exploratória, através do modelo OLS, sendo que a medida utilizada para averiguar a existência de multicolinearidade global foi o Fator de Inflação de Variância, ou em inglês *Variance Inflation Factor* (VIF). O VIF mede o quanto a variância do coeficiente estimado de uma variável independente é inflacionada devido à correlação com as demais variáveis explicativas do modelo de regressão linear. Por outras palavras, quando uma variável apresenta um VIF elevado, regra geral $VIF > 7,5$, significa que grande parte da sua variabilidade pode ser explicada pelas outras variáveis do modelo, o que será indicativo de redundância. Idealmente, um $VIF < 4$ revela um nível baixo de multicolinearidade (O’Brien, 2007). O método mais comum usado para evitar a multicolinearidade passa por remover, uma a uma, as variáveis explicativas com $VIF > 7,5$, analisando sequencialmente a reação provocada nas restantes, até que todas as variáveis explicativas não eliminadas obtenham um $VIF < 7,5$.

3.2.5. MODELO MGWR

A última grande etapa deste processo metodológico consiste na aplicação, calibração e validação do modelo MGWR. A Regressão Geograficamente Ponderada Multiescala (MGWR) é uma extensão da Regressão Geograficamente Ponderada (GWR), comumente utilizada para explorar a heterogeneidade espacial e modelar processos locais (Fotheringham et al., 2024). A sua característica distintiva consiste no facto de, além de permitir que os coeficientes variem no espaço, assegura, ainda, que cada variável explicativa opere em diferentes escalas geográficas, através da utilização de bandas de largura (i.e., vizinhanças) específicas para cada variável. A calibração do modelo MGWR segue uma série de etapas de diagnóstico que visam a identificação adequada das variáveis explicativas, a minimização da multicolinearidade e a validação dos coeficientes locais.

Primeiramente, verificam-se as larguras de banda individuais e as estatísticas dos coeficientes locais de cada variável explicativa. As larguras de banda ótimas (i.e., *optimal bandwidth*) são identificadas, aquando da aplicação do modelo, através do algoritmo *Golden Search*, que visa encontrar o número de vizinhos mais próximos (i.e., freguesias) ideal, estabelecendo um equilíbrio entre a incerteza (erro-padrão) e o enviesamento dos parâmetros (Fotheringham et al., 2024). Assim, as variáveis explicativas cuja relação com o valor médio mensal das rendas seja relativamente homogénea no espaço terão larguras de banda maiores, enquanto processos com elevada heterogeneidade espacial requerem vizinhanças menores. Uma largura de banda próxima do máximo (100% das freguesias) resulta em estimativas mais próximas de um modelo global e larguras de banda excessivamente reduzidas poderão alertar para a modelação de padrões aleatórios, ao invés de padrões espaciais significativos.

Voltando a frisar a importância da minimização da multicolinearidade, o passo seguinte consiste em testar a multicolinearidade local através do mapeamento do número de condição, em inglês *condition number* (CN), para cada um dos fatores explicativos. O CN é um método que mede a estabilidade numérica da matriz de variáveis explicativas em cada local, identificando regiões onde a multicolinearidade poderá estar a comprometer a robustez das estimativas dos coeficientes, sendo que, regra geral, um $CN > 30$ alerta para a existência de multicolinearidade local (Fotheringham et al., 2024). Neste caso, uma das três ações seguintes deve ser considerada:

- Executar novamente o modelo utilizando um maior número mínimo de vizinhos;
- Remover a(s) variável(eis) que revela(m) agrupamento espacial de valores idênticos;
- Combinar essa(s) variável(eis) com outras variáveis explicativas para aumentar a variação do valor.

De seguida, deve-se verificar a existência de observações influentes, isto é, observações cujos valores exercem um impacto desproporcional sobre as estimativas dos coeficientes locais do modelo. A medida usada para determinar a existência de tais observações é, entre outras, a distância de Cook (em inglês, *Cook's Distance* ou *Cook's D*), que mede o impacto que a remoção de uma observação teria sobre os coeficientes estimados do modelo. Assim, ao mapear os valores da distância de Cook, poderão ser identificadas observações influentes, ou seja, observações em que $Cook's D > 1$. Uma das possíveis soluções para lidar com observações influentes poderá passar pela sua remoção. No entanto, neste caso, remover uma observação implicaria remover uma freguesia,

o que provoca perda de informação e descontinuidade espacial na análise. Outra opção passa por simplesmente ajustar as larguras de banda do modelo MGWR, o que poderá reduzir o impacto da observação sem excluí-la. Não obstante, uma observação influente não necessariamente deverá ser removida, podendo-se optar por retificar a qualidade dos dados e anotar possíveis descobertas relacionadas com tal observação, sem ser necessária qualquer ação de maior calibre.

Entrando numa última fase de diagnóstico do modelo MGWR, avaliar a sua complexidade torna-se essencial para garantir que este representa adequadamente a variabilidade espacial sem comprometer sua robustez e interpretabilidade. Modelos excessivamente complexos tendem a descrever o erro aleatório nos dados ao invés de detetarem os padrões reais das relações entre as variáveis, causando um sobreajuste (*overfitting*) do modelo. Por outro lado, um modelo mais simples será mais fácil de interpretar e terá maior adaptabilidade para diferentes conjuntos de dados, isto é, terá melhor desempenho preditivo quando aplicado a diferentes períodos ou regiões.

A complexidade do modelo pode ser avaliada pelo número de variáveis explicativas, o Número Efetivo de Parâmetros (ENP) e os Graus de Liberdade Efetivos (*Effective Degrees of Freedom*). Se ao remover a variável considerada menos relevante (de acordo com a revisão de literatura ou com base na significância dos coeficientes), a complexidade do modelo (i.e., ENP) decrescer significativamente e não se registar um grande impacto negativo no desempenho do modelo, então deverá optar-se pelo modelo com o menor número de variáveis explicativas. Deste modo, em simultâneo, estar-se-á a analisar a adequação do modelo, medida essencialmente pelo *Corrected Akaike's Information Criterion* (AICc) e pelo coeficiente de determinação ajustado (Adj-R^2) (Ma & Gopal, 2018), sendo que modelos com valores de AICc menores indicam melhor equilíbrio entre desempenho e complexidade, e valores mais elevados de Adj-R^2 revelam um maior ajuste do modelo. Importa, no entanto, frisar que ambos os modelos MGWR finais aqui desenvolvidos incluíram menos de 10 variáveis independentes, pelo que não foi investigado o impacto da remoção de variáveis no desempenho dos modelos.

Ainda nesta última fase de verificação da adequação do modelo, deve-se testar a aleatoriedade espacial dos resíduos, já que a existência de autocorrelação espacial estatisticamente significativa poderá indicar que o modelo estará especificado incorretamente. Para tal, recorre-se novamente à estatística de *Global Moran's I*.

Em última instância, de modo a interpretar os resultados, isto é, analisar os efeitos locais de cada variável nos valores médios das rendas nos alojamentos familiares arrendados, são analisados e mapeados os coeficientes locais do modelo final.

4. RESULTADOS DO ESTUDO DE 2001

Neste capítulo são apresentados os resultados obtidos na análise dos preços do arrendamento de habitação nas freguesias da Área Metropolitana de Lisboa, em 2001. Para tal, a variável dependente considerada foi o valor médio mensal das rendas dos alojamentos familiares clássicos arrendados, registados no ano censitário em causa.

Este estudo inicia-se com uma análise exploratória de dados espaciais (4.1) que, com recurso a métodos de estatística descritiva e visualização dos dados, permite reconhecer previamente anomalias ou *outliers*, bem como algumas tendências e padrões que poderão auxiliar na seleção dos métodos e técnicas mais adequados às características dos dados. De seguida, por forma a testar a multicolinearidade global dos dados (4.2), recorre-se ao modelo OLS, auxiliando, assim, o processo de seleção das variáveis explicativas. Por fim, é descrito o processo de aplicação do modelo MGWR (4.3), desde a sua calibração e diagnóstico, até à sua validação, culminando na análise dos coeficientes locais, que permitem analisar a espacialidade dos fatores que, em 2001, influenciaram os preços das rendas.

4.1. ANÁLISE EXPLORATÓRIA DOS DADOS ESPACIAIS

A análise exploratória dos dados é o primeiro passo fundamental em qualquer trabalho de análise de dados. Por meio de métodos como a estatística descritiva (4.1.1), a visualização dos dados (4.1.2), a análise de padrões espaciais (4.1.3) e a análise das relações entre variáveis (4.1.4), será possível conhecer os dados e, concomitantemente, decidir sobre os métodos, técnicas e variáveis explicativas mais apropriados para o desenvolvimento do modelo de regressão sobre os preços das rendas em 2001. Para tal, foram consideradas as 18 potenciais variáveis explicativas descritas na Tabela 4.1.1.

Tabela 4.1.1 Descrição das potenciais variáveis explicativas consideradas para 2001

Variável	Descrição
2001_DENSALOJ	Densidade de alojamentos (N.º/ km²)
2001_PAFPO	Proporção de alojamentos familiares clássicos de residência habitual propriedade dos ocupantes (%)
2001_PAF_ANTES1960	Proporção de alojamentos familiares clássicos de residência habitual construídos antes de 1960 (%)
2001_PAFSAZ	Proporção de alojamentos familiares clássicos de uso sazonal (%)
2001_PAFV	Proporção de alojamentos familiares clássicos vagos (%)
2001_PASUPERLOT	Proporção de alojamentos superlotados (%)
2001_PAFQCENT	Proporção de alojamentos familiares com aquecimento central (%)
2001_NDIV	Divisões por alojamento familiar clássico de residência habitual (N.º)
2001_PED_10ANOS	Proporção de edifícios construídos nos últimos 10 anos (%)
2001_DENSPPOP	Densidade populacional (N.º/ km²)
2001_PPOPRES_65MAIS	Proporção da população residente com 65 ou mais anos (%)
2001_PPOPRESNE	Proporção da população residente de nacionalidade estrangeira (%)
2001_PPOPRESSESSUP	Proporção da população residente com ensino superior completo (%)
2001_DURMOVPEND	Duração média dos movimentos pendulares da população residente empregada ou estudante (min)
2001_TX_DESEMP	Taxa de desemprego (%)
NEAR_DIST_COS2007_PARQUES_JARDINS	Distância à classe da COS "Parques e Jardins" mais próxima (m)
NEAR_DIST_COS2007_PRAIAS	Distância à classe da COS "Praias, Dunas e Areais Costeiros" mais próxima (m)
NEAR_DIST_COS2007_REDE_VIARIA	Distância à classe da COS "Rede Viária e Espaços Associados" mais próxima (m)

4.1.1. ESTATÍSTICA DESCRITIVA

O cálculo das medidas de tendência central e de dispersão para o valor médio mensal das rendas, nas 207 freguesias que compunham a AML em 2001, bem como a sua representação gráfica (Figura 4.1.1), mostraram que a distribuição da variável dependente é assimétrica positiva, uma vez que a média (128,23€) é ligeiramente superior à mediana (121€).

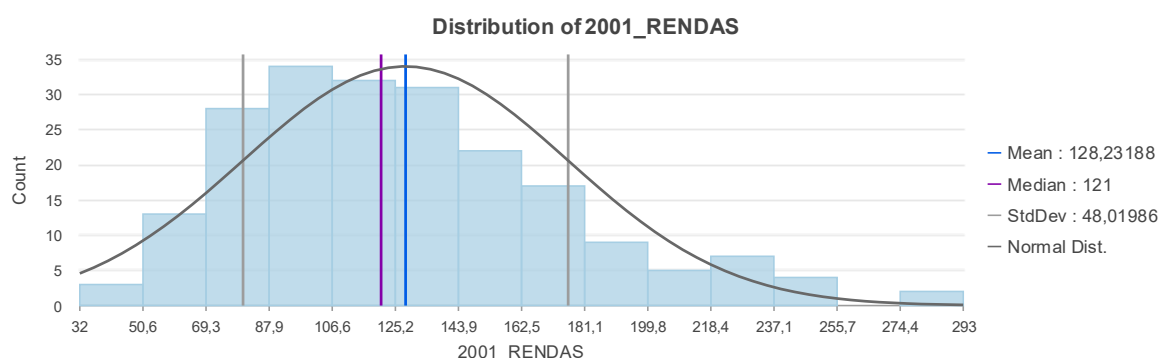


Figura 4.1.1 Histograma da distribuição do valor médio mensal das rendas dos alojamentos familiares clássicos arrendados, em 2001

Além disso, observa-se uma grande variabilidade dos valores médios mensais das rendas na AML, já que o valor mínimo registado em 2001 era de 32€ na freguesia de Sarilhos Pequenos, no concelho da Moita, enquanto o valor máximo foi de 293€, na freguesia de Alforneiros, concelho da Amadora. Esta variabilidade é, também, atestada pelo valor do desvio-padrão de 48,02.

Os histogramas referentes à distribuição dos vários possíveis fatores explicativos considerados estão representados no Apêndice A: Histogramas da Distribuição dos Potenciais Fatores Explicativos de 2001. Importa destacar que, de um modo geral, são revelados os seguintes padrões de distribuição:

- Distribuição normal (média $\approx$ mediana): 2001_NDIV, 2001_TX_DESEMP;
- Distribuição assimétrica positiva (média $>$ mediana): 2001_DENSALoj, 2001_PAF_ANTES1960, 2001_PAFSAZ, 2001_PAFV, 2001_PASUPERLOT, 2001_PAFQCENT, 2001_PED_10ANOS, 2001_DENSPOP, 2001_PPOPRES_65MAIS, 2001_PPOPRESNE, 2001_PPOPRESSENSUP, 2001_DURMOVPEND, NEAR_DIST_COS2007_PARQUES_JARDINS, NEAR_DIST_COS2007_PRAIAS, NEAR_DIST_COS2007_REDE_VIARIA;
- Distribuição assimétrica negativa (média $<$ mediana): 2001_PAFPO.

4.1.2. VISUALIZAÇÃO DOS DADOS

Como verificado anteriormente, os dados da variável dependente não seguem uma distribuição absolutamente normal, pelo que, para promover uma leitura mais intuitiva das diferenças espaciais nos preços do arrendamento, procedeu-se ao seu mapeamento (Figura 4.1.2), a partir do método de classificação por quantis, neste caso organizado em quintis (5 classes com aproximadamente o mesmo número de freguesias em cada uma).

À primeira vista, não é possível identificar um padrão claro de tendência de distribuição espacial do valor médio mensal das rendas nos alojamentos familiares clássicos arrendados, em 2001. No entanto, é possível distinguir um conjunto de valores mais reduzidos a Este e Sudeste do município central, bem como uma tendência de preços de arrendamento mais elevados na costa ocidental da AML, especialmente nos municípios da margem Norte, Oeiras, Cascais, Sintra e Mafra. De uma forma algo rudimentar, pode-se identificar uma tendência de concentração de valores mais reduzidos a Este e valores mais elevados a Oeste, embora se possam identificar freguesias que refutam esta afirmação, nomeadamente a freguesia de Poceirão, localizada na secção Sudeste da AML, no município de Palmela, que se inclui no quintil de preços de arrendamento mais elevados.

Os mapas temáticos desenvolvidos para visualizar a distribuição das potenciais variáveis explicativas encontram-se no Apêndice B: Mapas Temáticos dos Potenciais Fatores Explicativos.

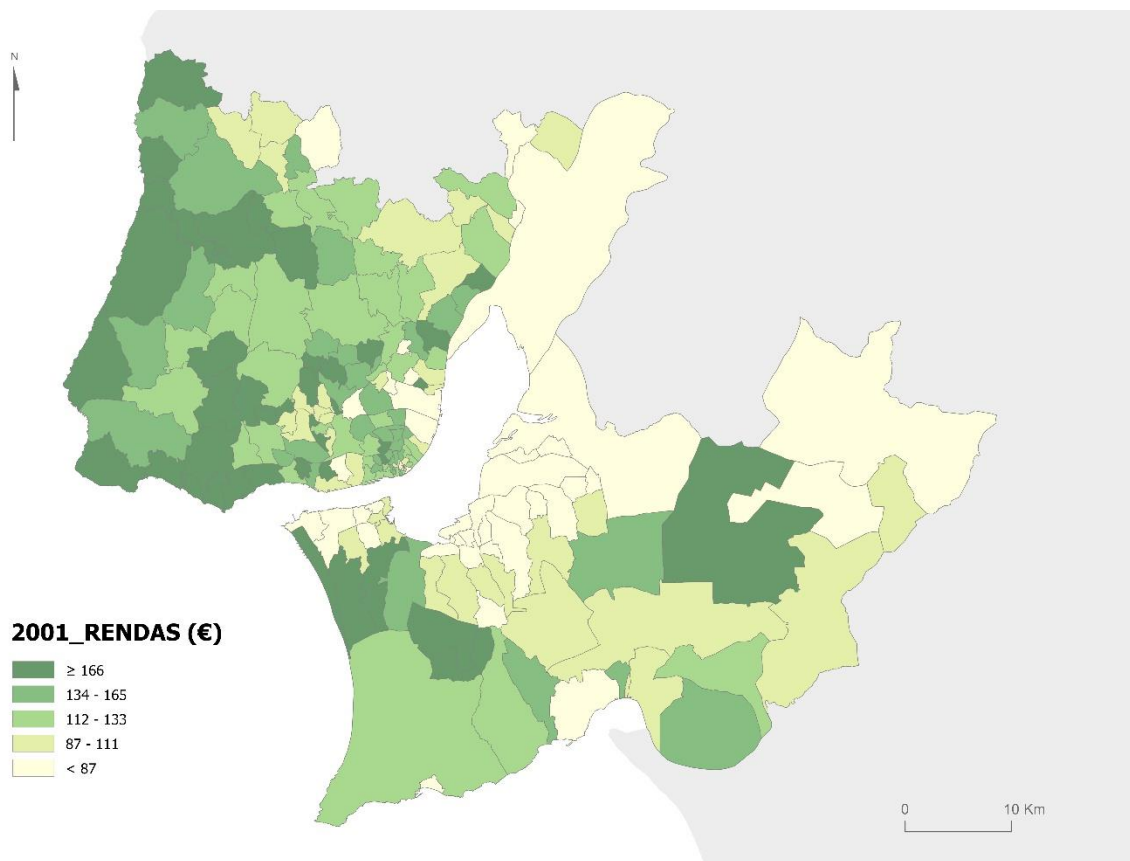


Figura 4.1.2 Valor médio mensal das rendas nos alojamentos familiares clássicos arrendados, em 2001

4.1.3. ANÁLISE DE PADRÕES ESPACIAIS

Para complementar esta análise exploratória, recorreu-se ao teste de *Global Moran's I* com o objetivo de analisar a dependência espacial da variável em estudo e, por forma a identificar *clusters* e *outliers*, foram utilizadas ferramentas de análise de padrões espaciais locais: *Local Moran's I* e *Getis-Ord Gi statistic*. Neste caso, o teste de *Global Moran's I*, atesta a existência de autocorrelação espacial positiva, com o valor de 0,238470. O valor de p (0,000000), inferior ao nível de significância de 0,05, indica que os valores da variável dependente tendem a apresentar-se em agrupamentos (*clusters*), sendo estatisticamente significativa a existência de autocorrelação espacial. Por outras palavras, os valores médios mensais das rendas tendem a ser semelhantes nas freguesias vizinhas. Deste modo, o uso de um modelo de regressão espacial, como MGWR e GWR é imperativo, já que o tradicional modelo de regressão linear OLS não atende aos efeitos da dependência espacial.

Por outro lado, os testes de *Local Moran's I* e *Getis-Ord Gi* (Figura 4.1.3) revelam que a variável dependente apresenta não-estacionaridade espacial, isto é, os padrões e processos espaciais dos preços de arrendamento não são constantes em toda a área de estudo. De acordo com os resultados do teste de *Local Moran's I*: 24 freguesias apresentam-se como clusters de valores reduzidos, localizados essencialmente nos municípios do Montijo, Moita e Barreiro; 17 freguesias são classificadas como clusters de valores elevados, principalmente localizadas em Cascais, Sintra, Amadora e Sul de Mafra; e 9 freguesias, em Oeiras, Sintra e Amadora, são identificadas como *outliers*, cujos valores reduzidos correlacionam-se com os valores elevados das freguesias vizinhas (*Low-High Outlier*).

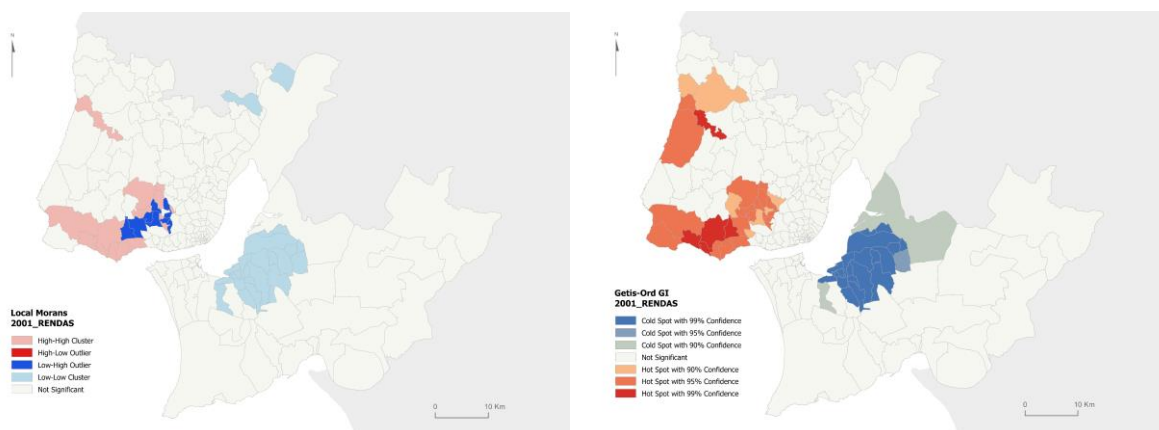


Figura 4.1.3 Local Moran's I e Getis-Ord Gi da variável 2001_RENDAS

As estatísticas de *Getis-Ord Gi* corroboram os mesmos padrões espaciais. Verifica-se um total de 33 freguesias identificadas como *clusters* de valores elevados (*hot spots*), das quais, 5 com 99% de confiança, 21 com 95% de confiança, e 7 com 90% de confiança, e um total de 23 freguesias consideradas *clusters* de valores reduzidos (*cold spots*), 20 das quais com 99% de confiança, 1 com 95% de confiança e 2 com 90% de confiança.

4.1.4. RELAÇÕES BIVARIADAS

A parte inferior esquerda da Figura 4.1.4 mostra os gráficos de dispersão e correspondentes retas de regressão linear, que permitem verificar a existência e intensidade das relações globais entre a variável dependente (2001_RENDAS) e 18 potenciais variáveis explicativas. O lado superior direito mostra os coeficientes de correlação de Pearson, adicionando informação sobre o sentido da relação entre as variáveis.

De acordo com estes coeficientes de regressão linear simples, o valor médio mensal das rendas em 2001 (2001_RENDAS) aparenta estar positivamente relacionado com o número de divisões dos alojamentos familiares (2001_NDIV) e com a proporção de população residente com o ensino superior completo (2001_PPOPRESSENSUP). Por outro lado, a variável dependente parece estar negativamente relacionada com a taxa de desemprego (2001_TX_DESEMP), a proporção de alojamentos superlotados (2001_PASUPERLOT), e a proporção de população residente com 65 ou mais anos de idade. Em contrapartida, as rendas mostram-se estar globalmente pouco ou nada relacionadas com a densidade de alojamentos (2001_DENSALOJ), a densidade populacional (2001_DENSPOP), a duração dos movimentos pendulares (2001_DURMOVPEND) ou qualquer uma

das variáveis de localização consideradas (NEAR_DIST_COS2007_REDE_VIÁRIA, NEAR_DIST_COS2007_PRAIAS e NEAR_DIST_COS2007_PARQUES_JARDINS).

A análise da Figura 4.1.4 permite igualmente identificar correlações entre as potenciais variáveis explicativas. Os valores elevados do coeficiente de Pearson entre as variáveis - 2001_DENSALOJ e 2001_DENSPOP; 2001_PPOPRES_65MAIS e 2001_PAF_ANTES1960; 2001_PAF_ANTES1960 e 2001_PAFPO; e 2001_NDIV e 2001_PASUPERLOT -, poderão indicar redundância entre estes pares.

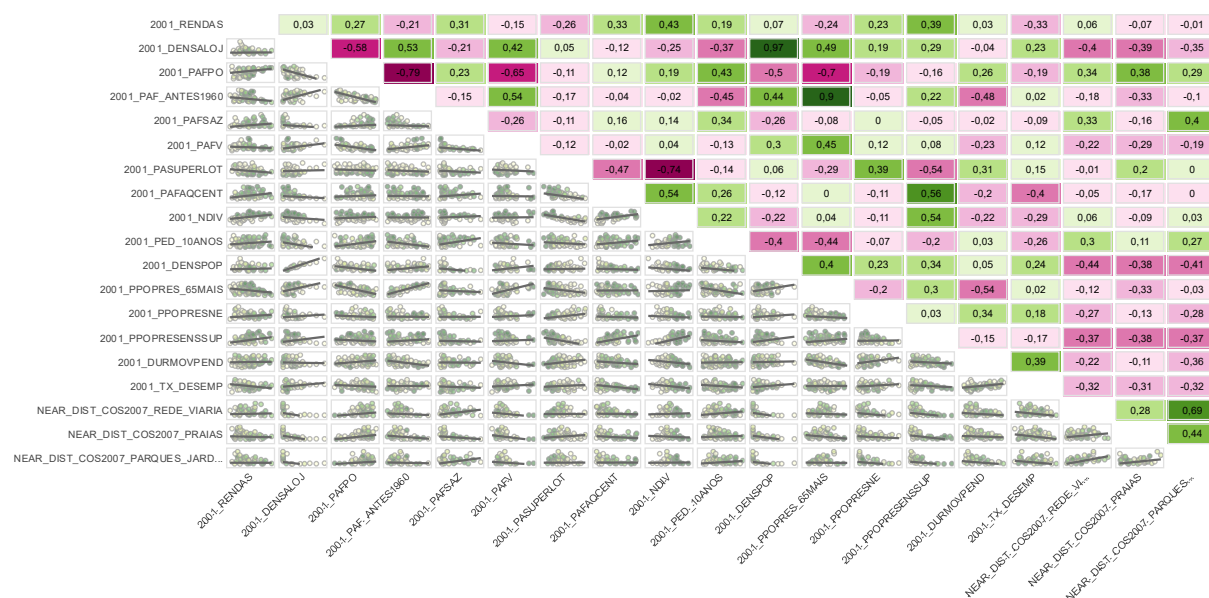


Figura 4.1.4 Gráficos de dispersão e matriz de correlação de Pearson do valor médio mensal das rendas e 18 potenciais variáveis explicativas, dados de 2001

Deste modo, torna-se útil compreender as relações bivariadas locais entre cada uma das potenciais variáveis explicativas e a variável dependente já que, mesmo que uma variável não mostre qualquer relação global com a variável dependente, poderá revelar uma relação linear significativa em determinados locais. Para além disso, esta análise auxilia a seleção das variáveis explicativas, uma vez que apenas aquelas que apresentem relações lineares com o valor médio mensal das rendas devem ser consideradas para o desenvolvimento do modelo de regressão.

A Tabela 4.1.2, que sintetiza as relações bivariadas locais mapeadas no Apêndice C: Relações bivariadas locais significantes entre fatores explicativos e o valor médio mensal das rendas, em 2001, faz notar que, das 18 potenciais variáveis explicativas consideradas apenas 10 apresentam algum tipo de relação linear local com a variável dependente, sendo elas: 2001_PAFPO; 2001_PAF_ANTES1960; 2001_PASUPERLOT; 2001_NDIV; 2001_PED_10ANOS; 2001_PPOPRES_65MAIS; 2001_PPOPRESNSUP; 2001_DURMOVPEND; 2001_TX_DESEMP; NEAR_DIST_COS2007_PRAIAS. As restantes 8 variáveis, que não apresentaram relações significativas locais com a variável dependente, foram excluídas do modelo e qualquer análise posterior não as terá em consideração.

Tabela 4.1.2 Classificação das relações bivariadas locais entre variáveis explicativas e a variável dependente (2001_RENDAS)

Variável	Tipo de relação										Total de freguesias com relação linear significante	
	Positiva linear		Negativa linear		Côncava		Convexa		Complexa			
	N.º	%	N.º	%	N.º	%	N.º	%	N.º	%	N.º	%
2001_PAFPO	62	29,95	4	1,93	19	9,18	70	33,82	4	1,93	66	31,88
2001_PAF_ANTES1960	5	2,42	18	8,70	1	0,48	38	18,36	4	1,93	23	11,11
2001_PASUPERLOT	0	0	60	28,99	0	0	5	2,42	1	0,48	60	28,99
2001_NDIV	62	29,95	0	0	0	0	15	7,25	1	0,48	62	29,95
2001_PED_10ANOS	2	0,97	0	0	0	0	15	7,25	0	0	2	0,97
2001_PPOPRES_65MAIS	0	0	21	10,14	1	0,48	28	13,53	2	0,97	21	10,14
2001_PPOPRESSESSUP	83	40,10	0	0	8	3,86	0	0,00	9	4,35	83	40,10
2001_DURMOVPEND	1	0,48	50	24,15	2	0,97	5	2,42	27	13,04	51	24,64
2001_TX_DESEMP	1	0,48	38	18,36	5	2,42	1	0,48	7	3,38	39	18,84
NEAR_DIST_COS2007_PRAIAS	0	0	7	3,38	0	0	0	0	4	1,93	7	3,38

Embora globalmente a duração média dos movimentos pendulares (2001_DURMOVPEND) não aparentasse relacionar-se com o valor médio mensal das rendas, a nível local identificam-se algumas relações significativas. Na verdade, em 50 freguesias (24,15%), localizadas maioritariamente no município de Lisboa e arredores, regista-se uma relação negativa linear, o que poderá indicar que, nestas localidades, quanto maior a duração dos movimentos pendulares, menor será o preço do arrendamento. Não obstante, em 35 outras freguesias, foram identificados outros tipos de relação significativa entre estas variáveis, das quais: 27 (13,04%) complexa não identificada; 5 (2,42%) convexa; 2 (0,97%) côncava; e 1 (0,48%) positiva linear. Similarmente, a variável que representa a distância euclidiana entre o centroide das freguesias e a classe de uso e ocupação do solo “praias, dunas e areais costeiros” (NEAR_DIST_COS2007_PRAIAS) não revelava qualquer relação global com a variável dependente, mas quando observadas as relações locais, identificam-se 7 freguesias (3,38%) em que existe uma relação negativa linear, isto é, quanto maior a distância às praias, menor o preço de arrendamento, nestas freguesias.

A proporção de população residente com o ensino superior completo (2001_PPOPRESSESSUP) revela estar positivamente relacionada com os preços das rendas em 83 freguesias (40,10%), especialmente nas freguesias litorais da margem Norte do rio Tejo, facto que corrobora a, embora fraca, relação positiva global identificada nos gráficos de dispersão ($r^2=0,16$) e na matriz de correlação de Pearson ($r=0,39$). De forma semelhante, a proporção de alojamentos familiares propriedade dos ocupantes (2001_PAFPO) apresenta uma relação positiva linear com a variável dependente em 62 freguesias (29,95%) e uma negativa linear em 4 outras (1,93%). Verificam-se ainda outros tipos de relação: convexa em 70 freguesias (33,82%); côncava em 19 freguesias (9,18%); e complexa em 4 freguesias (1,93%). O número de divisões dos alojamentos familiares (2001_NDIV) tende a ter uma relação positiva linear com o valor médio das rendas em 62 freguesias (29,95%), uma relação convexa em 15 (7,25%) e uma relação complexa em apenas 1 (0,48%).

Relativamente às variáveis que apresentam uma maior predominância de relações locais negativas, destaca-se a proporção de alojamentos superlotados (2001_PASUPERLOT), com 60 (28,99%) freguesias a apresentar uma

relação negativa linear, essencialmente localizadas no município central de Lisboa, 5 (2,42%) com uma relação convexa e 1 (0,48%) com relação indefinida complexa. No mesmo seguimento, a taxa de desemprego (2001_TX_DESEMP) também aparenta ter valores mais elevados onde os preços das rendas tendem a ser mais reduzidos, em 38 (18,36%) freguesias, localizadas maioritariamente ao longo da autoestrada Circular Regional Interior de Lisboa (CRIL), que demarca sensivelmente o limite da periferia do centro urbano de Lisboa, e, a Sul, na continuidade da Ponte 25 de Abril até Sesimbra. Esta variável apresenta, ainda, outros tipos de relações locais com a variável dependente: positiva linear em 1 (0,48%); côncava em 5 (2,42%); convexa em 1 (0,48%); e complexa em 7 (3,38%). A proporção de população residente com 65 ou mais anos (2001_PPOPRES_65MAIS) apresenta também uma relação linear negativa em 21 (10,14%) freguesias, apesar da relação convexa ser a mais predominante, presente em 28 (13,53%) freguesias.

Finalmente, ainda que se observe apenas uma relação linear significativa em 2 (0,97%) freguesias, a proporção de edifícios construídos nos 10 anos anteriores aos censos de 2001 (2001_PED_10ANOS) apresenta uma relação convexa com o valor médio das rendas em 15 (7,25%) freguesias.

Tal como detetado anteriormente na análise da matriz de correlação de Pearson, a variável 2001_PAF_ANTES1960 aparenta estar muito correlacionada com duas outras (2001_PAFPO e 2001_PPOPRES_65MAIS), pelo que, apesar de apresentar uma relação linear com a variável dependente em 23 freguesias (11,11%), será excluída do modelo, por forma a evitar multicolinearidade.

4.2. MULTICOLINEARIDADE GLOBAL

De modo a testar a multicolinearidade global das variáveis selecionadas anteriormente, procedeu-se à aplicação do modelo OLS. Tal como a Tabela 4.2.1 demonstra, nenhum dos fatores explicativos revela um VIF > 7,5 pelo que se pode considerar que o modelo não exhibe níveis de multicolinearidade global preocupantes, ainda que, a variável 2001_PPOPRES_65MAIS apresente um VIF ligeiramente superior a 4, o que, em análises mais rigorosas, poderá indicar um nível moderado de multicolinearidade, sendo necessário ter este facto em mente para próximas análises.

Tabela 4.2.1 Resultados da multicolinearidade global para os dados de 2001

Variável	VIF
Intercept	-----
2001_NDIV	2,789098
2001_PAFPO	2,85014
2001_PASUPERLOT	3,411745
2001_PED_10ANOS	1,784255
2001_PPOPRES_65MAIS	4,251506
2001_PPOPRESSENSUP	2,268468
2001_DURMOV PEND	2,125479
2001_TX_DESEMP	1,684539
NEAR_DIST_COS2007_PRAIAS	1,766934

4.3. MODELO MGWR

Nesta secção serão analisados os resultados obtidos pela aplicação do modelo MGWR, seguindo minuciosamente todos os passos de calibração, diagnóstico e validação. Tendo por base a análise exploratória anteriormente realizada, que levou à seleção das 9 variáveis listadas na Tabela 4.2.1, o modelo em causa pode ser representado pela seguinte equação:

$$Y_i = \beta_{bw0}(u_i, v_i) + \beta_{bw1}(u_i, v_i)X_{1i} + \beta_{bw2}(u_i, v_i)X_{2i} + \beta_{bw3}(u_i, v_i)X_{3i} + \beta_{bw4}(u_i, v_i)X_{4i} + \beta_{bw5}(u_i, v_i)X_{5i} \\ + \beta_{bw6}(u_i, v_i)X_{6i} + \beta_{bw7}(u_i, v_i)X_{7i} + \beta_{bw8}(u_i, v_i)X_{8i} + \beta_{bw9}(u_i, v_i)X_{9i} + \varepsilon_i, \quad i = 1, \dots, 207$$

Em que:

- Y corresponde ao valor médio mensal das rendas nos alojamentos familiares arrendados, em 2001 (2001_RENDAS);
- X₁ corresponde ao número de divisões nos alojamentos familiares clássicos (2001_NDIV);
- X₂ corresponde à proporção de alojamentos familiares de residência habitual propriedade dos ocupantes (2001_PAFPO);
- X₃ corresponde à proporção de alojamentos superlotados (2001_PASUPERLOT);
- X₄ corresponde à proporção de edifícios construídos nos últimos 10 anos (2001_PED_10ANOS);
- X₅ corresponde à proporção de população residente com 65 ou mais anos (2001_PPOPRES_65MAIS);
- X₆ corresponde à proporção de população residente com o ensino superior completo (2001_PPOPRESSENSUP);
- X₇ corresponde à duração média dos movimentos pendulares da população residente empregada ou estudante (2001_DURMOVPEND);
- X₈ corresponde à taxa de desemprego (2001_TX_DESEMP);
- X₉ corresponde à distância à classe da COS “Praias, dunas e areais costeiros” mais próxima (NEAR_DIST_COS2007_PRAIAS);
- (u_i, v_i) refere-se às coordenadas dos centroides de cada freguesia;
- β_{bw} representa os coeficientes das variáveis explicativas ajustados de acordo com a largura de banda (bw) de cada uma;
- ε_i corresponde ao erro residual.

4.3.1. LARGURAS DE BANDA E ESTATÍSTICAS DOS COEFICIENTES

As estatísticas dos coeficientes e as larguras de banda de cada variável encontram-se representadas na Tabela 4.3.1. Quanto às larguras de banda individuais, observa-se uma relação global entre a proporção de população residente com 65 ou mais anos (2001_PPOPRES_65MAIS) e os valores médios das rendas, uma vez que a largura de banda desta variável inclui a totalidade das freguesias em estudo. A taxa de desemprego (2001_TX_DESEMP) aparenta ter uma relação regional com a variável dependente, já que a vizinhança ótima representa 84,54% das freguesias da AML. As restantes variáveis (2001_NDIV; 2001_PAFPO; 2001_PASUPERLOT; 2001_PED_10ANOS;

2001_PPOPRESSESSUP; 2001_DURMOVPEPEND; e NEAR_DIST_COS2007_PRAIAS) aparentam ter uma relação mais local com os preços do arrendamento. O *intercept*, isto é, a variabilidade inexplicável de Y na área de estudo parece operar também a uma escala local, já que a largura de banda representa 27,05% das freguesias.

Tabela 4.3.1 Larguras de banda e estatísticas dos coeficientes para o modelo de 2001

Variável (scaled)	Largura de banda		Freguesias com coeficiente significativo		Média	Desvio-padrão	Mínimo	Mediana	Máximo
	N.º	%	N.º	%					
Intercept	56	27,05	38	18,36	-0,3083	0,2761	-0,7786	-0,3227	0,2215
2001_NDIV	66	31,88	0	0,00	0,0786	0,1079	-0,2159	0,1098	0,3564
2001_PAFPO	34	16,43	57	27,54	0,3638	0,2786	-0,1007	0,3188	1,1848
2001_PASUPERLOT	78	37,68	0	0,00	-0,0708	0,1020	-0,2077	-0,0815	0,2397
2001_PED_10ANOS	82	39,61	26	12,56	-0,1063	0,1280	-0,4078	-0,1165	0,1955
2001_PPOPRES_65MAIS	207	100,00	207	100,00	-0,2978	0,0074	-0,3152	-0,2974	-0,2851
2001_PPOPRESSESSUP	60	28,99	29	14,01	0,1597	0,2096	-0,4302	0,1127	0,5847
2001_DURMOVPEPEND	56	27,05	20	9,66	-0,0341	0,3229	-0,5608	-0,0637	0,4429
2001_TX_DESEMP	175	84,54	40	19,32	-0,0944	0,0739	-0,2394	-0,0800	0,0201
NEAR_DIST_COS2007_PRAIAS	46	22,22	38	18,36	-0,2346	0,4838	-0,9634	-0,3501	0,7885

Note-se que larguras de banda mais amplas indicam menor variabilidade dos coeficientes e, portanto, desvios-padrões mais reduzidos, assim como amplitudes entre os valores mínimos e máximos mais reduzidos também. Veja-se o exemplo dos coeficientes da variável 2001_PPOPRES_65MAIS que, com uma largura de banda que abrange as 207 freguesias, apresentam o menor desvio-padrão (0,0074) e a menor amplitude entre o mínimo e o máximo (0,0301). Por outro lado, atesta-se que a influência negativa da população idosa nos preços das rendas é efetivamente estacionária no espaço, sendo que todos os coeficientes têm valor abaixo de 0.

Em média, a variável 2001_PAFPO é a que mais influencia os preços das rendas, sendo que o facto da média dos coeficientes (0,3638) ser positiva indica que a percentagem de alojamentos familiares propriedade dos ocupantes tende a elevar as rendas. Contrariamente, variáveis como 2001_DURMOVPEPEND, 2001_PASUPERLOT e 2001_NDIV tendem a ter uma menor influencia nos valores das rendas, sendo que as duas últimas não revelaram qualquer associação estatisticamente significativa em nenhuma das freguesias.

4.3.2. MULTICOLINEARIDADE LOCAL

Por forma a testar a multicolinearidade local, foi mapeada a distribuição dos valores de CN (Figura 4.3.1), verificando-se que em toda a área de estudo os valores de CN são inferiores a 30, o que indica que a multicolinearidade local não será um problema neste modelo.

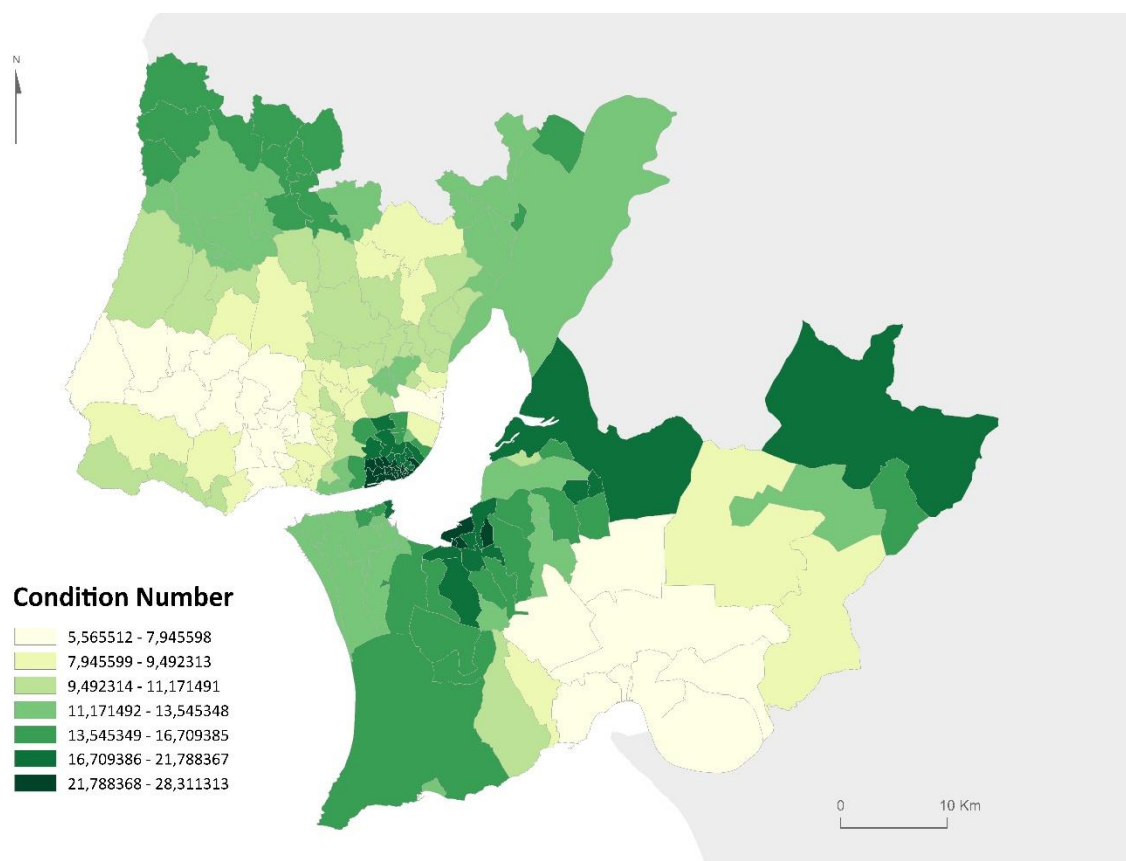


Figura 4.3.1 Multicolinearidade local no modelo de 2001

4.3.3. OBSERVAÇÕES INFLUENTES

Relativamente ao teste sobre a existência de observações influentes, o mapeamento da distância de Cook (Figura 4.3.2) permite concluir que as observações influentes não aparentam ser um problema neste modelo, já que todas as freguesias revelam um Cook's $D < 1$.

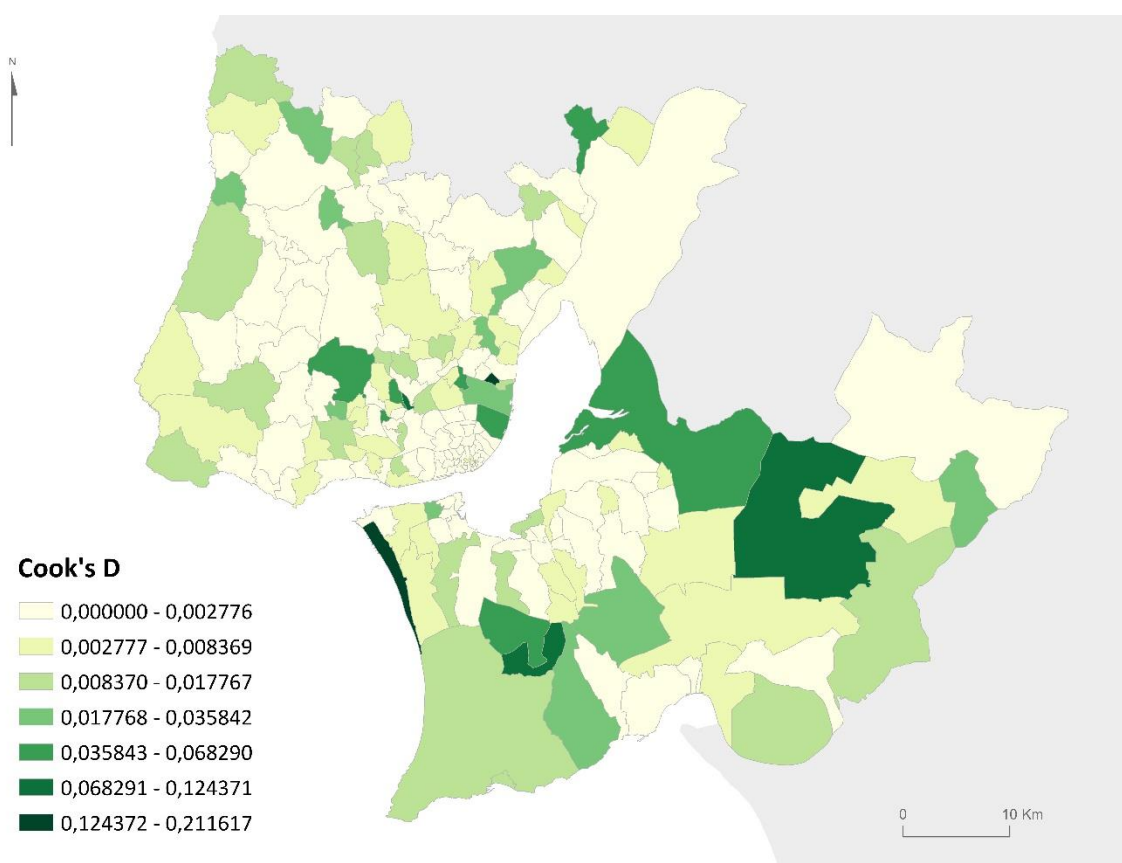


Figura 4.3.2 Distância de Cook no modelo de 2001

4.3.4. ADEQUAÇÃO DO MODELO

Quando se executa o modelo MGWR no *ArcGIS Pro*, o próprio programa apresenta também os resultados das estatísticas de adequação do modelo GWR, permitindo comparar a precisão dos dois modelos. Deste modo, a análise da Tabela 4.3.2 possibilita verificar que o modelo MGWR revela melhor adequação, tanto pelo facto de apresentar um Adj-R^2 superior (0,7479), como por exibir um AICc consideravelmente inferior (393,5349, face aos 435,0693 do modelo GWR).

Tabela 4.3.2 Medidas de adequação do modelo de 2001

Modelo	R ²	Adj-R ²	AICc	ENP	Graus de Liberdade Efetivos
GWR 9 variáveis	0,7910	0,6705	435,0693	-	131,6174
MGWR 9 variáveis	0,8158	0,7479	393,5349	55,45	151,5568

Por outro lado, ao mapear-se os valores do coeficiente de determinação (*Local R*²) é possível identificar os locais onde o modelo melhor se adequa. A Figura 4.3.3 mostra uma melhor precisão do modelo no centro da área de estudo, nas freguesias junto à margem Norte do rio Tejo (Lisboa, Oeiras e Cascais), bem como nas freguesias a Este da Baía do Seixal (Barreiro, Moita, ...). Na Grande Lisboa, a precisão do modelo decresce gradualmente à

medida que se afasta do centro urbano de Lisboa, em direção aos municípios de Sintra e Mafra, enquanto na Península de Setúbal a adequação do modelo é mais reduzida nas freguesias a Este.

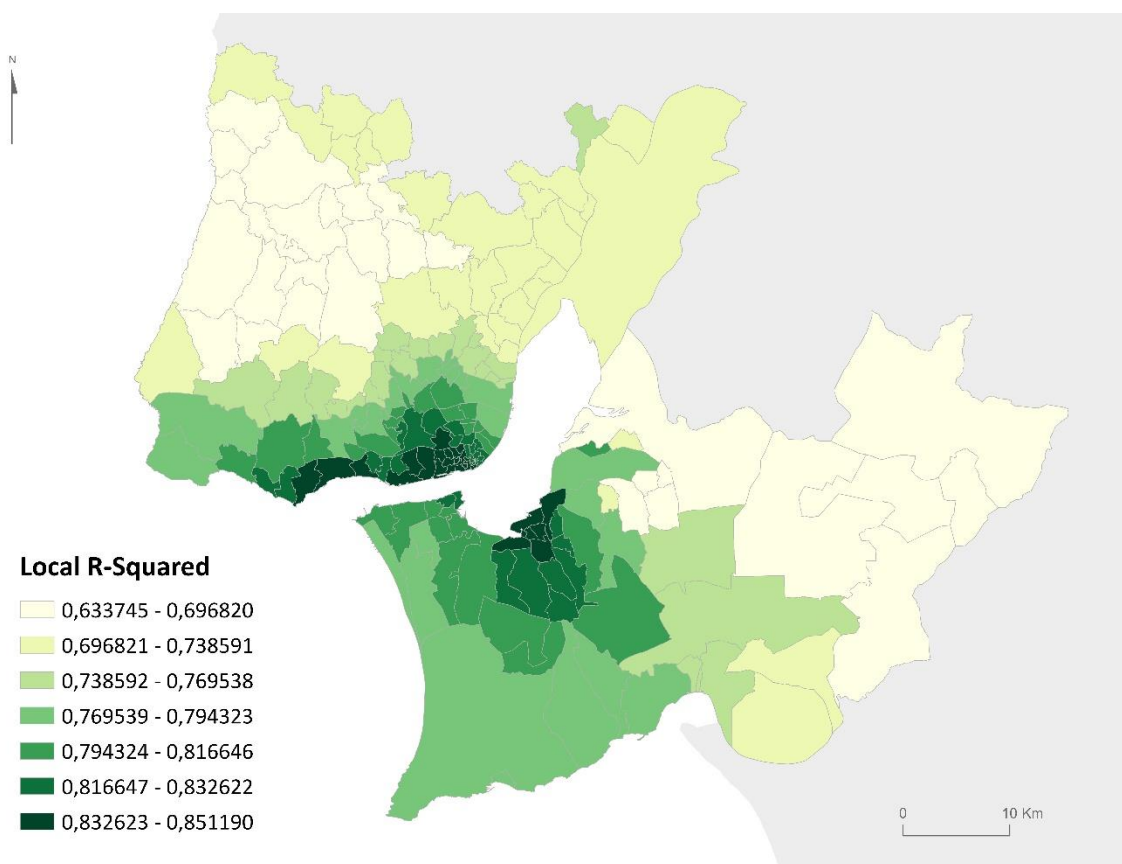


Figura 4.3.3 Local R-squared no modelo de 2001

4.3.5. ANÁLISE DE RESÍDUOS

Como última fase de diagnóstico deste modelo, é crucial verificar que os resíduos se distribuem de forma aleatória no espaço, caso contrário, a existência de autocorrelação espacial nos resíduos poderá sugerir que este se encontra mal especificado. Como tal, tanto o mapa de distribuição dos resíduos estandardizados (Figura 4.3.4), como os resultados do teste de *Global Moran's I* atestam a aleatoriedade estatisticamente significativa dos resíduos, este último com um índice de -0,029847 e um valor de p (0,312923), superior ao nível de significância de 0,05, o que propõe uma não rejeição da hipótese nula da aleatoriedade espacial da distribuição dos resíduos.

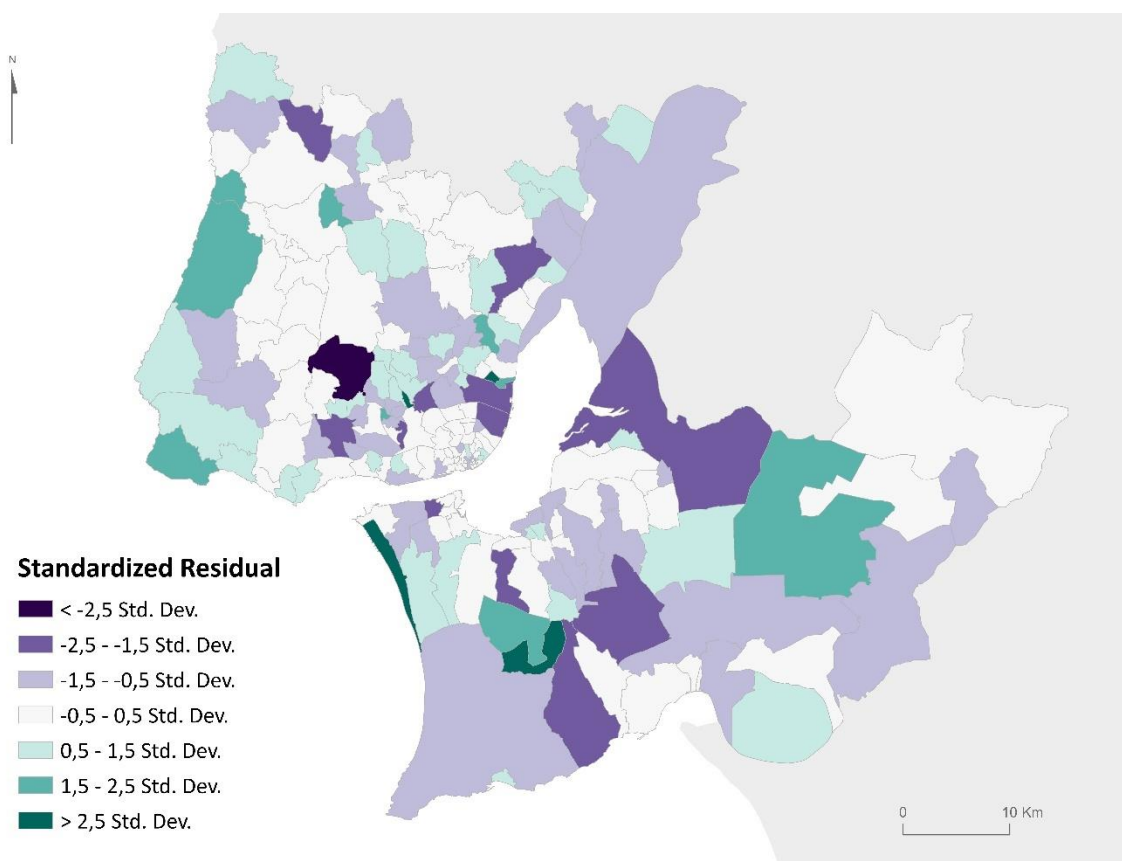


Figura 4.3.4 Resíduos estandardizados no modelo de 2001

4.3.6. COEFICIENTES LOCAIS

Os mapas da Figura 4.3.5 mostram a distribuição espacial dos coeficientes estimados para cada variável explicativa no modelo MGWR de 2001. De um modo geral, os resultados fundamentam a importância de um modelo local, que capta as nuances do impacto das variáveis em regiões distintas e revela padrões espaciais impossíveis de visualizar através de um modelo global.

Os coeficientes das variáveis 2001_NDIV e 2001_PASUPERLOT são todos não significativos, o que sugere que nem o número de divisões dos alojamentos nem a proporção de alojamentos superlotados revelam um efeito estatisticamente relevante sobre os preços das rendas.

A proporção de alojamentos familiares propriedade dos ocupantes (2001_PAFPO) apresenta coeficientes positivos na maior parte do território da AML, em especial nas freguesias circundantes do centro urbano, junto à margem Noroeste do rio (Cascais), ao longo da radial periférica (Loures, Odivelas, Amadora e Oeiras) e na parte Oeste da Península de Setúbal (Almada, Seixal e Sesimbra). De destacar, no entanto, o agrupamento onde a influência desta variável é especialmente significativa (coeficiente $> 1,0364$), constituído pelas freguesias de Reboleira, Falagueira, Venda Nova, Brandoa, Alforneiros e Pontinha.

A percentagem de edifícios construídos nos últimos 10 anos (2001_PED_10ANOS) não apresenta uma influência significativa na maior parte do território, mas onde existe significância estatística, os coeficientes são negativos, concentrados nas freguesias imediatamente a Norte de Lisboa (essencialmente no município de Loures).

Tal como já identificado nas análises anteriores sobre as estatísticas dos coeficientes e das larguras de banda, a proporção de população com 65 ou mais anos (2001_PPOPRES_65MAIS) influencia de forma global todo o território e de forma relativamente homogénea, isto é, áreas com maior proporção de idosos tendem a registar rendas mais baixas.

A proporção de população com ensino superior completo revela alguma variabilidade espacial na significância dos coeficientes, apresentando coeficientes estatisticamente significativos positivos no extremo Sudoeste da Grande Lisboa (Cascais) e, na margem Sul do rio, sensivelmente nos concelhos de Alcochete, Montijo e Moita. Curiosamente, destaca-se a freguesia de Montelavar, em Sintra, como a única onde o coeficiente é negativo e estatisticamente significativo, embora esteja rodeada de freguesias onde o coeficiente é igualmente negativo, mas não significativo.

Já a duração dos movimentos pendulares revela uma influência positiva relevante nas freguesias de Loures, Caneças, Ramada, Famões, Pontinha, Casal de Cambra, São Brás, Brandoa, Falagueira e Mina (Loures, Odivelas, Sintra e Amadora), enquanto que nas freguesias de São Francisco Xavier, Santa Maria de Belém, Caparica, Sobreda, Pragal, Almada, Cacilhas, Cova da Piedade, Laranjeiro e Feijó (Lisboa e Almada), quanto menor a duração dos movimentos pendulares, mais inflacionados serão os valores das rendas.

Naturalmente, a taxa de desemprego (2001_TX_DESEMP) apresenta coeficientes negativos na grande maioria do território da AML, sendo estes significativos em toda a faixa Oeste e Norte da Grande Lisboa (incluindo as freguesias de Mafra e parte das de Sintra, Cascais, Loures e Vila Franca de Xira), o que indica que taxas de desemprego elevadas tendem a estar associadas a valores de arrendamento mais reduzidos.

Finalmente, a distância às praias (NEAR_DIST_COS2007_PRAIAS) regista uma influência negativa significativa em toda a costa Oeste da Grande Lisboa e na freguesia da Costa da Caparica, isto é, os preços das rendas tendem a ser inflacionados pela proximidade às praias. Esta influência deteta-se de igual forma nas freguesias de Vila Franca de Xira, possivelmente devido à sua proximidade ao Estuário do Tejo.

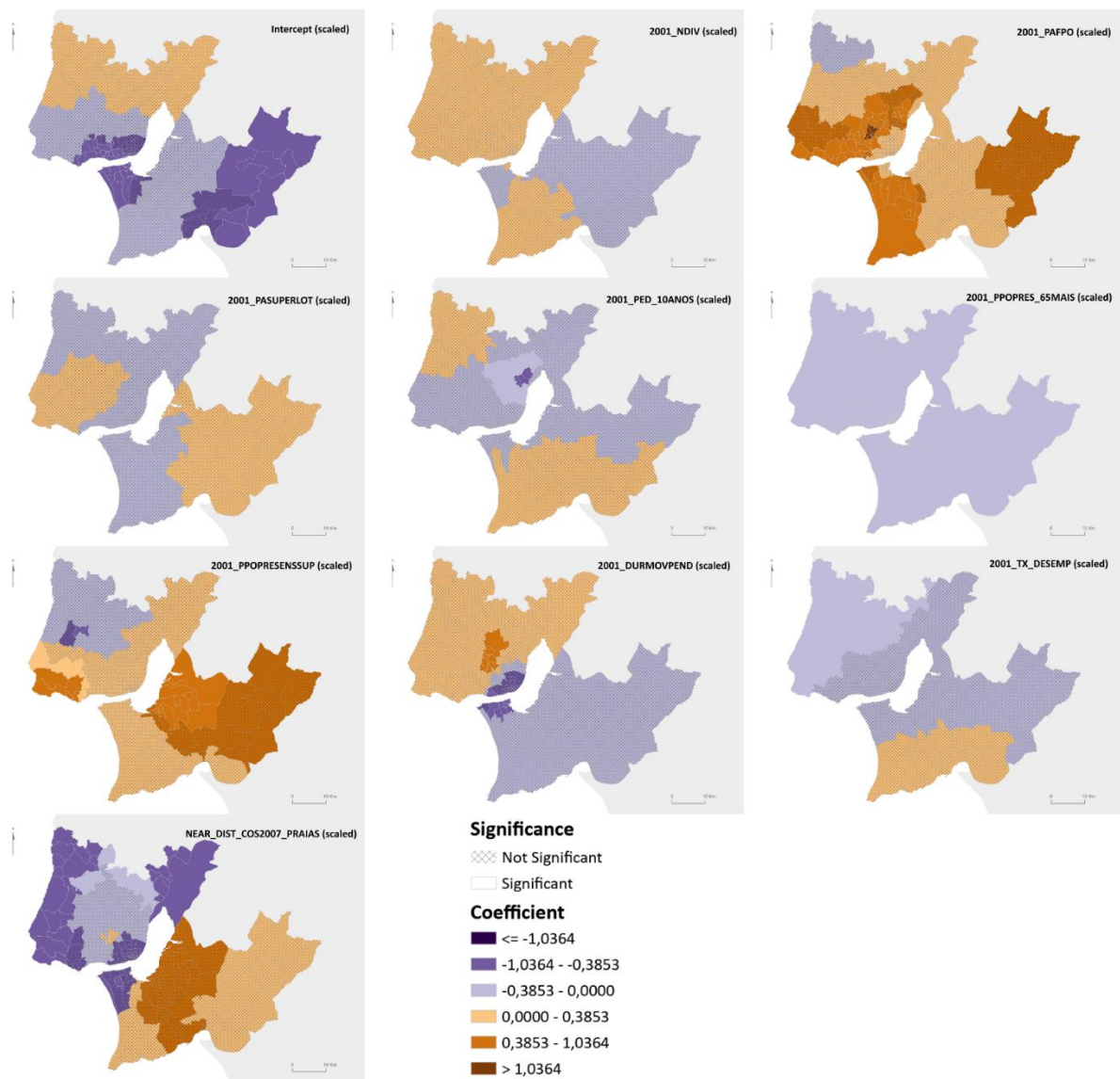


Figura 4.3.5 Coeficientes locais do modelo de 2001

5. RESULTADOS DO ESTUDO DE 2021

Neste capítulo são apresentados os resultados obtidos na análise espacial dos preços do arrendamento de habitação nas freguesias da Área Metropolitana de Lisboa, em 2021, considerando-se a variável dependente “valor médio mensal das rendas dos alojamentos familiares clássicos arrendados, em 2021”. Para tal, segue-se sensivelmente a mesma estrutura do capítulo anterior, começando-se com a análise exploratória dos dados espaciais (5.1), seguida da análise de multicolinearidade global (5.2), e terminando com a especificação do modelo MGWR adotado (5.3).

5.1. ANÁLISE EXPLORATÓRIA DOS DADOS ESPACIAIS

Ao longo deste subcapítulo será possível verificar que, não só as distribuições e padrões espaciais das rendas e dos potenciais fatores explicativos se transformaram face a 2001, mas também as relações entre as variáveis sofrem alterações substanciais. Estas diferenças fazem com que o modelo de 2021, embora mantenha a estrutura metodológica, culmine em resultados distintos, nomeadamente na seleção de algumas variáveis explicativas diferentes das do modelo anterior. Não obstante, as potenciais variáveis explicativas consideradas para o modelo de 2021 (Tabela 5.1.1) são as mesmas do modelo anterior, adicionando-se a variável 2021_RENDIMENTO, referente ao rendimento bruto declarado por agregado fiscal (€) e removendo a proporção de alojamentos superlotados.

Tabela 5.1.1 Descrição das potenciais variáveis explicativas consideradas para 2021

Variável	Descrição
2021_DENSALOI	Densidade de alojamentos (N.º/ km ²)
2021_PAFPO	Proporção de alojamentos familiares clássicos de residência habitual propriedade dos ocupantes (%)
2021_PAF_ANTES1960	Proporção de alojamentos familiares clássicos de residência habitual construídos antes de 1960 (%)
2021_PAFSAZ	Proporção de alojamentos familiares clássicos de uso sazonal (%)
2021_PAFV	Proporção de alojamentos familiares clássicos vagos (%)
2021_PAFQCENT	Proporção de alojamentos familiares com aquecimento central (%)
2021_NDIV	Divisões por alojamento familiar clássico de residência habitual (N.º)
2021_PED_10ANOS	Proporção de edifícios construídos nos últimos 10 anos (%)
2021_DENSPOP	Densidade populacional (N.º/ km ²)
2021_PPOPRES_65MAIS	Proporção da população residente com 65 ou mais anos (%)
2021_PPOPRESNE	Proporção da população residente de nacionalidade estrangeira (%)
2021_PPOPRESSESSUP	Proporção da população residente com ensino superior completo (%)
2021_DURMOVPEND	Duração média dos movimentos pendulares da população residente empregada ou estudante (min)
2021_TX_DESEMP	Taxa de desemprego (%)
2021_RENDIMENTO	Rendimento bruto declarado por agregado fiscal (€)
NEAR_DIST_COS2018_PARQUES_JARDINS	Distância à classe da COS "Parques e Jardins" mais próxima (m)
NEAR_DIST_COS2018_PRAIAS	Distância à classe da COS "Praias, Dunas e Areais Costeiros" mais próxima (m)
NEAR_DIST_COS2018_REDE_VIARIA	Distância à classe da COS "Rede Viária e Espaços Associados" mais próxima (m)

5.1.1. ESTATÍSTICA DESCRITIVA

De forma semelhante à distribuição de 2001, o valor médio mensal das rendas em 2021 (Figura 5.1.1) apresenta uma distribuição assimétrica positiva, uma vez que a média (379,00€) é superior à mediana (364,24€).

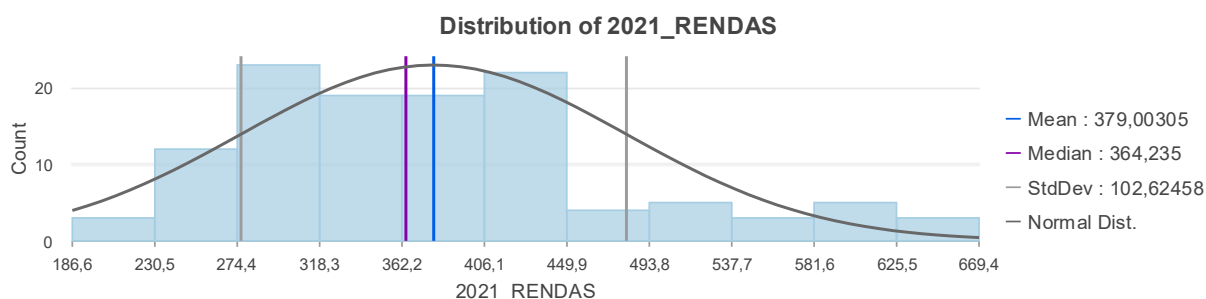


Figura 5.1.1 Histograma da distribuição do valor médio mensal das rendas dos alojamentos familiares clássicos arrendados, em 2021

Se em 2001 já se observava alguma variabilidade nos dados dos valores mensais das rendas, em 2021 tal agravou-se, registando-se um valor mínimo de 186,59€ na freguesia de Marvila, no concelho de Lisboa, e um valor máximo de 669,41€ na freguesia das Avenidas Novas, também localizada no município central. O incremento do desvio-padrão para 102,62 revela, igualmente, o aumento das assimetrias regionais, sobretudo no centro urbano de Lisboa.

Relativamente aos restantes possíveis fatores explicativos, representados no Apêndice D: Histogramas da Distribuição dos Potenciais Fatores Explicativos de 2021, observam-se os seguintes padrões de distribuição:

- Distribuição normal (média $\approx$ mediana): 2021_NDIV, 2021_PPOPRES_65MAIS, 2021_TX_DESEMP, 2021_DURMOVPEND;
- Distribuição assimétrica positiva (média $>$ mediana): 2021_DENSALoj, 2021_PAF_ANTES1960, 2021_PAFSAZ, 2021_PAFV, 2021_PAFACENT, 2021_PED_10ANOS, 2021_DENSPOP, 2021_PPOPRESNE, 2021_PPOPRESSENSUP, 2021_RENDIMENTO, NEAR_DIST_COS2018_PARQUES_JARDINS, NEAR_DIST_COS2018_PRAIAS, NEAR_DIST_COS2018_REDE_VIARIA;
- Distribuição assimétrica negativa (média $<$ mediana): 2021_PAFPO.

5.1.2. VISUALIZAÇÃO DOS DADOS

A Figura 5.1.2 permite visualizar de forma clara a distribuição heterogénea dos dados referentes ao valor médio mensal das rendas, através de uma classificação em cinco classes (quintis). Observa-se um padrão de concentração de valores mais elevados nas freguesias ao longo da margem Norte do rio Tejo, desde a União de Freguesias de Cascais e Estoril (Cascais) até à baixa de Lisboa (freguesia da Misericórdia), onde se estende para Norte, sensivelmente ao longo da estrada IP7 (Eixo Norte-Sul) até à freguesia do Lumiar (Lisboa). Já os valores mais reduzidos não seguem um padrão espacial tão claro, evidenciando-se alguns agrupamentos no Norte do município de Loures e, na Península de Setúbal, nos municípios de Moita, Barreiro, Palmela e Montijo.

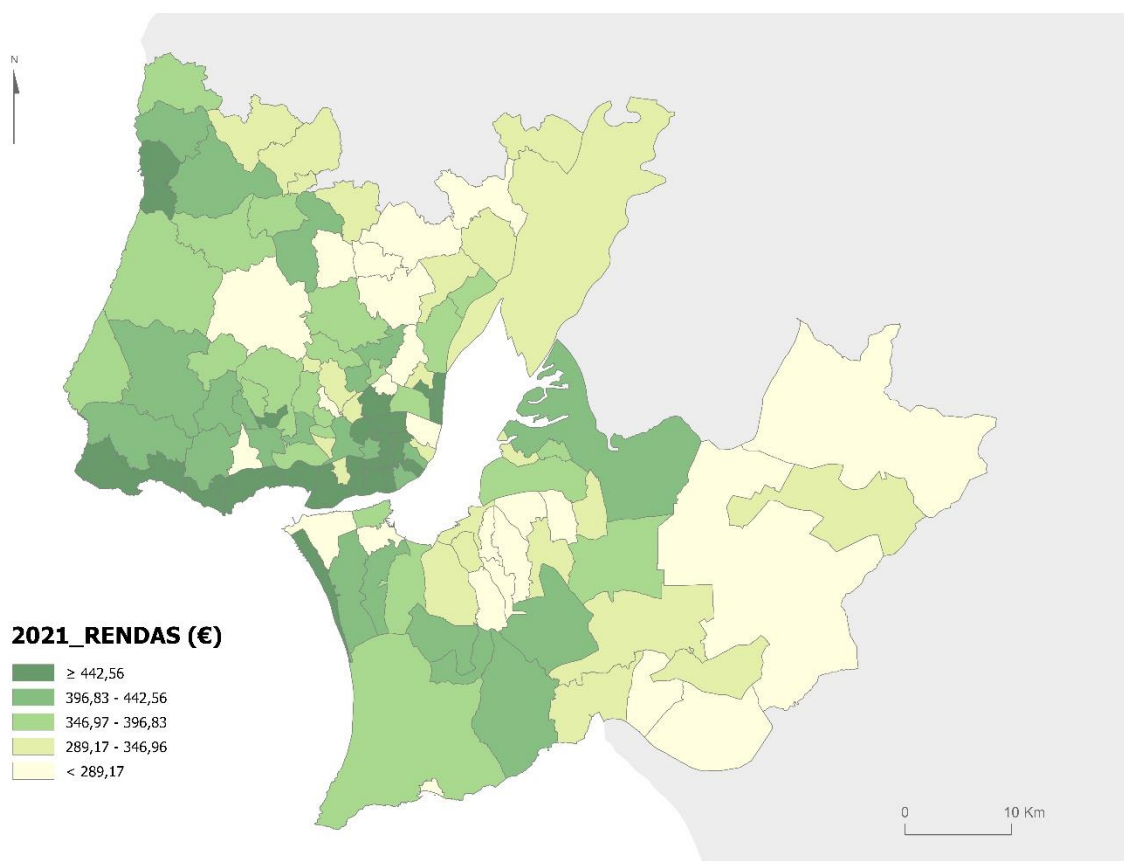


Figura 5.1.2 Valor médio mensal das rendas nos alojamentos familiares clássicos arrendados, em 2021

5.1.3. ANÁLISE DE PADRÕES ESPACIAIS

Os resultados do teste de *Global Moran's I* evidenciam a existência de autocorrelação espacial positiva no conjunto de dados sobre o valor médio das rendas, apresentando um valor do índice de 0,243369 e um valor de p de 0,000000, inferior ao nível de significância de 0,05. Interpreta-se daqui, que os valores médios das rendas em 2021 tendem a apresentar-se em agrupamentos, isto é, valores elevados tendem a estar próximos entre si, e o mesmo tende a acontecer com os valores reduzidos.

Tal como se observa na Figura 5.1.3, além da autocorrelação espacial, esta variável é caracterizada pela não-estacionaridade, isto é, os padrões espaciais identificados não são homogéneos no espaço.

Os resultados do teste de Local Moran's I revelam que: 25 freguesias, localizadas essencialmente no município de Lisboa e Cascais, são identificadas como *clusters* de valores elevados; 13 são *clusters* de valores reduzidos, agrupados junto à margem Sul do rio Tejo, nos municípios do Montijo, Moita e Barreiro; 15 freguesias apresentam-se como *outliers* de valores reduzidos próximos de valores elevados (*Low-High Outlier*), em torno do centro urbano de Lisboa; e a União das Freguesias de Póvoa de Santa Iria e Forte da Casa, em Vila Franca de Xira, é o único polígono identificado como *outlier* de valores elevados próximo de valores reduzidos (*High-Low Outlier*).

De um modo geral, a estatística de Getis-Ord GI corrobora estes mesmos padrões espaciais, na medida em que se identifica, sensivelmente nas mesmas localizações, um total de 43 freguesias consideradas *clusters* de valores elevados (*hot spots*), das quais, 32 com 99% de confiança, 8 com 95% de confiança, e 3 com 90% de confiança, e um total de 17 freguesias consideradas *clusters* de valores reduzidos (*cold spots*), 2 das quais com 99% de confiança, 8 com 95% de confiança e 7 com 90% de confiança.

Estes resultados atestam a importância do uso de um modelo local, como GWR ou MGWR, tornando-se desapropriado o uso de um modelo de regressão linear global (i.e., OLS) que não atende às especificidades espaciais da variável.

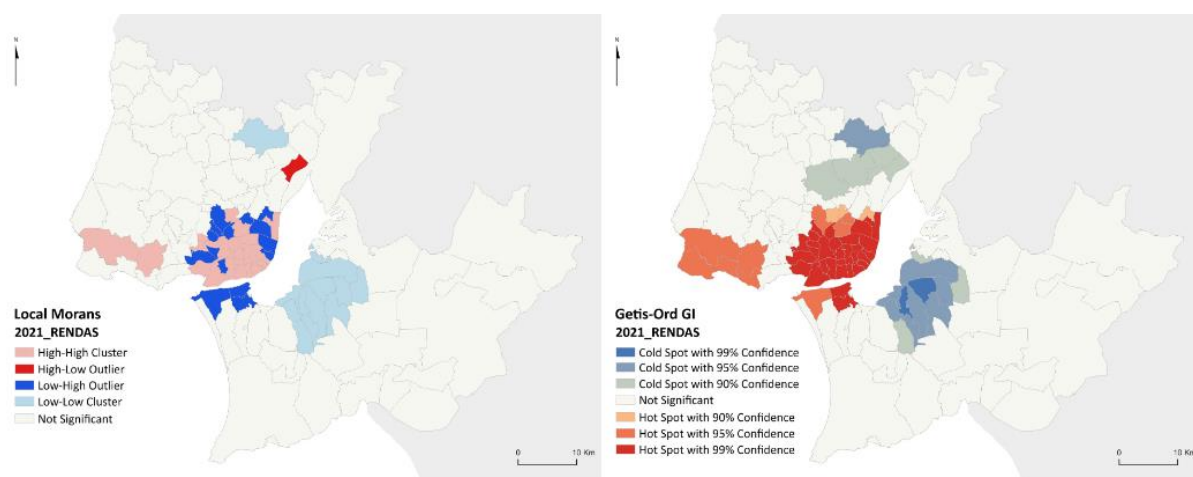


Figura 5.1.3 Local Moran's I e Getis-Ord GI da variável 2021_RENDAS

5.1.4. RELAÇÕES BIVARIADAS

Seguindo o mesmo processo metodológico, compreender as relações entre variáveis torna-se um passo fundamental para identificar o melhor conjunto possível de variáveis independentes que possam explicar a variação dos valores médios das rendas. Para tal, os gráficos de dispersão e a matriz de correlação de Pearson representados na Figura 5.1.4 permitem verificar a existência, intensidade e sentido das relações globais entre as 18 potenciais variáveis explicativas e a variável dependente.

De acordo com estes coeficientes de regressão linear simples, o valor médio mensal das rendas em 2021 (2021_RENDAS) apresenta uma correlação positiva forte com duas variáveis: a proporção de população residente com o ensino superior completo (2021_PPOPRESNSUP); e o rendimento bruto declarado por agregado fiscal (2021_RENDIMENTO). Identifica-se também uma relação global moderada a fraca, ainda que positiva, com as variáveis: 2021_PAF_ANTES1960; 2021_PAFAQCENT; 2021_DENSAL0J; 2021_DENSPOP; 2021_PPOPRESNE. Para além disso, verificam-se, ainda, algumas variáveis que aparentam estar negativamente relacionadas com a variável dependente, embora tais relações mostrem uma intensidade moderada a fraca: 2021_DURMOVPEND; 2021_TX_DESEMP; NEAR_DIST_COS2018_PRAIAS; NEAR_DIST_COS2018_PARQUES_JARDINS; NEAR_DIST_COS2018_REDE_VIARIA. Contrariamente, as rendas aparentam estar pouco ou nada relacionadas

com a proporção de edifícios construídos nos últimos 10 anos (2021_PED_10ANOS) e a proporção de população residente com 65 ou mais anos (2021_PPOPRES_65MAIS).

Quanto às relações entre as potenciais variáveis explicativas, encontram-se alguns valores elevados do coeficiente de Pearson, que poderão indicar redundância entre os seguintes pares: 2021_PAFPO e 2021_PAF_ANTES1960; 2021_PAF_ANTES1960 e 2021_PAFV; 2021_DENSALOJ e 2021_DENSPOP; 2021_PAFAQCENT e 2021_RENDIMENTO; e 2021_PPOPRESSENSUP e 2021_RENDIMENTO.

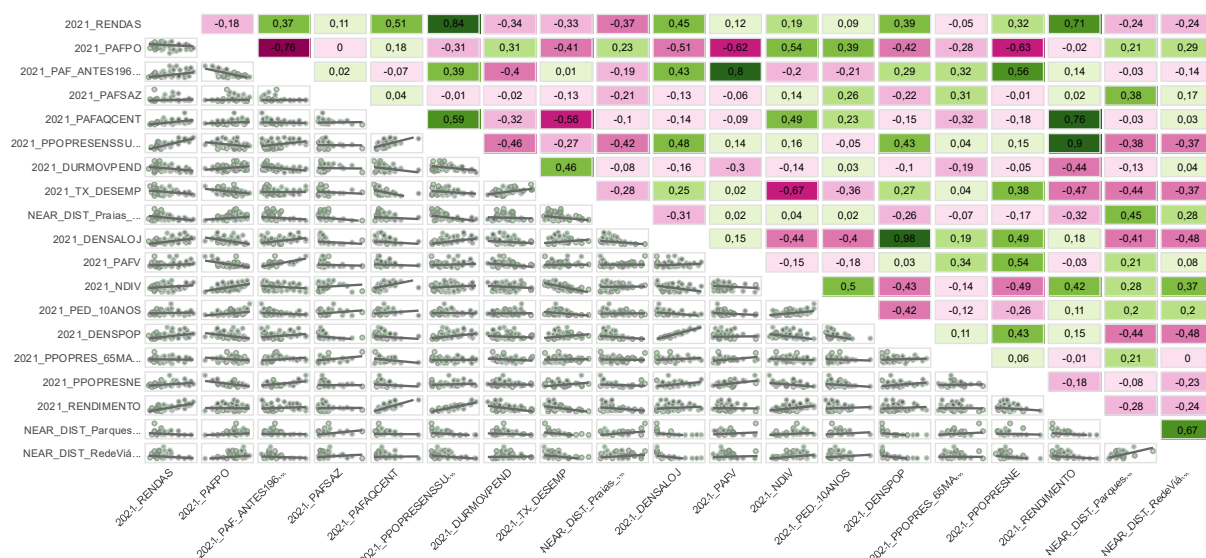


Figura 5.1.4 Gráficos de dispersão e matriz de correlação de Pearson do valor médio mensal das rendas e 18 potenciais variáveis explicativas, dados de 2021

Simultaneamente a esta análise de relações bivariadas globais, deve-se analisar também as relações bivariadas locais, já que o uso de um modelo de regressão local como MGWR implica obrigatoriamente que existam relações lineares locais significantes entre as variáveis independentes e a variável dependente. Assim, a Tabela 5.1.2 sintetiza as relações bivariadas locais mapeadas no Apêndice F: Relações bivariadas locais significantes entre fatores explicativos e o valor médio mensal das rendas, em 2021.

Dos 18 potenciais fatores explicativos considerados, apenas 9 revelam alguma relação linear local significativa com a variável dependente, sendo elas: 2021_PAFPO; 2021_PAF_ANTES1960; 2021_PAFAQCENT; 2021_PAFSAZ; 2021_PPOPRESSENSUP; 2021_DURMOVPEND; 2021_TX_DESEMP; 2021_RENDIMENTO; NEAR_DIST_COS2018_PRAIAS. Não obstante, as restantes 9 variáveis, que não apresentaram relações significativas locais com a variável dependente, foram excluídas do modelo e qualquer análise posterior não as terá em consideração.

Globalmente, a proporção de alojamentos familiares propriedade dos ocupantes (2021_PAFPO) apresentava uma relação muito fraca, de sentido negativo ($r = -0,18$), com o valor médio das rendas. No entanto, localmente observam-se diferentes tipologias de relações: positiva linear em 25 freguesias (21,19%); negativa linear em 10

(8,47%); côncava em 3 (2,54%); convexa em 6 (5,08%) e indefinida complexa em 18 (15,25%). Constatase, contudo, que as relações positivas lineares se concentram preponderantemente na zona Norte e Oeste da Grande Lisboa, enquanto as negativas lineares localizam-se junto à margem Sul do rio Tejo, nos municípios de Almada, Seixal, Barreiro e Moita, evidenciando novamente os contrastes entre as duas grandes regiões que compõem a AML. Similarmente, a variável 2021_PAFSAZ também registava uma relação global muito fraca ($r = 0,11$) com a variável dependente, mas quando observadas as relações locais identificam-se 55 freguesias (46,61%) onde existe uma relação linear positiva, isto é, quanto maior a proporção de alojamentos familiares de uso sazonal, mais inflacionados serão os preços das rendas. Além disso, em 17 outras freguesias, foram identificados outros tipos de relação estatisticamente significativa, das quais: 2 (1,69%) côncava; 1 (0,85%) convexa; e 14 (11,86%) complexa.

Tabela 5.1.2 Classificação das relações bivariadas locais entre variáveis explicativas e a variável dependente (2021_RENDAS)

Variável	Tipo de relação										Total de freguesias com relação linear significativa	
	Positiva linear		Negativa linear		Côncava		Convexa		Complexa			
	N.º	%	N.º	%	N.º	%	N.º	%	N.º	%	N.º	%
2021_PAFPO	25	21,19	10	8,47	3	2,54	6	5,08	18	15,25	35	29,66
2021_PAF_ANTES1960	8	6,78	0	0	0	0	4	3,39	0	0	8	6,78
2021_PAFAQCENT	40	33,90	0	0	9	7,63	2	1,69	0	0	40	33,90
2021_PAFSAZ	55	46,61	0	0	2	1,69	1	0,85	14	11,86	55	46,61
2021_PPOPREENSSUP	108	91,53	1	0,85	0	0	0	0	0	0	109	92,37
2021_DURMOVPEND	0	0	38	32,20	2	1,69	21	17,80	6	5,08	38	32,20
2021_TX_DESEMP	0	0	69	58,47	0	0	0	0	0	0	69	58,47
2021_RENDIMENTO	66	55,93	0	0	0	0	0	0	0	0	66	55,93
NEAR_DIST_COS2018_PRAIAS	0	0	16	13,56	0	0	1	0,85	1	0,85	16	13,56

A forte correlação global positiva entre a proporção de população residente com o ensino superior completo (2021_PPOPREENSSUP) e os valores médios das rendas, identificada nos gráficos de dispersão ($r^2=0,7$) e na matriz de correlação de Pearson ($r=0,84$), é corroborada pela existência de uma relação local positiva linear em 108 freguesias (91,53%). Intrinsecamente relacionado com a escolaridade da população está o seu rendimento. Por este motivo, não surpreende que a variável 2021_RENDIMENTO apresente também uma relação positiva linear em 66 freguesias (55,93%). Adicionalmente, a proporção de alojamentos familiares com aquecimento central (2021_PAFAQCENT) mostra estar positivamente relacionada com os valores das rendas em 40 freguesias (33,90%) localizadas essencialmente no centro urbano de Lisboa, com continuidade ao longo da margem Norte do rio Tejo até Cascais, e na faixa Noroeste da Península de Setúbal (Almada). Além do mais, verificam-se outras relações não lineares: côncava em 9 freguesias (7,63%) e convexa em 2 (1,69). Por sua vez, a proporção de alojamentos familiares construídos antes de 1960 (2021_PAF_ANTES1960) aparenta influenciar positivamente os preços das rendas em 8 freguesias (6,78%) localizadas predominantemente junto à Baía do Seixal, observando-se, ainda, uma relação convexa em 4 outras freguesias (3,39%).

Contrariamente, de entre as variáveis com maior predominância de associações locais negativas destaca-se a taxa de desemprego (2021_TX_DESEMP). Esta variável apresenta uma relação negativa linear em 69 freguesias

(58,47%), especialmente aglomeradas ao longo de ambas as margens do rio Tejo. Por outro lado, a duração média dos movimentos pendulares (2021_DURMOVPEND) tende a estar negativamente relacionada com a variável dependente em 38 freguesias (32,20%), verificando-se, ainda, uma relação convexa em 21 freguesias (17,80%), côncava em 2 (1,69%) e complexa em 6 (5,08%). Finalmente, observam-se alguns locais, num total de 16 freguesias (13,56%), onde quanto maior a distância às praias, menor o valor médio das rendas: em torno das freguesias da Ericeira e Mafra (Noroeste da área de estudo); junto à freguesia de Alcabideche (Cascais); e na zona de Vila Franca de Xira.

5.2. MULTICOLINEARIDADE GLOBAL

Tal como constatado na análise das relações bivariadas, a proporção de população residente com o ensino superior completo (2021_PPOPREENSSUP) tende a estar muito correlacionada com o rendimento dos agregados familiares (2021_RENDIMENTO), o que poderá indicar redundância entre as duas variáveis. Na verdade, a regressão linear exploratória (Tabela 5.2.1), que testa a multicolinearidade global, revela que ambas apresentam níveis elevados de multicolinearidade, com $VIF > 7,5$. Deste modo, procedeu-se à remoção da variável 2021_RENDIMENTO e, ao executar novamente o OLS, verifica-se que nenhuma das 8 variáveis acusa um nível de multicolinearidade global preocupante, visto que todas possuem um $VIF < 7,5$. Contudo, deve-se tomar atenção às variáveis 2021_PAFPO e 2021_TX_DESEMP, pois um $VIF > 4$ poderá indicar algum nível moderado de multicolinearidade. Neste sentido, o modelo prossegue com as 8 variáveis selecionadas.

Tabela 5.2.1 Resultados da multicolinearidade global para os dados de 2021

Variável	VIF (9 variáveis)	VIF (8 variáveis)
<i>Intercept</i>	-----	-----
2021_PAFPO	5,470609	5,397638
2021_PAF_ANTES1960	3,934369	3,720534
2021_PAFSAZ	1,215462	1,214605
2021_PAFQCENT	3,109443	2,571138
2021_PPOPREENSSUP	11,025913	3,054530
2021_DURMOVPEND	2,185318	2,146841
2021_TX_DESEMP	4,565256	4,467851
2021_RENDIMENTO	12,779127	-----
NEAR_DIST_COS2018_PRAIAS	1,923616	1,910927

5.3. MODELO MGWR

Tendo em conta as variáveis selecionadas através da análise exploratória realizada, o modelo MGWR de 2021, que será devidamente validado e analisado nos próximos subcapítulos, pode ser expresso pela seguinte equação:

$$Y_i = \beta_{bw0}(u_i, v_i) + \beta_{bw1}(u_i, v_i)X_{1i} + \beta_{bw2}(u_i, v_i)X_{2i} + \beta_{bw3}(u_i, v_i)X_{3i} + \beta_{bw4}(u_i, v_i)X_{4i} + \beta_{bw5}(u_i, v_i)X_{5i} \\ + \beta_{bw6}(u_i, v_i)X_{6i} + \beta_{bw7}(u_i, v_i)X_{7i} + \beta_{bw8}(u_i, v_i)X_{8i} + \varepsilon_i, \quad i = 1, \dots, 118$$

Em que:

- Y corresponde ao valor médio mensal das rendas nos alojamentos familiares arrendados, em 2021 (2021_RENDAS);
- X_1 corresponde à proporção de alojamentos familiares de residência habitual propriedade dos ocupantes (2021_PAFPO);
- X_2 corresponde à proporção de alojamentos familiares construídos antes de 1960 (2021_PAF_ANTES1960);
- X_3 corresponde à proporção de alojamentos familiares de uso sazonal (2021_PAFSAZ);
- X_4 corresponde à proporção de alojamentos familiares com aquecimento central (2021_PAFAQCENT);
- X_5 corresponde à proporção de população residente com o ensino superior completo (2021_PPOPRESSENSUP);
- X_6 corresponde à duração média dos movimentos pendulares da população residente empregada ou estudante (2021_DURMOV PEND);
- X_7 corresponde à taxa de desemprego (2021_TX_DESEMP);
- X_8 corresponde à distância à classe da COS “Praias, dunas e areais costeiros” mais próxima;
- (u_i, v_i) refere-se às coordenadas dos centroides de cada freguesia;
- β_{bw} representa os coeficientes das variáveis explicativas ajustados de acordo com a largura de banda (bw) de cada uma;
- ε_i corresponde ao erro residual.

5.3.1. LARGURAS DE BANDA E ESTATÍSTICAS DOS COEFICIENTES

Tal como já evidenciado, uma das principais características do modelo MGWR consiste no facto de este permitir a designação de uma largura de banda diferente para cada uma das variáveis independentes, atendendo à heterogeneidade das relações entre os fatores explicativos e a variável dependente. Para tal, a Tabela 5.3.1 expõe as larguras de banda ótimas usadas no modelo.

Quanto às larguras de banda individuais, observa-se uma relação global entre a duração dos movimentos pendulares e o valor médio das rendas, visto que a largura de banda inclui todas as freguesias da área de estudo. A proporção de alojamentos familiares propriedade dos ocupantes, a proporção de alojamentos familiares de uso sazonal, e a taxa de desemprego aparentam operar a uma escala regional com a variável dependente, já que a vizinhança ótima representa, respetivamente, 58,47%, 82,20% e 50%, das freguesias da AML. As restantes variáveis (2021_PAF_ANTES1960; 2021_PAFAQCENT; 2021_PPOPRESSENSUP; e NEAR_DIST_COS2018_PRAIAS) revelam uma relação mais local com a variável dependente, com larguras de banda mais reduzidas (<30%).

A mesma Tabela 5.3.1 permite igualmente analisar as estatísticas dos coeficientes. Naturalmente, uma vez que larguras de banda mais amplas frequentemente resultam em menor variabilidade dos coeficientes, a duração dos movimentos pendulares é a variável que apresenta menor desvio-padrão (0,0035).

Tabela 5.3.1 Larguras de banda e estatísticas dos coeficientes para o modelo de 2021

Variável (<i>scaled</i>)	Largura de banda		Freguesias com coeficiente significativo		Média	Desvio-padrão	Mínimo	Mediana	Máximo
	N.º	%	N.º	%					
<i>Intercept</i>	97	82,20	107	90,68	0,2135	0,0369	0,1107	0,2282	0,2662
2021_PAFPO	69	58,47	87	73,73	0,2474	0,0593	0,0504	0,2558	0,3868
2021_PAF_ANTES1960	33	27,97	67	56,78	0,3712	0,1856	-0,0545	0,4453	0,6125
2021_PAFSAZ	97	82,20	80	67,80	0,1460	0,0478	0,0557	0,1519	0,2118
2021_PAFQCENT	31	26,27	9	7,63	-0,0537	0,1749	-0,4673	-0,0611	0,2008
2021_PPOPREENSSUP	35	29,66	117	99,15	0,5617	0,1632	0,3112	0,5069	0,9793
2021_DURMOVPEND	118	100,00	9	7,63	0,1250	0,0035	0,1154	0,1248	0,1347
2021_TX_DESEMP	59	50,00	53	44,92	-0,2122	0,1224	-0,3848	-0,2514	0,0688
NEAR_DIST_COS2018_PRAIAS	30	25,42	48	40,68	0,1088	0,3806	-0,5159	0,1293	0,8693

Em média, a variável 2021_PPOPREENSSUP é a que mais influencia os preços das rendas, apresentando um total de 117 freguesias (99,15%) com coeficiente significativo. Além disso, uma vez que tanto o coeficiente mínimo como o máximo são positivos, pode-se inferir que quanto maior a proporção de população residente com o ensino superior completo, mais inflacionados serão os preços das rendas. Opostamente, tanto a proporção de alojamentos com aquecimento central, como a duração dos movimentos pendulares apresentam menor influência sobre os valores das rendas, observando-se apenas 9 freguesias com coeficiente significativo em ambas as variáveis.

5.3.2. MULTICOLINEARIDADE LOCAL

A Figura 5.3.1 representa a distribuição espacial dos valores de CN, permitindo testar a multicolinearidade local do modelo. Uma vez que todas as freguesias apresentam valores de CN inferiores a 30, é possível aferir que a multicolinearidade local não será um problema neste modelo.

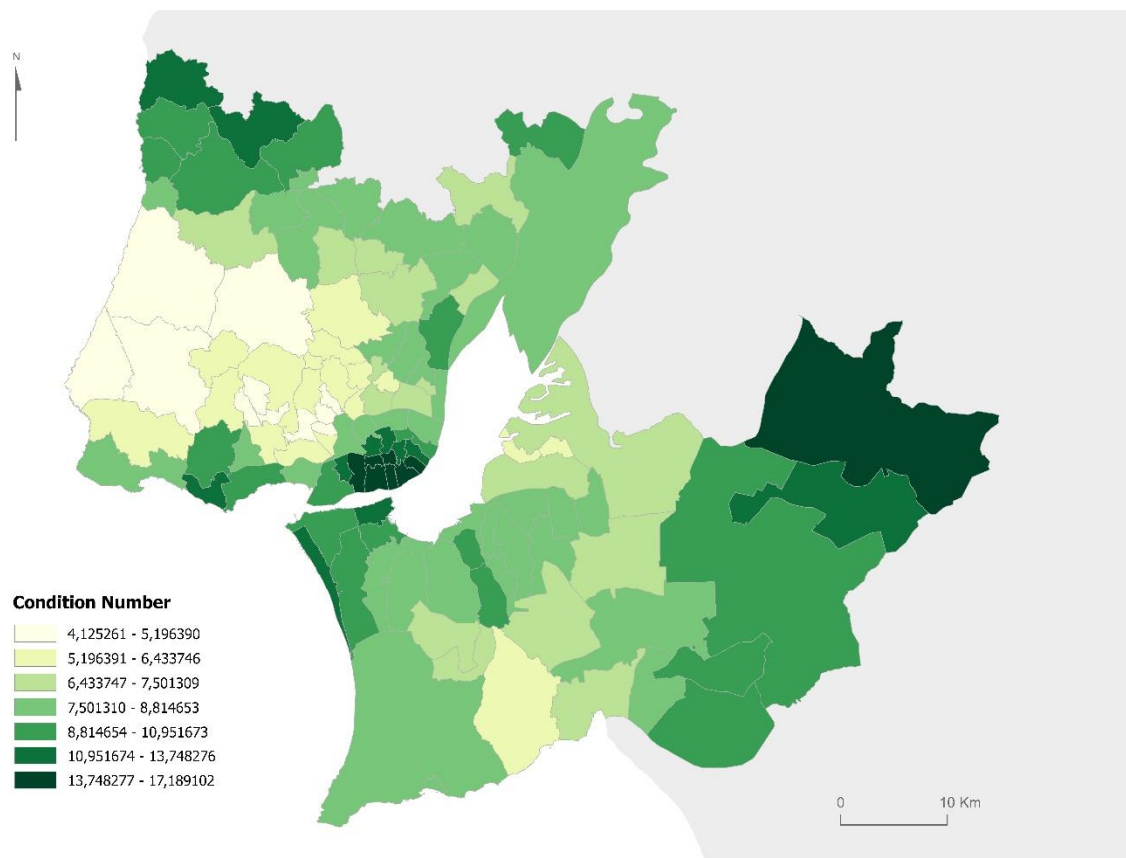


Figura 5.3.1 Multicolinearidade local no modelo de 2021

5.3.3. OBSERVAÇÕES INFLUENTES

De modo a testar a existência de observações influentes, foi mapeada a distância de Cook (Figura 5.3.2). Deste modo, é possível concluir que as observações influentes não são um problema neste modelo, uma vez que nenhuma freguesia apresenta um Cook's D > 1.

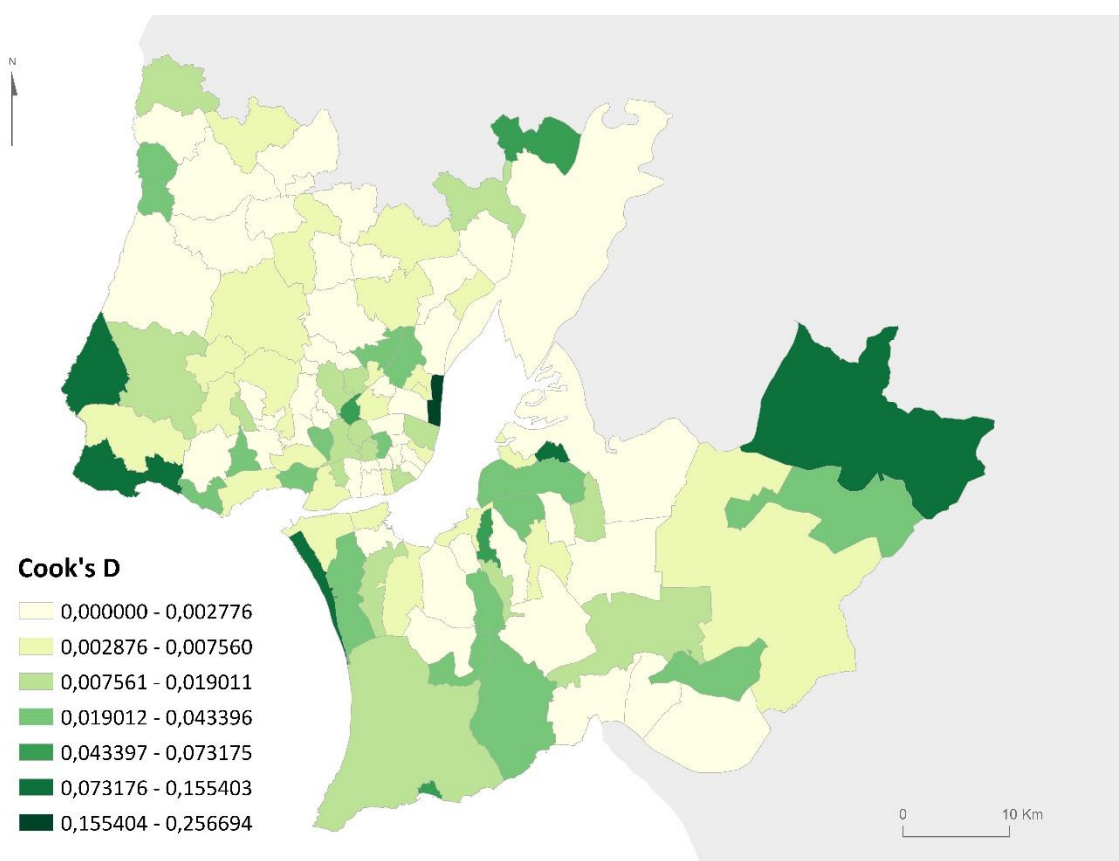


Figura 5.3.2 Distância de Cook no modelo de 2021

5.3.4. ADEQUAÇÃO DO MODELO

Ao comparar as medidas de adequação dos modelos GWR e MGWR obtidas através da execução do modelo de 2021 (Tabela 5.3.2), verifica-se que o modelo MGWR revela melhor adequação, uma vez que simultaneamente apresenta um Adj- R^2 superior (0,8779, face aos 0,8158 do modelo GWR) e um AICc substancialmente inferior (151,3131, comparativamente aos 173,6824 do modelo GWR).

Tabela 5.3.2 Medidas de adequação do modelo de 2021

Modelo	R^2	Adj- R^2	AICc	ENP	Graus de Liberdade Efetivos
GWR 8 variáveis	0,8740	0,8158	173,6824		81,0536
MGWR 8 variáveis	0,9150	0,8779	151,3131	35,57	82,4193

Adicionalmente, a Figura 5.3.3 mostra que, na área de estudo, o R^2 varia entre 0,7752 e 0,9448, sendo que, na área da Grande Lisboa, o modelo é mais preciso no centro urbano de Lisboa, estendendo-se para Norte e, à medida que se afasta para Oeste (Sintra), o R^2 diminui. Na Península de Setúbal este padrão inverte: o modelo apresenta melhor precisão na União das Freguesias de Almada, Cova da Piedade, Pragal e Cacilhas, na União das Freguesias de Laranjeiro e Feijó, e na União das Freguesias de Barreiro e Lavradio, diminuindo o R^2 em direção a Sudeste.

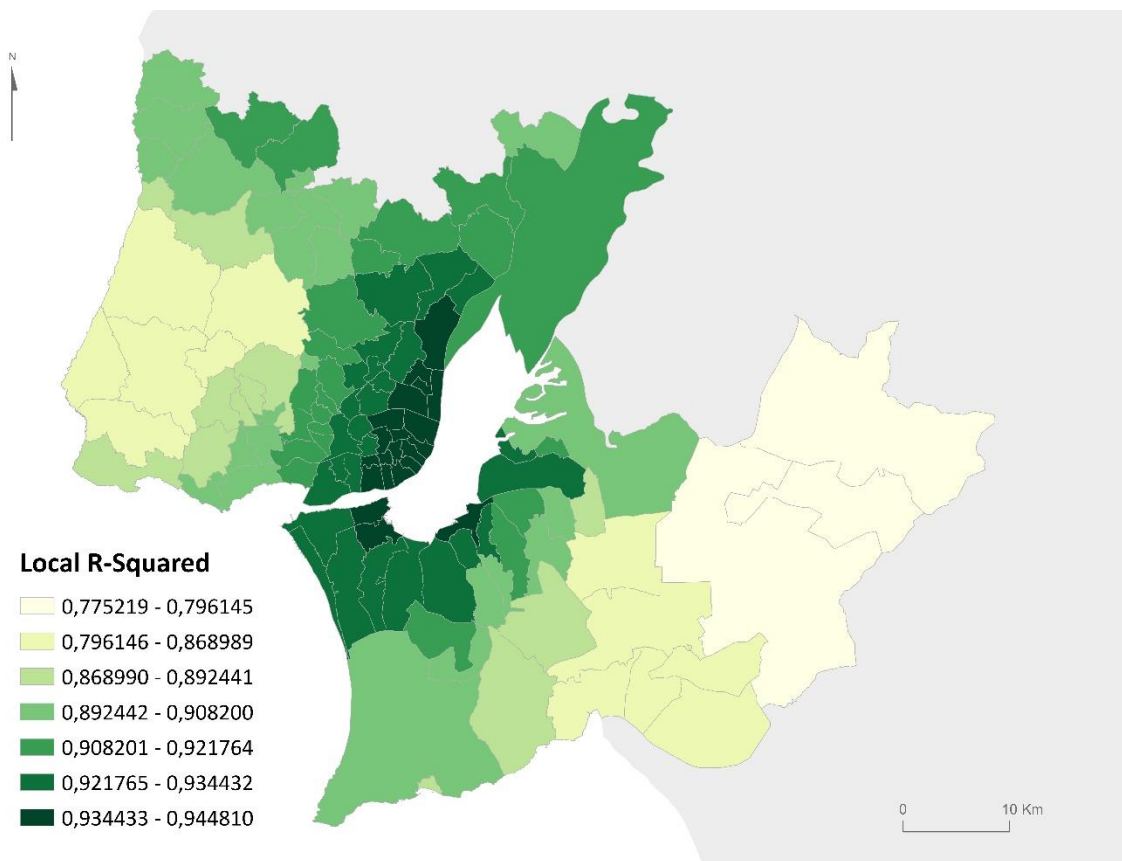


Figura 5.3.3 Local R-squared no modelo de 2021

5.3.5. ANÁLISE DE RESÍDUOS

Finalmente, a última etapa de diagnóstico assenta na análise dos resíduos do modelo. Tanto a distribuição espacial, representada na Figura 5.3.4, como o teste de *Global Moran's I* permitem concluir que não existe autocorrelação espacial nos resíduos. Com um índice de -0,064965 e um valor de p (0,220233) superior ao nível de significância de 0,05, o teste de *Global Moran's I* indica que os resíduos apresentam uma distribuição aleatória.

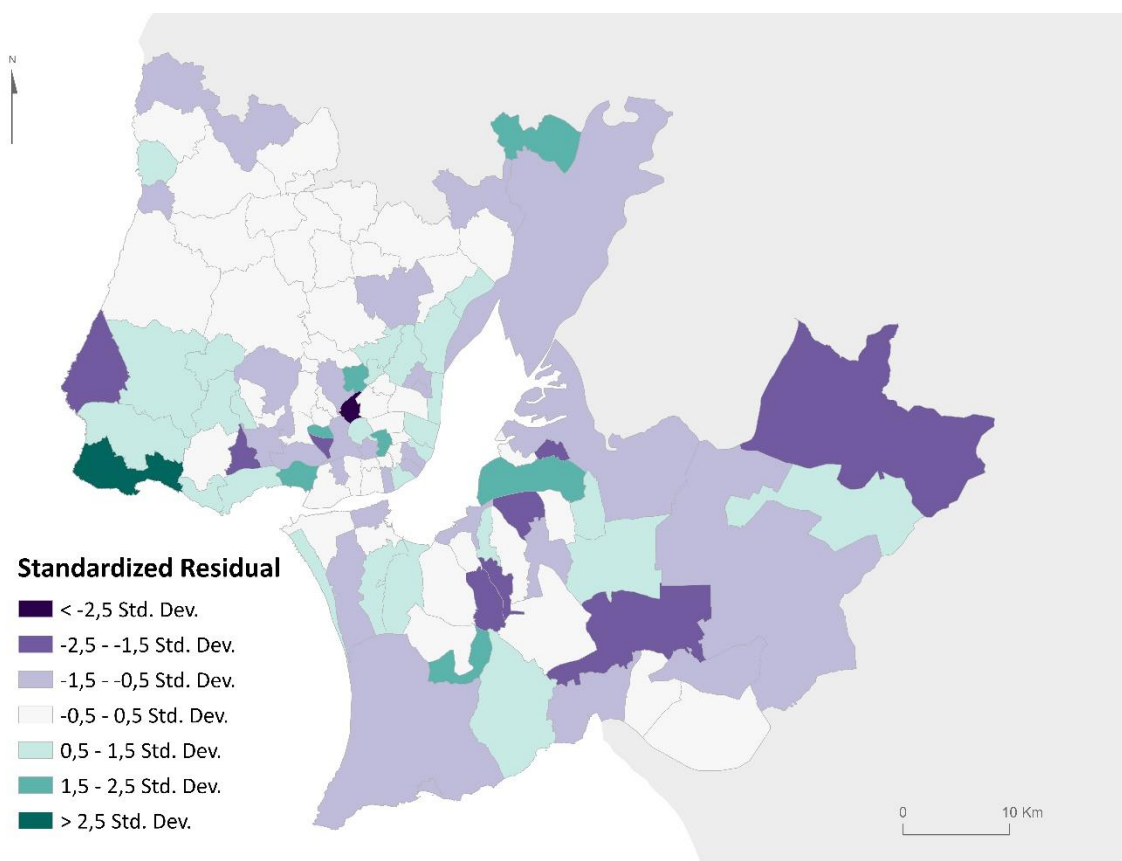


Figura 5.3.4 Resíduos estandardizados no modelo de 2021

5.3.6. COEFICIENTES LOCAIS

Por forma a analisar a espacialidade da influência de cada fator explicativo sobre a variável dependente, foram mapeados os coeficientes estimados para cada variável independente utilizada neste modelo de 2021 (Figura 5.3.5).

Tal como já constatado pela análise das estatísticas dos coeficientes, a duração dos movimentos pendulares é uma das variáveis que menos influencia os preços das rendas, com coeficientes significantes que variam entre 0 e 0,3853, somente na faixa Sul da área de estudo (Sesimbra, Setúbal e Palmela). Os coeficientes positivos em toda a área de estudo revelam que, movimentos pendulares mais longos tendem a estar associados a rendas mais caras, pressupondo-se que as áreas mais periféricas tendem a ser mais valorizadas. Por sua vez, a proporção de alojamentos com aquecimento central é igualmente uma variável em que a significância dos coeficientes é diminuta. No entanto, observam-se quatro uniões de freguesias em Sintra (Aqualva e Mira-Sintra, Massamá e Monte-Abraão, Queluz e Belas, e Cacém e São Marcos), cujo coeficiente é estatisticamente significativo e negativo, com valores entre -1,0634 e -0,3853, circundadas por freguesias com coeficiente igualmente significativo e negativo, mas com valores mais próximos de 0.

A proporção de alojamentos familiares propriedade dos ocupantes tem uma influência positiva significativa em toda a Grande Lisboa e extremo Noroeste da Península de Setúbal (Almada), destacando-se a União das

Freguesias de Sintra (Santa Maria e São Miguel, São Martinho e São Pedro de Penaferrim) e a freguesia de Colares, onde o coeficiente é mais elevado.

Além disso, uma maior proporção de alojamentos familiares construídos antes de 1960 aparenta estar associada a valores de rendas mais elevados na maior parte da AML, sendo que os coeficientes positivos significantes se concentram predominantemente no centro urbano de Lisboa e em grande parte das freguesias da margem Sul do rio Tejo (nos municípios de Almada, Seixal, Barreiro, Moita, Montijo e Palmela).

Por sua vez, a proporção de alojamentos de uso sazonal tende a influenciar positivamente os valores das rendas, apresentando coeficiente significantes no município de Lisboa e freguesias adjacentes (Odivelas, Amadora, Oeiras e Cascais), bem como em quase toda a Península de Setúbal.

Estando a variável 2021_PPOPREENSSUP intrinsecamente relacionada com os rendimentos familiares, é natural que valores mais elevados estejam associados a rendas mais elevadas também. De facto, observa-se que o efeito da proporção de população residente com o ensino superior completo é bastante visível em todo o território, com significância em 99,15% das freguesias, sendo que a maioria dos coeficientes detém valores entre 0,3853 e 1,0364.

Contrariamente, elevadas taxas de desemprego tendem a desvalorizar as rendas, sendo este fenómeno estatisticamente significativo nas freguesias do Sul da Grande Lisboa (municípios de Lisboa, Amadora, Oeiras, Cascais e Sintra), e no Noroeste da Península de Setúbal, nas freguesias de Amora e Corroios, bem como nas Uniões das Freguesias de: Almada, Cova da Piedade, Pragal e Cacilhas; Caparica e Trafaria; Charneca de Caparica e Sobreda; e Laranjeiro e Feijó.

Por fim, os coeficientes da variável NEAR_DIST_COS2018_PRAIAS revelam impactos distintos da distância às praias nos preços das rendas. Na Grande Lisboa, os coeficientes são predominantemente negativos, à exceção das freguesias do município de Lisboa e algumas adjacentes, onde os coeficientes são positivos, mas não significantes. No Norte da AML (Mafra, Loures e Vila Franca de Xira), na união das freguesias de Cascais e Estoril, bem como na freguesia de Alcabideche, quanto maior a distância às praias, menor serão os valores das rendas. Contudo, este padrão inverte-se quando se observa a região a Sul do rio Tejo, onde uma maior distância às praias reflete-se em valores de rendas mais valorizados, com coeficientes estatisticamente significantes em quase toda a Península de Setúbal, à exceção dos extremos Noroeste (Almada) e Sudeste (Setúbal, Palmela e Montijo).

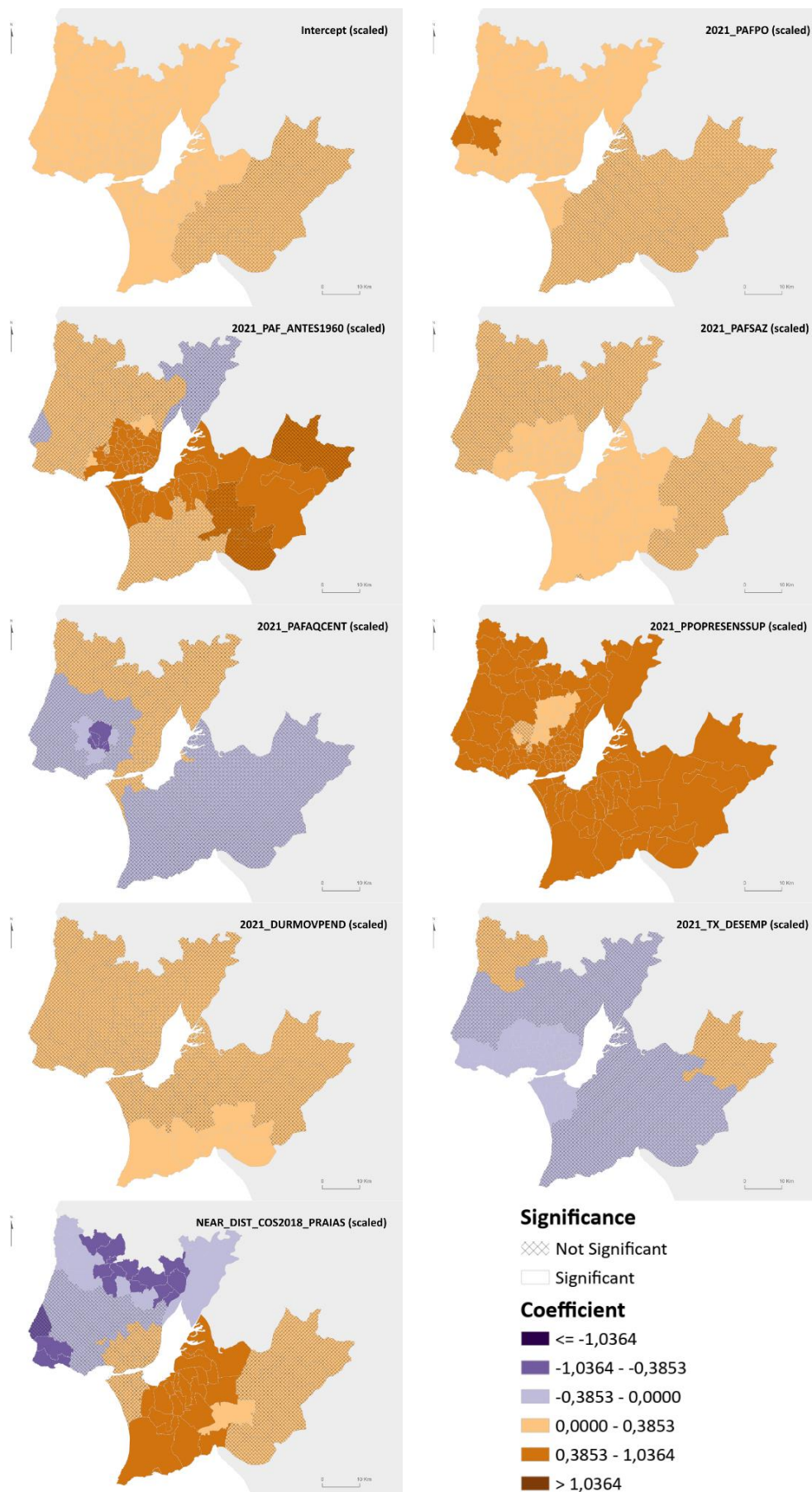


Figura 5.3.5 Coeficientes locais do modelo de 2021

6. DISCUSSÃO DE RESULTADOS

Tendo por base as análises realizadas nos capítulos supra, esta secção tem por objetivo interpretar os resultados obtidos nos dois modelos MGWR de 2001 e 2021 através de uma reflexão sobre as variações espaciais e temporais que ocorreram nos fatores que explicam os preços das rendas na AML. Primeiramente, são comparados os modelos (6.1), destacando-se as diferenças no ajuste e adequação, bem como as mudanças dos fatores que melhor explicam a variável dependente. De seguida, aprofunda-se a análise dos efeitos da gentrificação (6.2), detalhando as alterações que ocorreram nas variáveis que caracterizam este fenómeno.

Contudo, uma questão importante que deve ser tida em conta nesta análise comparativa entre 2001 e 2021 prende-se com as alterações nos limites administrativos das freguesias, decorrentes da reforma administrativa de 2013, que em muito dificultam a interpretação e comparação dos resultados. Atente à impossibilidade de uma comparação fiável entre freguesias, este capítulo tenta compreender as alterações visíveis a nível regional, municipal e, sempre que possível, local, ocorridas neste período de duas décadas.

6.1. COMPARAÇÃO DOS MODELOS (2001-2021)

A evolução do mercado imobiliário e das dinâmicas urbanas ao longo das últimas décadas na AML justifica a necessidade de comparar os modelos de 2001 e 2021. Através desta análise, pretende-se compreender não apenas como os fatores que influenciam os preços das rendas se modificaram ao longo do tempo, mas também avaliar a adequação dos modelos utilizados para captar tais mudanças.

Com base nas medidas de adequação presentes na Tabela 6.1.1, constata-se que, de um modo geral, o modelo de 2021 apresenta melhor qualidade de ajuste, isto é, capta melhor os padrões espaciais dos valores das rendas, com menos variáveis e menor complexidade. Tanto o R^2 , como o R^2 ajustado são mais elevados no modelo de 2021 ($R^2 = 0,9150$; $\text{Adj-}R^2 = 0,8779$), do que no modelo de 2001 ($R^2 = 0,8158$; $\text{Adj-}R^2 = 0,7479$), o que significa que o modelo mais recente explica, com as suas 8 variáveis independentes, uma maior proporção da variabilidade dos preços das rendas. Adicionalmente, o AICc decresceu substancialmente de 393,5349 (2001) para 151,3131 (2021), indicando um melhor equilíbrio entre adequabilidade e complexidade por parte do modelo de 2021. Finalmente, o número efetivo de parâmetros (ENP) e os graus de liberdade efetivos, significativamente mais reduzidos no modelo de 2021, sugerem que menos parâmetros foram necessários para alcançar um maior ajuste, o que se reflete numa maior estabilidade do modelo. Não obstante, a melhor adequação do modelo de 2021 poderá estar relacionada com uma maior capacidade explicativa das variáveis selecionadas, não se podendo excluir a possibilidade de omissão de alguma variável importante que poderia ter tido melhor capacidade explicativa sobre os preços das rendas em 2001.

Tabela 6.1.1 Comparação da adequação dos modelos de 2001 e 2021

Modelo	R^2	$\text{Adj-}R^2$	AICc	ENP	Graus de Liberdade Efetivos
MGWR 2001 (9 variáveis)	0,8158	0,7479	393,5349	55,45	151,5568
MGWR 2021 (8 variáveis)	0,9150	0,8779	151,3131	35,57	82,4193

Recuperando a classificação das tipologias de fatores que afetam os preços da habitação anteriormente referida, a Tabela 6.1.2 comprova a complexidade do mercado de arrendamento, sendo fundamental incluir variáveis das várias categorias para corretamente explicar a variação dos preços das rendas. Contudo, evidenciam-se algumas mudanças nas variáveis que foram utilizadas nos dois modelos.

As variáveis representativas das características estruturais das habitações sofreram alterações significativas. Por um lado, o número de divisões por alojamento familiar que, em 2001 aparentava ter correlação local positiva linear com a variável dependente, foi substituída, no modelo de 2021, pela proporção de alojamentos familiares com aquecimento central. Esta alteração retrata as constantes mudanças nas preferências dos consumidores. Por outro lado, em 2001, o modelo considerava a proporção de edifícios construídos nos últimos 10 anos como um fator relevante, possivelmente relacionado com a influência da expansão urbana, mas em 2021, esta variável foi trocada pela proporção de alojamentos familiares clássicos de residência habitual construídos antes de 1960, atribuindo-se maior importância às construções mais antigas. Este fenómeno poderá estar associado a estratégias de reabilitação urbana (Li et al., 2022), que priorizam a renovação de edifícios antigos, em detrimento das novas construções, tal como pode ser comprovado pelo decréscimo substancial do número de fogos concluídos em construções novas para habitação familiar: de acordo com os dados do INE⁷, em Portugal, este número diminuiu de 115 154, em 2001, para 19 565, em 2021.

Tabela 6.1.2 Variáveis explicativas dos modelos de 2001 e 2021, por categoria de determinantes dos preços das habitações

Categoria	Variáveis explicativas do modelo de 2001	Variáveis explicativas do modelo de 2021
Características Estruturais	Divisões por alojamento familiar clássico de residência habitual (2001_NDIV)	Proporção de alojamentos familiares com aquecimento central (2021_PAFAQCENT)
	Proporção de edifícios construídos nos últimos 10 anos (2001_PED_10ANOS)	Proporção de alojamentos familiares clássicos de residência habitual construídos antes de 1960 (2021_PAF_ANTES1960)
Características de Vizinhança	Proporção de alojamentos superlotados (2001_PASUPERLOT)	
	Proporção da população residente com ensino superior completo (2001_PPOPRESSENSUP)	Proporção da população residente com ensino superior completo (2021_PPOPRESSENSUP)
	Taxa de desemprego (2001_TX_DESEMP)	Taxa de desemprego (2021_TX_DESEMP)
	Percentagem de alojamentos familiares clássicos de residência habitual propriedade dos ocupantes (2001_PAFPO)	Percentagem de alojamentos familiares clássicos de residência habitual propriedade dos ocupantes (2021_PAFPO)
	Proporção da população residente com 65 ou mais anos (2001_PPOPRES_65MAIS)	Proporção de alojamentos familiares clássicos de uso sazonal (2021_PAFSAZ)
Atributos de Localização	Duração média dos movimentos pendulares da população residente empregada ou estudante (2001_DURMOV PEND)	Duração média dos movimentos pendulares da população residente empregada ou estudante (2021_DURMOV PEND)
	Distância à classe da COS "Praias, dunas e areais costeiros" mais próxima (NEAR_DIST_COS2007_PRAIAS)	Distância à classe da COS "Praias, dunas e areais costeiros" mais próxima (NEAR_DIST_COS2018_PRAIAS)

⁷ Consultado no dia 03 de abril de 2025, no website oficial do Instituto Nacional de Estatística: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0000076&contexto=bd&selTab=tab2.

Relativamente às características de vizinhança, a proporção de população residente com o ensino superior completo, a taxa de desemprego e a percentagem de alojamentos familiares clássicos de residência habitual propriedade dos ocupantes mantêm-se como fatores explicativos relevantes em ambos os anos. Tanto Lee et al. (2016), como Ma & Gopal (2018) indicam que a escolaridade da população residente tende a influenciar positivamente os preços da habitação e, de facto, observa-se claramente esta relação em ambos os modelos, especialmente em 2021, onde os coeficientes da variável 2021_PPOPREENSSUP são estatisticamente significantes e positivos em 99,15% da área de estudo. Tal como Fotheringham & Sachdeva (2022) e Al-Masum & Lee (2019) também concluem, em ambos os modelos, os preços das rendas revelam ser negativamente influenciadas pelas taxas de desemprego, isto é, locais com elevadas taxas de desemprego tendem a apresentar valores médios das rendas mais reduzidos. Para além disso, Ma & Gopal (2018) verificam, no seu estudo, que a taxa de desemprego revela ser, consistentemente ao longo do tempo, uma variável estatisticamente significativa na explicação dos padrões de valorização do mercado habitacional. Os resultados obtidos no presente trabalho corroboram esta constatação, evidenciando que, tanto em 2001 como em 2021, a taxa de desemprego surge como um fator estatisticamente relevante na determinação dos valores médios das rendas, reforçando a premissa de que o desemprego funciona como um indicador estrutural das dinâmicas socioeconómicas do território. Por sua vez, a percentagem de alojamentos familiares clássicos de residência habitual propriedade dos ocupantes, embora se mantenha um fator explicativo importante, revela alterações significativas nos padrões espaciais da sua influência, o que poderá confirmar a importância do regime de ocupação dos alojamentos como característica da gentrificação. Adicionalmente, verifica-se que tanto a proporção de alojamentos superlotados, como a proporção de população com 65 ou mais anos deixam de ser relevantes para o modelo mais recente e são substituídas pela proporção de alojamentos de uso sazonal, que revela ter grande influência nos preços das rendas, com coeficientes significantes em 67,80% das freguesias.

Tal como vários autores defendem (Bitter, 2008; Droj et al., 2024; Fotheringham et al., 2015), os atributos de localização assumem um importante papel na explicação dos preços da habitação. Neste caso, ambas as variáveis - duração média dos movimentos pendulares e distância às praias – apresentam uma influência significativa nos dois modelos.

6.2. OS EFEITOS DA GENTRIFICAÇÃO

Uma vez que este estudo tenta focar-se nos impactos da gentrificação no mercado de arrendamento, o presente subcapítulo examina as mudanças que ocorreram entre o modelo de 2001 e o de 2021, em particular nas variáveis que poderão, direta ou indiretamente caracterizar a gentrificação. Neste contexto, a gentrificação consiste num processo de revalorização imobiliária em áreas urbanas, geralmente associado à substituição gradual da população residente de menor rendimento por grupos com maior capacidade económica, resultando em alterações significativas na composição socioeconómica e no tecido urbano local. Não obstante, a complexidade deste fenómeno e a dificuldade em defini-lo exatamente (Finio, 2022), leva a que seja necessária uma interpretação conjunta de várias variáveis. Atente às diferenças que se observam na distribuição espacial

dos coeficientes e nas estatísticas das larguras de banda e coeficientes das variáveis comuns aos dois modelos, é possível inferir algumas possíveis influências do fenómeno da gentrificação.

Em primeiro lugar, verifica-se que a proporção de alojamentos familiares propriedade dos ocupantes apresentava, em 2001, coeficientes significantes em 27,54% das freguesias, valor que aumentou para 73,73% em 2021. Além deste aumento direto na influência desta variável, observa-se, igualmente, uma alteração nos padrões espaciais da sua influência: passa de uma influência regionalizada, com coeficientes significantes positivos nos municípios periféricos ao centro urbano de Lisboa e na faixa Oeste da Península de Setúbal, para uma influência igualmente positiva no município de Almada e em toda a Grande Lisboa, onde se destacam apenas duas localidades do município de Sintra (freguesia de Colares e União das Freguesias de Sintra - Santa Maria e São Miguel, São Martinho e São Pedro de Penaferrim), com coeficientes mais elevados. Contudo, apesar do aumento das freguesias com coeficiente significativo e da tendência de uma influência positiva em todo o território, de 2001 para 2021, observa-se uma ligeira diminuição nos valores dos coeficientes, nomeadamente a média diminuiu de 0,3638 para 0,2474. Este fenómeno pode, à primeira vista, parecer contraditório, dado que a variável ganha expressão em mais freguesias, mas simultaneamente apresenta uma relação menos intensa com os valores das rendas. No entanto, quando contextualizada à luz da literatura sobre gentrificação, esta constatação adquire um significado revelador. Como referem vários autores (Álvarez-Herranz & Macedo-Ruiz, 2021; Carvalho et al., 2019), um dos traços estruturantes da gentrificação é precisamente a transformação no regime de ocupação da habitação, marcada pela substituição de residentes proprietários por arrendatários. Assim, a redução da magnitude dos coeficientes pode estar a refletir tal reconfiguração do mercado habitacional no contexto da AML, em que se estará a assistir a uma crescente predominância dos regimes de arrendamento, em detrimento da habitação própria, que perde peso como determinante dos preços. Conquanto, mantém-se a análise que, de um modo geral, proporções elevadas de alojamentos propriedade dos ocupantes tendem a inflacionar os valores das rendas.

Por sua vez, a forte associação entre a proporção de população com ensino superior completo e o aumento das rendas é um fenómeno amplamente documentado nos processos de gentrificação (Hawkins et al., 2022). Esta variável ilustra as transformações socioeconómicas dos espaços urbanos, refletindo a substituição gradual de populações historicamente mais vulneráveis, como idosos ou indivíduos com baixos rendimentos, por novos com maiores capacidades financeiras. Estes últimos, geralmente detentores de níveis de escolaridade mais elevados, possuem maior capacidade financeira para suportar os aumentos significativos nos valores das rendas, contribuindo para a valorização do território e para a exclusão habitacional dos grupos sociais pré-existentes. Este facto é ilustrado nos modelos desenvolvidos pelo aumento muito expressivo da percentagem de freguesias com coeficiente significativo, de 14,01% em 2001, para 99,15% em 2021, que evidencia a crescente influência da proporção de população residente com ensino superior na estruturação do mercado de arrendamento na AML.

Embora diretamente não se associe a duração dos movimentos pendulares aos efeitos da gentrificação, a ligeira redução da proporção de freguesias com coeficiente significativo (de 9,66%, em 2001, para 7,63%, em 2021) pode reforçar a ideia de uma metrópole em reconfiguração, especialmente de uma deslocação da população para as

zonas periféricas e de uma melhoria na rede de transportes públicos. Inicialmente a duração dos movimentos pendulares tinha, em média, um efeito negativo, isto é, quanto menor a duração dos movimentos, mais elevados os valores das rendas, salvo as localidades da Grande Lisboa periféricas ao centro urbano, onde a relação se invertia. Contudo, em 2021, os coeficientes são sempre positivos, significantes apenas no Sul da área de estudo.

A variável distância às praias revela, também, um aumento da proporção de freguesias com coeficiente significativo, de 19,32% em 2001, para 44,92% em 2021, concordante com a premissa de que o fenómeno da gentrificação está associado ao aumento do turismo. Contudo, esta variável apresenta alterações significativas na distribuição espacial da sua influência. Em 2001, todas as freguesias com coeficiente significativo apresentavam um valor negativo, ou seja, quanto maior a distância às praias, menores seriam os valores das rendas, tal como constatado por Fotheringham & Sachdeva (2022). No entanto, em 2021, observa-se um acentuar das discrepâncias entre as zonas Norte e Sul da AML: na Grande Lisboa os coeficientes mantêm-se negativos, à exceção do centro urbano de Lisboa, onde a influência desta variável não é estatisticamente significativa; na Península de Setúbal os coeficientes são positivos, revelando uma tendência de desvalorização imobiliária com a proximidade às praias. Esta última constatação poderá estar relacionada com o fraco desenvolvimento da rede de transportes públicos na Península de Setúbal, acompanhado por uma menor expansão suburbana destas zonas (Lestegás, 2019).

No seguimento da influência do turismo como fator intrinsecamente ligado à gentrificação em Lisboa, importa destacar a influência da proporção de alojamentos familiares de uso sazonal. Esta variável que, em 2001, não aparentava ter qualquer relação com os valores médios das rendas, em 2021 passou a ser significativa em 67,80% das freguesias da AML. Esta evolução pode ser interpretada como reflexo do crescimento da pressão turística sobre o mercado habitacional, com o aumento do alojamento local e a reconfiguração de áreas residenciais para fins turísticos (Carvalho et al., 2019; Franco & Santos, 2021).

Não obstante, os padrões espaciais da influência da distância às praias e da proporção de alojamentos de uso sazonal revelam que o impacto do turismo se alastra por toda a área metropolitana e não se concentra somente no centro urbano de Lisboa, onde tradicionalmente se percecionaria o impacto da gentrificação.

Finalmente, a perda de significância estatística da variável que representa a proporção de população com 65 ou mais anos, entre 2001 e 2021, revela uma possível deslocação progressiva da população idosa das zonas com maior pressão imobiliária. Em 2001, esta variável apresentava uma relação negativa estatisticamente significativa com os valores das rendas na totalidade das freguesias da AML. No entanto, no modelo de 2021 esta variável não é sequer incluída, pois não apresenta qualquer relação linear local com a variável dependente. Esta ausência de significância em 2021 sugere que esta faixa etária perdeu centralidade explicativa na estrutura espacial dos preços, possivelmente por ter sido gradualmente substituída por outros grupos populacionais, nomeadamente indivíduos mais jovens, com maior nível de escolaridade e poder de compra, fenómeno coerente com os efeitos esperados da gentrificação (Hawkins et al., 2022; Wilhelmsson et al., 2021).

7. CONCLUSÕES E INVESTIGAÇÃO FUTURA

De um modo geral este trabalho atesta a complexidade do mercado de habitação e, em particular do mercado de arrendamento, alertando para a necessidade de compreender os fatores que influenciam os preços das habitações, bem como a sua variação no tempo e no espaço. Por forma a responder à questão de investigação previamente estabelecida - “De que forma os processos de gentrificação influenciam os preços da habitação em Lisboa, e quais são os fatores socioeconómicos e urbanísticos determinantes deste fenómeno?” -, esta dissertação seguiu um processo metodológico estruturado que visava cumprir os vários objetivos estipulados, utilizando técnicas de modelação geoespacial.

O primeiro objetivo consistiu em distinguir as variáveis quantitativas que caracterizam a gentrificação. Para cumprir este propósito, foi realizada uma revisão da literatura crítica que permitiu identificar um conjunto de variáveis habitualmente associadas ao fenómeno da gentrificação, como o aumento do nível de escolaridade, o aumento do turismo urbano ou a mudança na estrutura da propriedade dos alojamentos. Nos modelos desenvolvidos, destacaram-se três variáveis que representam diretamente a influência da gentrificação: a percentagem de alojamentos familiares clássicos de residência habitual propriedade dos ocupantes, a proporção da população residente com 65 ou mais anos e, a proporção de alojamentos familiares de uso sazonal. No entanto, estas variáveis não devem ser analisadas isoladamente, sendo a sua interpretação mais robusta quando em articulação com outras dimensões, como a proximidade às praias ou a proporção da população com ensino superior completo, que também revelam tendências de transformação urbana associadas ao processo de gentrificação.

O segundo objetivo passou por testar a existência de autocorrelação e heterogeneidade espacial no mercado de arrendamento da AML. Através da análise exploratória dos dados espaciais, que recorreu ao uso de mapas temáticos, testes de *Local e Global Moran's I*, bem como a estatística de *Getis-Ord Gi*, foi possível confirmar, para ambos os anos em análise, a existência de padrões espaciais bem definidos na distribuição das rendas. Estes resultados validam o pressuposto de que os valores das rendas não estão distribuídos aleatoriamente no espaço, mas sim sujeitos à influência heterogénea de vários fatores espaciais, socioeconómicos e urbanos.

No que respeita ao terceiro objetivo, a identificação dos fatores que mais influenciam os preços das rendas, foi alcançada através da modelação MGWR. De um modo geral, os resultados indicam que, em 2001, os principais fatores explicativos significantes foram: a proporção de alojamentos familiares propriedade dos ocupantes, a proporção de edifícios construídos nos últimos dez anos, a percentagem da população com 65 ou mais anos, a proporção da população residente com o ensino superior completo, a duração média dos movimentos pendulares, a taxa de desemprego e a distância às praias mais próximas. Em 2021, observou-se uma evolução nas variáveis explicativas, destacando-se a proporção de alojamentos construídos antes de 1960, a proporção de alojamentos de uso sazonal e a proporção de alojamentos com aquecimento central, mantendo-se outros fatores como a taxa de desemprego, a escolaridade e a duração dos movimentos pendulares. Esta mudança no perfil

das variáveis evidencia, não só as constantes mudanças nos padrões de consumo, mas também a transformação estrutural e funcional da AML ao longo destas duas décadas.

O quarto objetivo consistiu na construção de um modelo hedónico adequado para avaliar a influência destes fatores no preço das rendas. Através dos dois modelos MGWR desenvolvidos, foi possível analisar a variabilidade espacial dos coeficientes, permitindo perceber que o impacto das variáveis explicativas não é homogéneo em toda a região. Por conseguinte, estes dois modelos permitiram compreender de forma mais detalhada as dinâmicas locais, tendo em atenção a complexidade do mercado de arrendamento e a diversidade urbana da AML.

Por fim, o quinto objetivo prendeu-se com a análise da relação entre a gentrificação e a variação espacial dos preços da habitação. Embora os modelos tenham incluído variáveis representativas da gentrificação, a identificação direta deste processo exige uma abordagem mais longitudinal e integrada. Ainda assim, os resultados sugerem uma alteração significativa nos fatores que influenciam os valores das rendas, o que aponta para uma possível influência deste fenómeno no processo de valorização imobiliária.

Não obstante, este estudo não é isento de limitações que, na verdade, mais do que insuficiências, devem ser reconhecidas como oportunidades para futuras investigações e desenvolvimentos.

Em primeiro lugar, deve-se sublinhar que os modelos aqui desenvolvidos não retratam uma análise temporal, mas sim apenas uma comparação entre dois anos (2001 e 2021) através de dois modelos independentes. A inexistência de um conjunto de dados contínuos ao longo do tempo impede uma análise evolutiva mais fina e rigorosa. Por outro lado, a escolha de comparação de dois momentos específicos que se distanciam temporalmente em 20 anos limita a capacidade de isolar a gentrificação como principal fator de mudança, uma vez que outros acontecimentos marcantes, como a crise financeira de 2007/2008 ou a pandemia da Covid-19, poderão ter tido, igualmente, um impacto significativo nas dinâmicas do mercado de habitação. A escolha de um intervalo temporal de duas décadas, embora metodologicamente justificável, acarreta riscos de omissão de tendências relevantes. Neste sentido, futuras investigações poderiam beneficiar da utilização de séries temporais mais longas, que permitam uma análise longitudinal da evolução dos preços do arrendamento e da transformação do território.

Além disso, a escolha da freguesia como unidade de análise não segue as instruções de autores como Finio (2022), que sugerem o uso de uma escala espacial ainda mais fina (i.e. bairros) para identificar corretamente as zonas gentrificadas. Assim, o uso da freguesia implica uma inevitável generalização dos dados, que pode diluir variações importantes à escala intraurbana, introduzindo incerteza nos modelos. De facto, sendo uma unidade administrativa relativamente extensa e heterogénea, a freguesia tende a agrupar áreas com características socioeconómicas distintas, o que dificulta a deteção de processos de gentrificação que, por norma, se manifestam de forma localizada. Esta limitação reflete-se nos resultados obtidos, que não permitem isolar zonas específicas onde os efeitos do fenómeno sejam diretamente observáveis. Ainda assim, a evolução dos coeficientes, em variáveis-chave como a proporção de população com ensino superior, os padrões de

propriedade e os alojamentos de uso sazonal, sugere a presença de dinâmicas associadas à gentrificação, que aparentam ter-se expandido para fora dos perímetros tradicionais do centro histórico de Lisboa. Como estratégia futura, e à semelhança do que alguns autores propõem (Wilhelmsson et al., 2021), poderia ser considerada a identificação prévia de bairros já reconhecidos como gentrificados, nomeadamente Chiado, Bairro Alto e Príncipe Real, de modo a incorporar nos modelos uma variável de distância a estes bairros. Esta abordagem permitiria testar de forma mais explícita se a proximidade a zonas gentrificadas exerce um efeito espacialmente difusor sobre os preços do arrendamento. Outra perspetiva que poderia facilitar a interpretação e distinção das áreas gentrificadas poderia passar por uma análise de *clusters* aos resultados dos modelos MGWR, por forma a perceber como se poderia organizar o mercado de arrendamento da AML em submercados, criando grupos de localidades consoante a influência das várias variáveis, tal como sugerem Kopczewska & Ćwiakowski (2021).

Para além das questões inerentes à escala de análise, importa reconhecer que os modelos desenvolvidos poderão ter sido afetados pela omissão de variáveis relevantes. Embora se tenha procurado integrar um conjunto abrangente de indicadores socioeconómicos e urbanos, existem outros fatores que, potencialmente, poderiam contribuir para uma explicação mais robusta da variação dos preços no mercado de arrendamento, nomeadamente variáveis como a distância a infraestruturas estruturantes (i.e. hospitais, escolas, universidades ou interfaces de transportes). Outra variável que poderá ter sido erroneamente omitida consiste na distância ao CBD. Contudo, no contexto da AML, este conceito é alvo de muita controvérsia, uma vez que a tradicional associação do CBD à baixa lisboeta não reflete a atual realidade policêntrica da área metropolitana, onde se destacam vários polos científico-tecnológicos relevantes, tais como o Parque das Nações ou o Taguspark. Ainda relativamente a possíveis variáveis interessantes de analisar, uma linha de investigação futura particularmente promissora passaria pelo desenvolvimento de um índice de qualidade das freguesias (Chica-Olmo et al., 2013), que agregasse múltiplas dimensões (como segurança, escolaridade da população, oferta de serviços e qualidade ambiental), oferecendo, assim, uma visão mais integrada e robusta das características da vizinhança. Importa também referir que os modelos MGWR não permitem captar relações não lineares. Tendo em consideração os resultados obtidos na análise das relações bivariadas locais, sugere-se para estudos futuros a utilização de modelos de *machine learning* geoespacial, como por exemplo o modelo de Floresta Aleatória Geograficamente Ponderada (*Geographically Weighted Random Forest*), proposto por Georganos et al. (2021) e Georganos & Kalogirou (2022).

Por último, importa salientar que a dinâmica do mercado habitacional é extremamente volátil, pelo que os resultados obtidos com base em dados de 2021 poderão já não refletir com rigor a realidade atual. De facto, segundo dados do Idealista⁸, entre fevereiro de 2020 e fevereiro de 2025, o preço mediano do metro quadrado no arrendamento em Portugal registou um aumento de aproximadamente 44%, passando de 11,4 €/m² para 16,4 €/m². Esta variação acentuada num curto espaço de tempo sugere que novos fatores poderão ter vindo a ganhar relevância na explicação da valorização habitacional. Entre estes, destaca-se a crescente importância do teletrabalho, que tem incentivado a procura por habitação fora dos grandes centros urbanos, alterando os

⁸ Consultado no dia 08 de abril de 2024, no *website* do portal imobiliário online Idealista: <https://www.idealista.pt/media/relatorios-preco-habitacao/arrendamento/>

padrões de mobilidade e a geografia da procura. Simultaneamente, os impactos crescentes das alterações climáticas, nomeadamente a frequência de fenómenos extremos como inundações ou ondas de calor, poderão desvalorizar as áreas mais vulneráveis. Assim, futuras investigações poderiam beneficiar da incorporação de variáveis que traduzam estas novas dinâmicas sociais e ambientais, contribuindo para a construção de modelos mais atualizados e ajustados aos desafios do contexto contemporâneo.

Em suma, esta dissertação confirma a importância de uma abordagem espacialmente sensível para a análise do mercado de arrendamento e aponta para a necessidade de modelos cada vez mais complexos e contextualizados, capazes de integrar as múltiplas dimensões que influenciam os preços da habitação. Espera-se, portanto, que este trabalho tenha contribuído para a investigação científica, quer em termos teóricos, através do aprofundamento dos conhecimentos relativos à importância da aplicação de métodos de geoestatística para a compreensão dos fenómenos urbanos, quer em termos práticos, no apoio à tomada de decisão, através do auxílio à definição de políticas de habitação territorialmente ajustadas, e identificação de áreas potencialmente mais vulneráveis aos impactos destes processos de transformação social, económica e cultural das cidades.

REFERÊNCIAS BIBLIOGRÁFICAS

- Al-Masum, M., & Lee, C. (2019). Modelling housing prices and market fundamentals: Evidence from the Sydney housing market. *INTERNATIONAL JOURNAL OF HOUSING MARKETS AND ANALYSIS*, 12(4), 746–762. <https://doi.org/10.1108/IJHMA-10-2018-0082>
- Álvarez-Herranz, A., & Macedo-Ruíz, E. (2021). An evaluation of the three pillars of sustainability in cities with high airbnb presence: A case study of the city of Madrid. *Sustainability (Switzerland)*, 13(6). <https://doi.org/10.3390/su13063220>
- an de Meulen, P., & Mitze, T. (2014). Exploring the spatial variation in quality-adjusted rental prices and identifying hot spots in Berlin's residential property market. *Regional Studies, Regional Science*, 1(1), 310–328. <https://doi.org/10.1080/21681376.2014.981577>
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. <https://doi.org/10.2307/143780>
- Anselin, L. (2002). Under the hood Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3), 247–267. <https://doi.org/10.1111/j.1574-0862.2002.tb00120.x>
- Anselin, L., & Lozano-Gracia, N. (2009). Spatial hedonic models. Em *Palgrave Handb. Of Econometrics: Volume 2: Appl. Econometrics* (pp. 1213–1250). Palgrave Macmillan; Scopus. https://doi.org/10.1057/9780230244405_26
- Antunes, G., & Ferreira, J. (2021). SHORT-TERM RENTALS: HOW MUCH IS TOO MUCH – SPATIAL PATTERNS IN PORTUGAL AND LISBON. *Tourism and Hospitality Management*, 27(3), 581–603. <https://doi.org/10.20867/thm.27.3.6>
- Antunes, G., & Seixas, J. (2022). *Impactos da pandemia na evolução do acesso à habitação na Área Metropolitana de Lisboa*. <https://doi.org/10.15847/cct.26600>
- Atuesta, L. H., & Hewings, G. J. D. (2019). Housing appreciation patterns in low-income neighborhoods: Exploring gentrification in Chicago. *Journal of Housing Economics*, 44, 35–47. <https://doi.org/10.1016/j.jhe.2018.08.005>
- Banabak, S. (2024). Closing the Rent Index Gap – A Quantitative Approach to Rental-Sector Gentrification. *Tijdschrift voor Economische en Sociale Geografie*, 115(1), 64–80. Scopus. <https://doi.org/10.1111/tesg.12607>
- Bao, S., & Anselin, L. (1999). Spatial Data Analysis Wiht Spatial Statistics & GIS. *The Proceedings of Geoinformatics'99 Conference*, 17.
- Basu, S., & Thibodeau, T. G. (1998). Analysis of Spatial Autocorrelation in House Prices. *The Journal of Real Estate Finance and Economics*, 17(1), 61–85. <https://doi.org/10.1023/A:1007703229507>

- Batista, P., & Marques, J. L. (2021). *Automated Housing Price Valuation and Spatial Data: Vol. 12952 LNCS* (Gervasi O., Murgante B., Misra S., Garau C., Blečić I., Taniar D., Apduhan B.O., Rocha A.M., Tarantino E., & Torre C.M., Eds.; pp. 366–381). Scopus. https://doi.org/10.1007/978-3-030-86973-1_26
- Bhattacharjee, A., Castro, E., Maiti, T., & Marques, J. (2016). Endogenous Spatial Regression and Delineation of Submarkets: A New Framework with Application to Housing Markets. *Journal of Applied Econometrics*, 31(1), 32–57. Scopus. <https://doi.org/10.1002/jae.2478>
- Bhattacharjee, A., Castro, E., & Marques, J. (2012). Spatial Interactions in Hedonic Pricing Models: The Urban Housing Market of Aveiro, Portugal. *Spatial Economic Analysis*, 7(1), 133–167. <https://doi.org/10.1080/17421772.2011.647058>
- Bitter, C. (2008). *Geography, Housing Prices, and Interregional Migration*. <http://hdl.handle.net/10150/194749>
- Bitter, C., Mulligan, G. F., & Dall’erba, S. (2007). Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems*, 9(1), 7–27. <https://doi.org/10.1007/s10109-006-0028-7>
- Bottero, M., Caprioli, C., Foth, M., Mitchell, P., Rittenbruch, M., & Santangelo, M. (2022). Urban parks, value uplift and green gentrification: An application of the spatial hedonic model in the city of Brisbane. *Urban Forestry and Urban Greening*, 74. Scopus. <https://doi.org/10.1016/j.ufug.2022.127618>
- Carvalho, L., Chamusca, P., Fernandes, J., & Pinto, J. (2019). Gentrification in Porto: Floating city users and internationally-driven urban change. *Urban Geography*, 40(4), 565–572. <https://doi.org/10.1080/02723638.2019.1585139>
- CCDR-LVT & AML. (2020). *Estratégia Regional de Lisboa 2030* (p. 115). <https://www.ccdr-lvt.pt/study-publications/rlvt-2030-para-a-estrategia-2030-lisboa-aml-junho-2020/>
- Chen, M., Chun, Y., & Griffith, D. A. (2023). Delineating Housing Submarkets Using Space–Time House Sales Data: Spatially Constrained Data-Driven Approaches. *Journal of Risk and Financial Management*, 16(6). Scopus. <https://doi.org/10.3390/jrfm16060291>
- Chica-Olmo, J., Cano-Guervos, R., & Chica-Olmo, M. (2013). A Coregionalized Model to Predict Housing Prices. *URBAN GEOGRAPHY*, 34(3), 395–412. <https://doi.org/10.1080/02723638.2013.778662>
- Ciocchetta, F., Guglielminetti, E., & Mistretta, A. (2024). What Drives House Prices in Europe? *Oxford Bulletin of Economics and Statistics*, 86(5), 1089–1121. Scopus. <https://doi.org/10.1111/obes.12601>
- Cunha, A. M., & Lobão, J. (2022). The effects of tourism on housing prices: Applying a difference-in-differences methodology to the Portuguese market. *International Journal of Housing Markets and Analysis*, 15(4), 762–779. Scopus. <https://doi.org/10.1108/IJHMA-04-2021-0047>

Daly, P., Dias, Á., & Patuleia, M. (2021). The impacts of tourism on cultural identity on Lisbon Historic Neighbourhoods. *The Impacts of Tourism on Cultural Identity on Lisbon Historic Neighbourhoods*, 1, 1–25. <https://doi.org/10.29333/ejecs/516>

Direção-Geral do Território. (2019). *Especificações Técnicas da Carta de Uso e Ocupação do Solo (COS) de Portugal Continental para 2018* [Relatório Técnico]. Direção-Geral do Território. https://www.dgterritorio.gov.pt/sites/default/files/documentos-publicos/2019-12-26-11-47-32-0__ET-COS-2018_v1.pdf

Droj, G., Kwartnik-Pruc, A., & Droj, L. (2024). A Comprehensive Overview Regarding the Impact of GIS on Property Valuation. *ISPRS International Journal of Geo-Information*, 13(6). Scopus. <https://doi.org/10.3390/ijgi13060175>

Farber, S., & Yeates, M. (2006). A Comparison of Localized Regression Models in a Hedonic House Price Context. *The Canadian Journal of Regional Science*.

Finio, N. (2022). Measurement and Definition of Gentrification in Urban Studies and Planning. *Journal of Planning Literature*, 37(2), 249–264. <https://doi.org/10.1177/08854122211051603>

Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Exploring, modelling and predicting spatiotemporal variations in house prices. *The Annals of Regional Science*, 54(2), 417–436. <https://doi.org/10.1007/s00168-015-0660-6>

Fotheringham, A. S., Oshan, T. M., & Li, Z. (2024). *Multiscale Geographically Weighted Regression: Theory and Practice*. CRC Press.

Fotheringham, A. S., & Sachdeva, M. (2022). On the importance of thinking locally for statistics and society. *Spatial Statistics*, 50, 100601. <https://doi.org/10.1016/j.spasta.2022.100601>

Franco, S. F., & Santos, C. D. (2021). The impact of Airbnb on residential property values and rents: Evidence from Portugal. *Regional Science and Urban Economics*, 88, 103667. <https://doi.org/10.1016/j.regsciurbeco.2021.103667>

G. Carvalho, P., & Ribeiro, A. (2007). Acknowledging for spatial effects in the Portuguese housing markets. *MPRA Paper*, Artigo 6132. <https://ideas.repec.org/p/pra/mprapa/6132.html>

Geng, J., Cao, K., Yu, L., & Tang, Y. (2011). *Geographically Weighted Regression model (GWR) based spatial analysis of house price in Shenzhen*. Proceedings - 2011 19th International Conference on Geoinformatics, Geoinformatics 2011. Scopus. <https://doi.org/10.1109/GeoInformatics.2011.5981032>

Georganos, S., Grippa, Tais, Niang Gadiaga, Assane, Linard, Catherine, Lennert, Moritz, Vanhuysse, Sabine, Mboga, Nicholus, Wolff, Eléonore, & and Kalogirou, S. (2021). Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, 36(2), 121–136. <https://doi.org/10.1080/10106049.2019.1595177>

Georganos, S., & Kalogirou, S. (2022). A Forest of Forests: A Spatially Weighted and Computationally Efficient Formulation of Geographical Random Forests. *ISPRS International Journal of Geo-Information*, 11(9), Artigo 9. <https://doi.org/10.3390/ijgi11090471>

Glass, R. (1964). *London: Aspects of change*. MacGibbon & Kee.

Haining, R. (2003). *Spatial Data Analysis—Theory and Practice* (1.^a ed.). Cambridge University Press.

Hawkins, J., Ahmed, U., Roorda, M., & Habib, K. N. (2022). Measuring the process of urban gentrification: A composite measure of the gentrification process in Toronto. *Cities*, 126, 103708. <https://doi.org/10.1016/j.cities.2022.103708>

Instituto Nacional de Estatística. (2024). *Inquérito às Despesas das Famílias 2022/2023* (p. 65). https://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=673105441&att_display=n&att_download=y

Kopczewska, K., & Ćwiakowski, P. (2021). Spatio-temporal stability of housing submarkets. Tracking spatial location of clusters of geographically weighted regression estimates of price determinants. *Land Use Policy*, 103, 105292. <https://doi.org/10.1016/j.landusepol.2021.105292>

Lee, G., Cho, D., & Kim, K. (2016). The modifiable areal unit problem in hedonic house-price models. *Urban Geography*, 37(2), 223–245. <https://doi.org/10.1080/02723638.2015.1057397>

Lestegás, I. (2019). Lisbon After the Crisis: From Credit-fuelled Suburbanization to Tourist-driven Gentrification. *International Journal of Urban and Regional Research*, 43(4), 705–723. <https://doi.org/10.1111/1468-2427.12826>

Lestegás, I., Lois-González, R. C., & Seixas, J. (2018). The global rent gap of Lisbon's historic centre. *International Journal of Sustainable Development and Planning*, 13(4), 683–694. <https://doi.org/10.2495/SDP-V13-N4-683-694>

Ley, D. (1986). Alternative Explanations for Inner-City Gentrification: A Canadian Assessment. *Annals of the Association of American Geographers*, 76(4), 521–535. Scopus. <https://doi.org/10.1111/j.1467-8306.1986.tb00134.x>

Li, X., Wang, J., Luo, K., Liang, Y., & Wang, S. (2022). Exploring the Spillover Effects of Urban Renewal on Local House Prices Using Multi-Source Data and Machine Learning: The Case of Shenzhen, China. *Land*, 11(9), Artigo 9. <https://doi.org/10.3390/land11091439>

Longley, P., Goodchild, M., Maguire, D., & Rhind, D. (2015). *Geographic Information Science and Systems*. John Wiley & Sons.

Lu, B., Ge, Y., Shi, Y., Zheng, J., & Harris, P. (2023). Uncovering drivers of community-level house price dynamics through multiscale geographically weighted regression: A case study of Wuhan, China. *Spatial Statistics*, 53, 100723. <https://doi.org/10.1016/j.spasta.2022.100723>

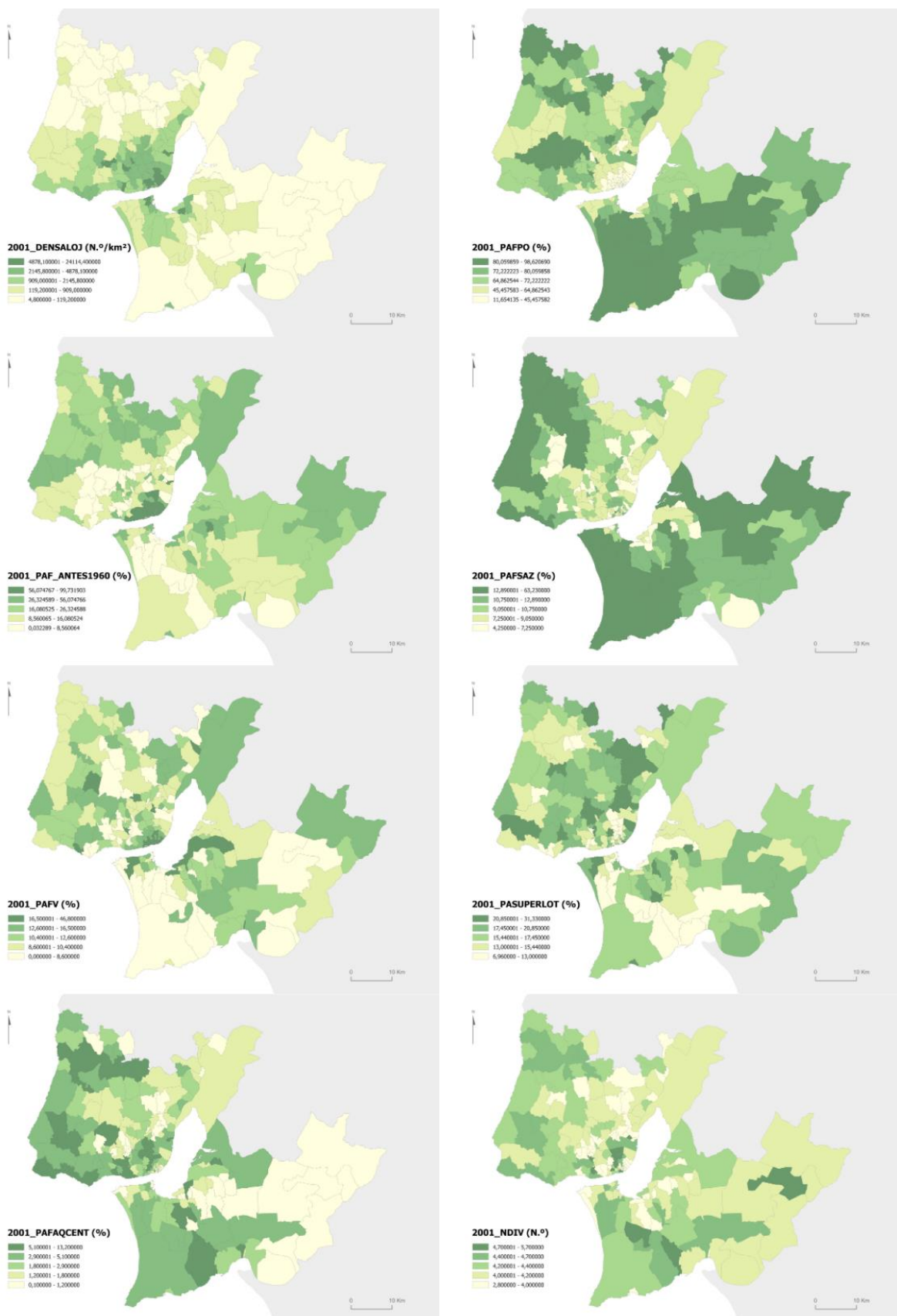
- Ma, Y., & Gopal, S. (2018). Geographically Weighted Regression Models in Estimating Median Home Prices in Towns of Massachusetts Based on an Urban Sustainability Framework. *Sustainability*, 10(4), Artigo 4. <https://doi.org/10.3390/su10041026>
- Macdonald, J., & Franco, S. (2016). *Tree canopies, urban green amenities and the residential real estate market: Remote sensing and spatial hedonic applications to Lisbon, Portugal*. European Regional Science Association. <https://EconPapers.repec.org/RePEc:wiw:wiwrsa:ersa16p640>
- Marques, J. L., Castro, E., & Bhattacharjee, A. (2013). Spatial patterns of the housing market: Multi-dimensional non-Euclidean notion of space. *Revista Portuguesa de Estudos Regionais*, 32(1), 5–24. Scopus.
- Mendes, L. (2017). Gentrificação turística em Lisboa: Neoliberalismo, financeirização e urbanismo austeritário em tempos de pós-crise capitalista 2008-2009. *Cadernos Metrópole*, 39(19), 479–512. <https://doi.org/10.1590/cm.v19i39.31531>
- Murphy, D. (2024). Housing cycles and gentrification. *Journal of Monetary Economics*, 144, 103550. <https://doi.org/10.1016/j.jmoneco.2024.01.003>
- O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, 41(5), 673–690. <https://doi.org/10.1007/s11135-006-9018-6>
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55. <https://doi.org/10.1086/260169>
- Samadani, S., & Costa, C. J. (2021). *Forecasting real estate prices in Portugal: A data science approach* (Rocha A., Goncalves R., Penalvo F.G., & Martins J., Eds.). Scopus. <https://doi.org/10.23919/CISTI52073.2021.9476447>
- Sampaio, E., Ramos, M., & Barroso, M. (2010). *Exercícios de Estatística Descritiva para as Ciências Sociais* (2.^a ed.). Edições Sílabo.
- Santos, A. D., & Castro, H. (2022). Housing and Setting Constraints: The Portuguese Evidence. *Sustainability*, 14(18), Artigo 18. <https://doi.org/10.3390/su141811720>
- Seixas, J., & Antunes, G. (2019). Tendências recentes de segregação habitacional na Área Metropolitana de Lisboa. *Cidades. Comunidades e Territórios*, 39, Artigo 39. <https://journals.openedition.org/cidades/1555>
- Shi, D., Guan, J., Zurada, J., & Levitan, A. S. (2022). Predicting home sale prices: A review of existing methods and illustration of data stream methods for improved performance. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(2). Scopus. <https://doi.org/10.1002/widm.1435>
- Soltani, A., Zali, N., Aghajani, H., Hashemzadeh, F., Rahimi, A., & Heydari, M. (2024). The nexus between transportation infrastructure and housing prices in metropolitan regions. *Journal of Housing and the Built Environment*, 39(2), 787–812. Scopus. <https://doi.org/10.1007/s10901-023-10085-3>

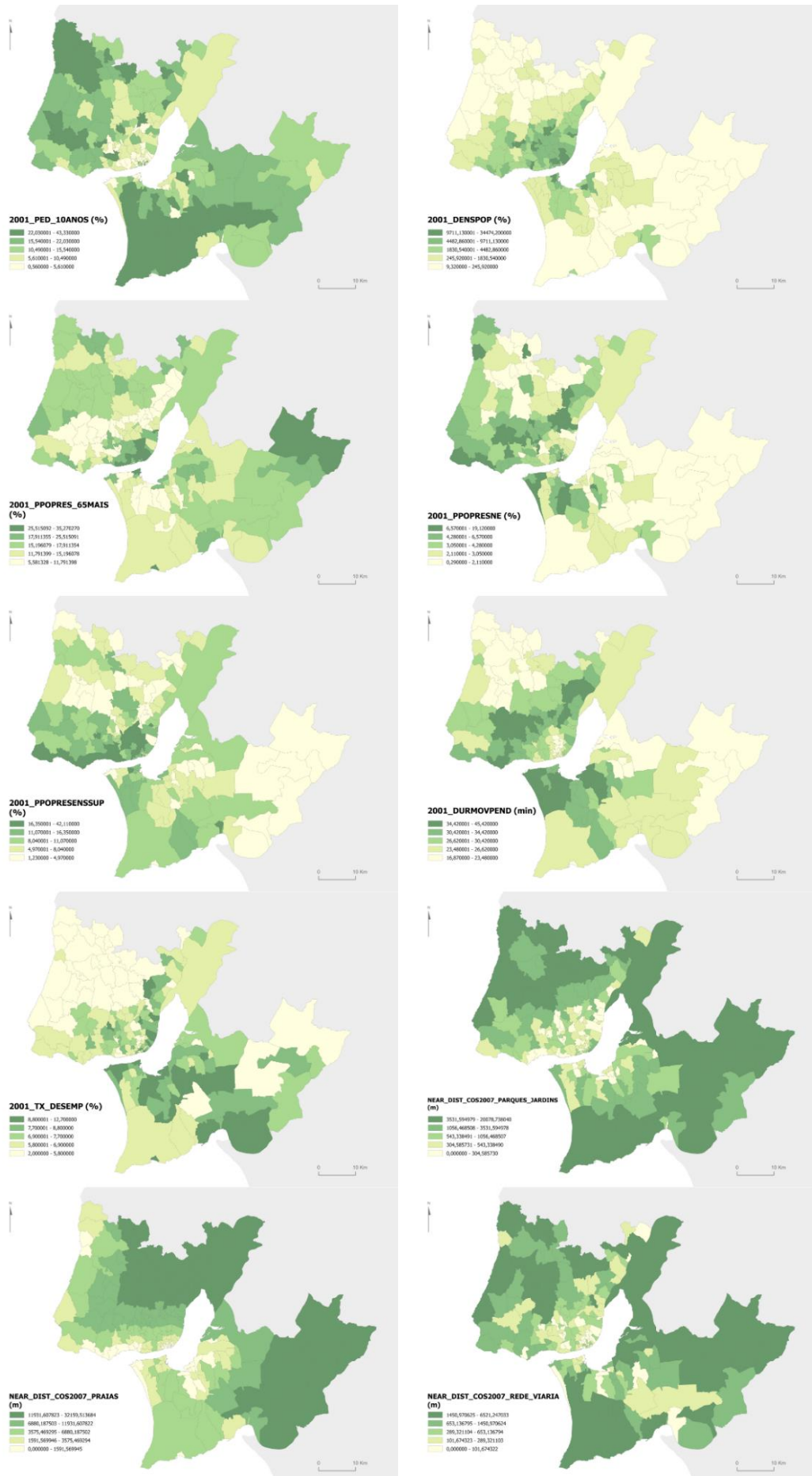
- Teixeira, M. C. C., Caridad y Ocerin, J. M., & Villamandos, N. C. (2011). Variáveis explicativas e a sua importância na formação do preço de um apartamento em Portugal: Uma abordagem com redes neuronais artificiais. *XXI Jornadas Hispano-Lusas de Gestión Científica, Cordoba, 2-4 Fev. 2011*. <https://repositorio.ipcb.pt/handle/10400.11/1033>
- Travasso, N., Oro, A. V., Almeida, M. R. de, & Ribeiro, L. S. (2020). ACESSO AO MERCADO DE ARRENDAMENTO EM PORTUGAL.: UM RETRATO A PARTIR DO PROGRAMA DE ARRENDAMENTO ACESSÍVEL. *Finisterra*, 55(114), Artigo 114. <https://doi.org/10.18055/Finis19380>
- Valier, A. (2020). Who performs better? AVMs vs hedonic models. *Journal of Property Investment and Finance*, 38(3), 213–225. Scopus. <https://doi.org/10.1108/JPIF-12-2019-0157>
- Wilhelmsson, M., Ismail, M., & Warsame, A. (2021). Gentrification effects on housing prices in neighbouring areas. *International Journal of Housing Markets and Analysis*, 15(4), 910–929. <https://doi.org/10.1108/IJHMA-04-2021-0049>
- Wu, C., Ren, F., Hu, W., & Du, Q. (2019). Multiscale geographically and temporally weighted regression: Exploring the spatiotemporal determinants of housing prices. *International Journal of Geographical Information Science*, 33(3), 489–511. Scopus. <https://doi.org/10.1080/13658816.2018.1545158>
- Wu, Y., Wei, Y. D., & Li, H. (2020). Analyzing Spatial Heterogeneity of Housing Prices Using Large Datasets. *Applied Spatial Analysis and Policy*, 13(1), 223–256. <https://doi.org/10.1007/s12061-019-09301-x>
- Yao, J., & Fotheringham, A. (2016). Local Spatiotemporal Modeling of House Prices: A Mixed Model Approach. *PROFESSIONAL GEOGRAPHER*, 68(2), 189–201. <https://doi.org/10.1080/00330124.2015.1033671>
- Zaki, J., Nayyar, A., Dalal, S., & Ali, Z. (2022). House price prediction using hedonic pricing model and machine learning techniques. *CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE*, 34(27). <https://doi.org/10.1002/cpe.7342>

APÊNDICE A: HISTOGRAMAS DA DISTRIBUIÇÃO DOS POTENCIAIS FATORES EXPLICATIVOS DE 2001

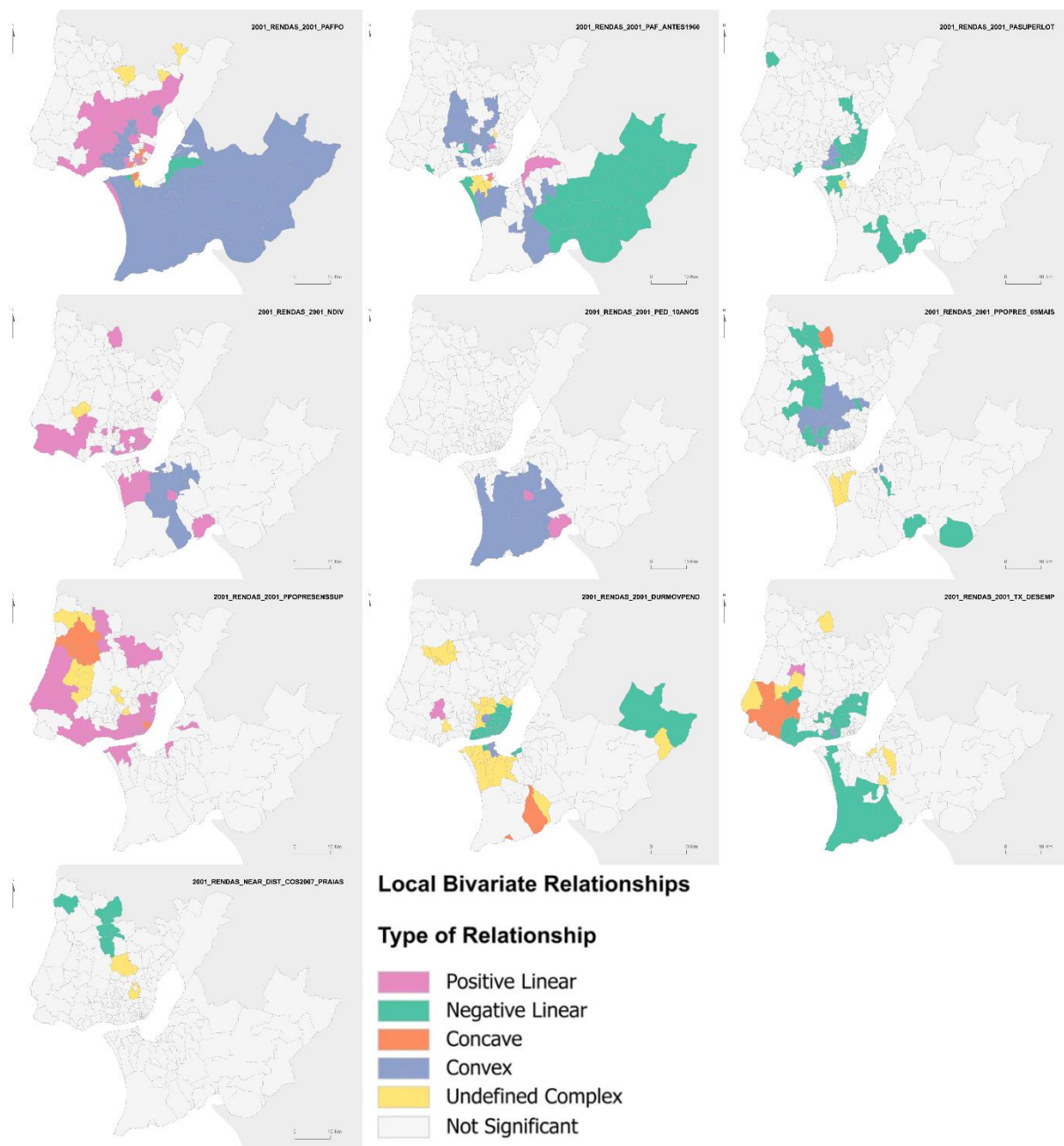


APÊNDICE B: MAPAS TEMÁTICOS DOS POTENCIAIS FATORES EXPLICATIVOS, EM 2001

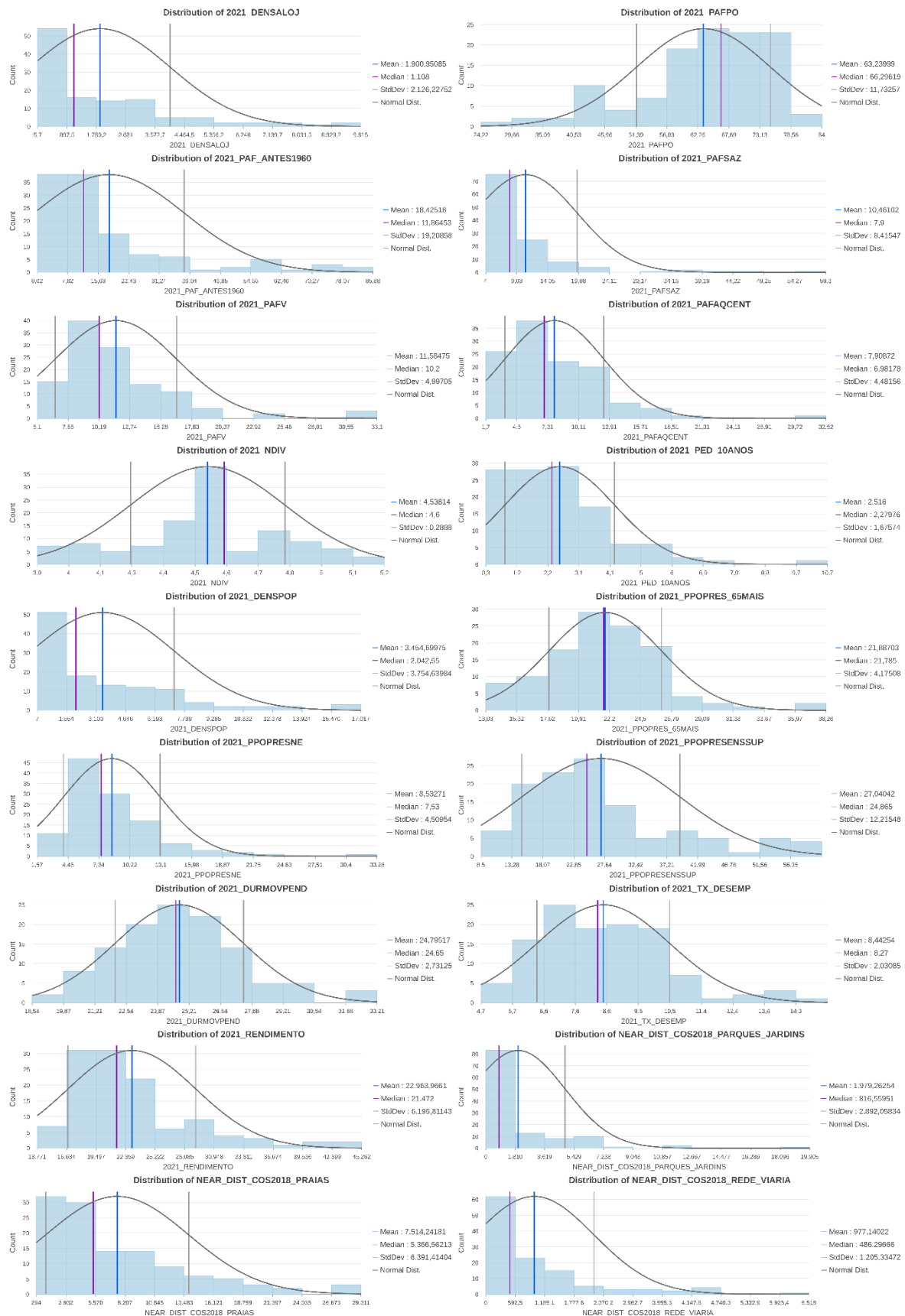




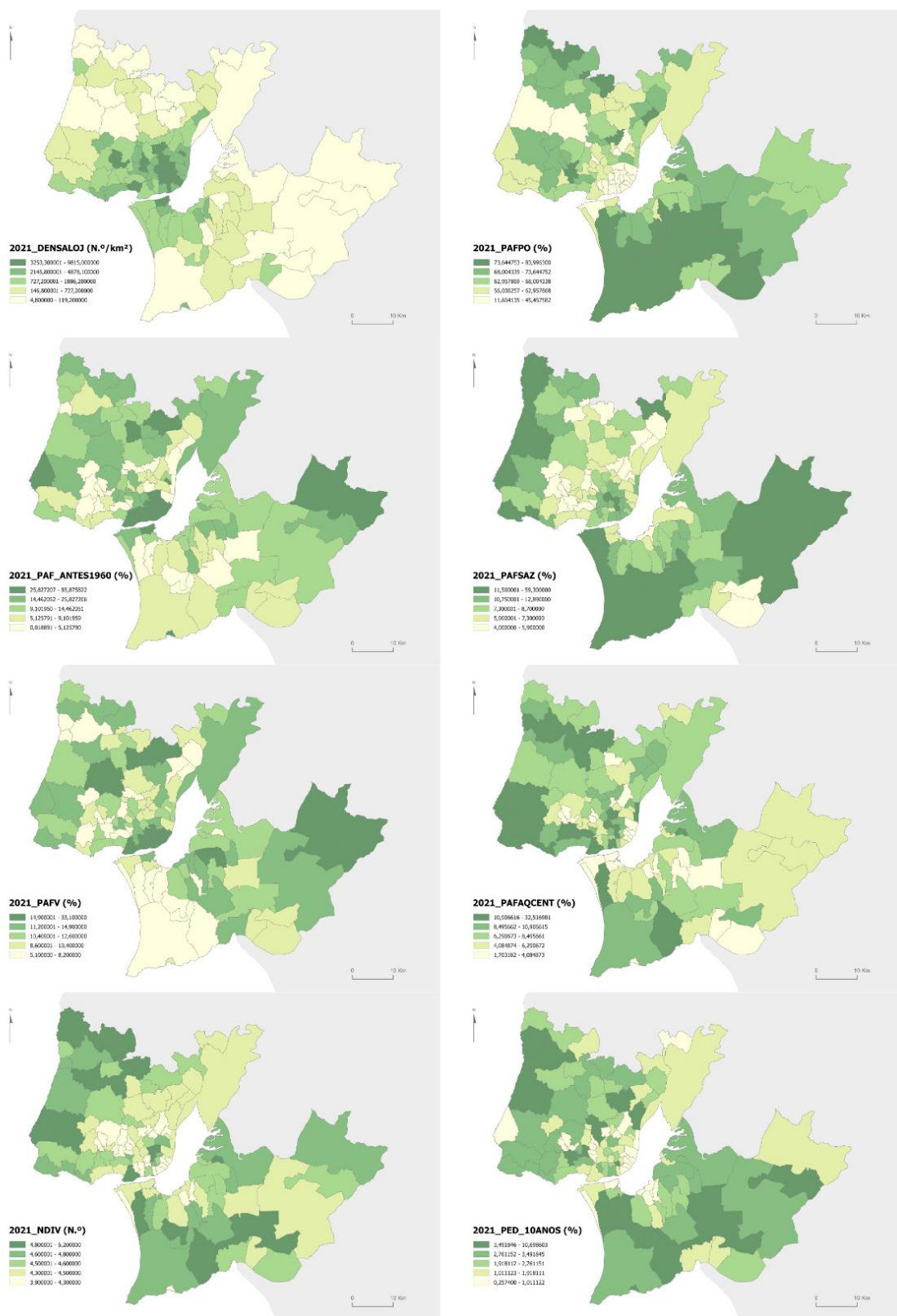
APÊNDICE C: RELAÇÕES BIVARIADAS LOCAIS SIGNIFICANTES ENTRE FATORES EXPLICATIVOS E O VALOR MÉDIO MENSAL DAS RENDAS, EM 2001

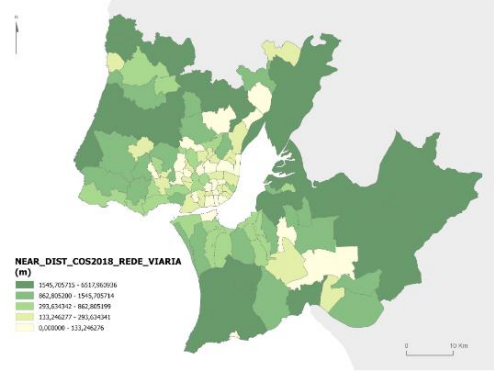
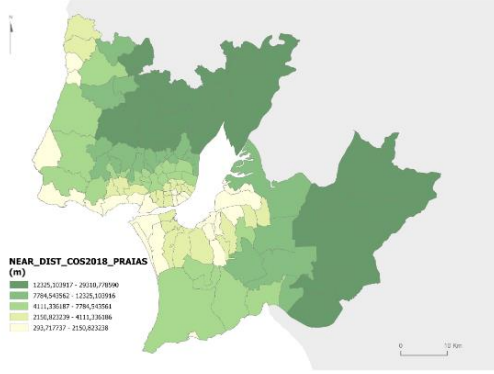
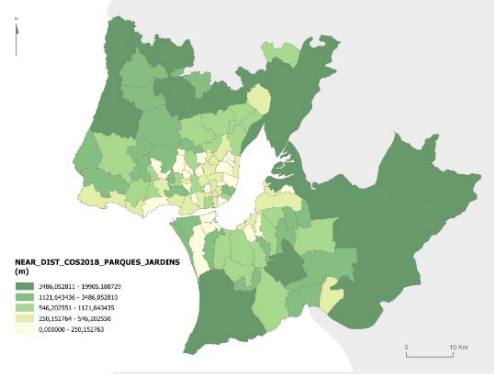
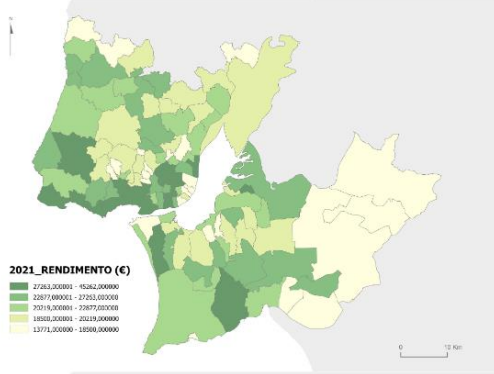
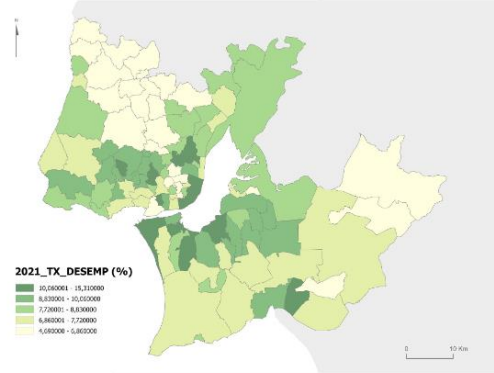
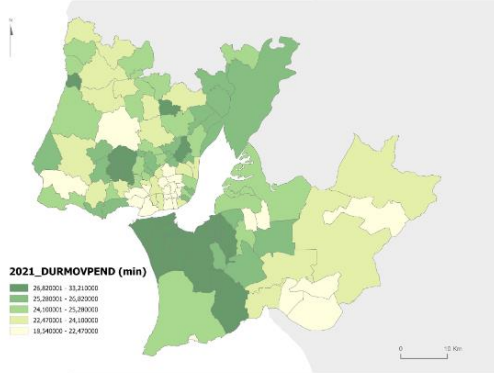
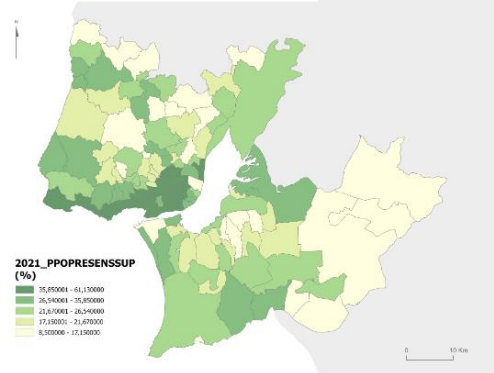
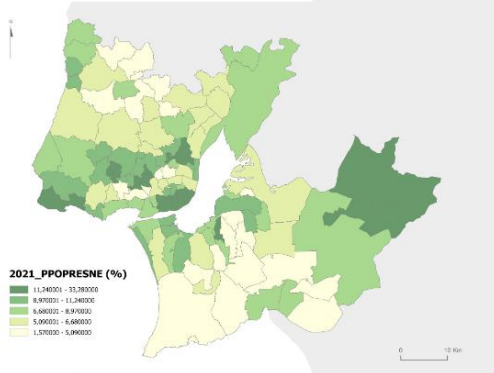
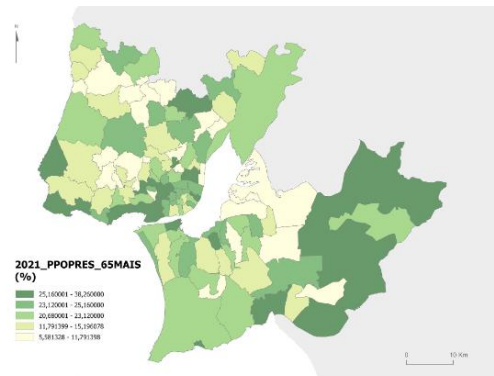
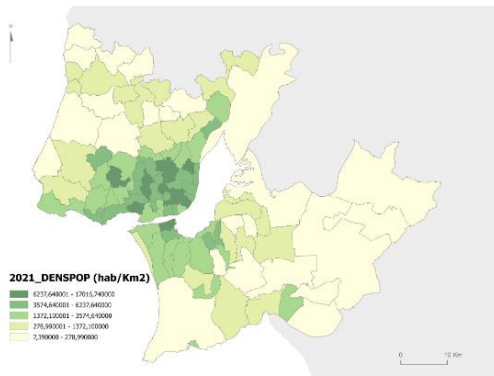


APÊNDICE D: HISTOGRAMAS DA DISTRIBUIÇÃO DOS POTENCIAIS FATORES EXPLICATIVOS DE 2021

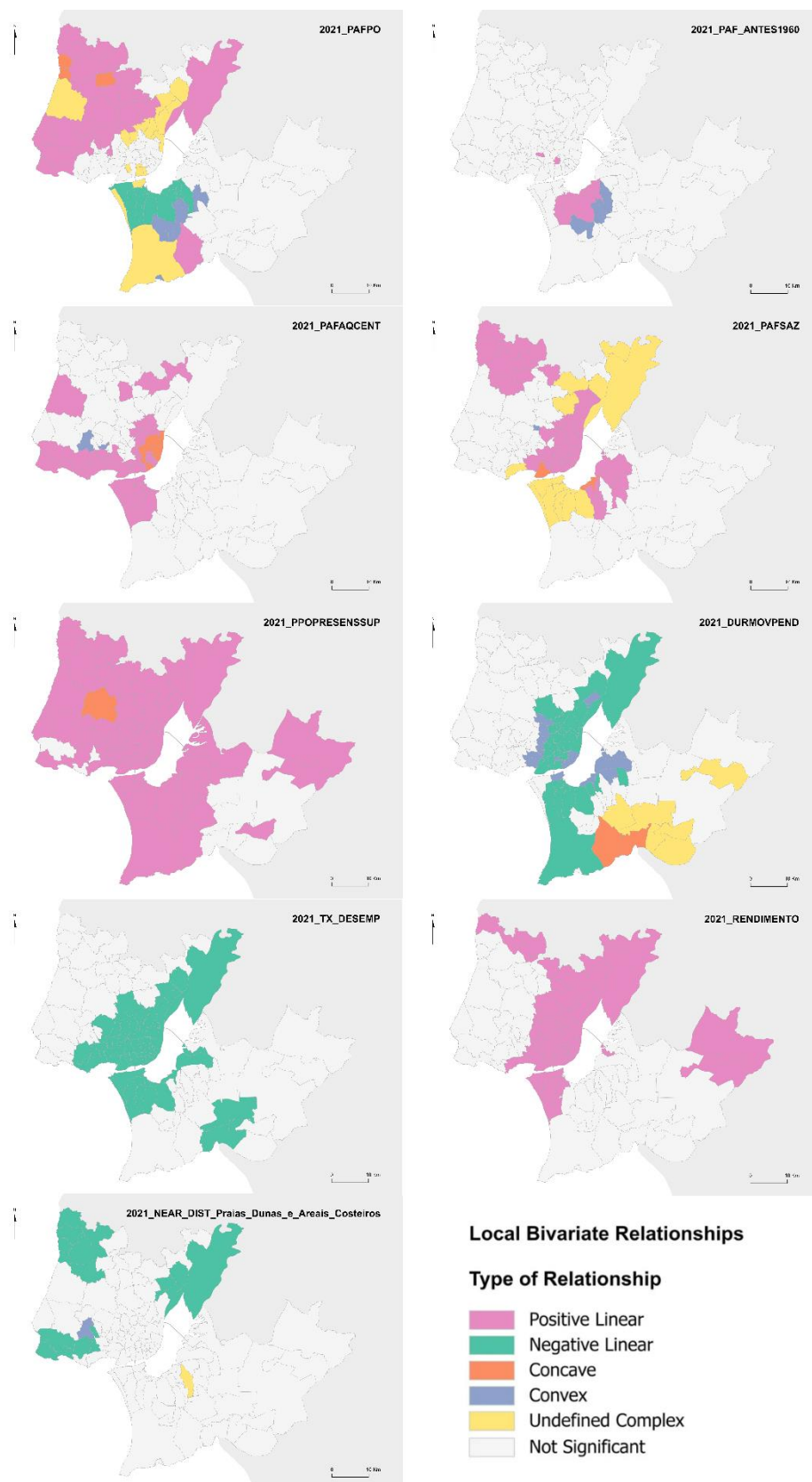


APÊNDICE E: MAPAS TEMÁTICOS DOS POTENCIAIS FATORES EXPLICATIVOS, EM 2021





APÊNDICE F: RELAÇÕES BIVARIADAS LOCAIS SIGNIFICANTES ENTRE FATORES EXPLICATIVOS E O VALOR MÉDIO MENSAL DAS RENDAS, EM 2021





C& SIG

UNIGIS PT

NOVA
IMS
Information
Management
School