

C&S SIG

Modelação dos dados abertos de bike-sharing de Bruxelas (Villo)
através de modelos de regressão espacial

Tiago Daniel Costa Pina

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

Modelação dos dados abertos de bike-sharing de Bruxelas
(Villo) através de modelos de regressão espacial

Dissertação orientada por
Professora Doutora Ana Cristina Costa

11 de Novembro de 2018

Declaração de Originalidade

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 assinalada 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.

Bruxelas, 11 de Novembro de 2018

Tiago Daniel Costa Pina

Agradecimentos

À professora Ana Cristina Costa que aceitou orientar-me neste trabalho, pelos conselhos, correções, material de estudo, assim como pela paciência e amabilidade que demonstrou sempre ao longo deste tempo.

Aos meus pais, que nunca descuidaram a minha formação e sempre me deram todos os meios para poder aprender mais.

À Ania, por tantas vezes ter vestido o fato de super heroína na luta contra o super-divagador implacável.

Ao Roi, que sempre me lembra que D. Afonso Henriques era Galego, por me ter pegado o gosto pelos Sistemas de Informação Geográfica.

A todos vós, muito obrigado.

Resumo

O ciclismo utilitário tem despertado renovado interesse nos últimos anos devido ao reconhecido potencial deste modo de transporte na redução da pegada energética e da poluição nas cidades. A etiqueta de meio de transporte sustentável é justificada, para além dos fatores citados, pelo reduzido espaço que a bicicleta ocupa em comparação com o automóvel, assim como pela pegada ecológica de toda a cadeia de produção. Neste contexto, várias cidades a nível mundial têm levado a cabo iniciativas para incitar a utilização quotidiana da bicicleta por parte dos habitantes. De entre estas iniciativas, os sistemas públicos de bicicletas partilhadas (*bike-sharing*) são talvez o melhor exemplo dada a sua universalidade. Os sistemas de bicicletas partilhadas além de possibilitarem a utilização da bicicleta sem necessidade de aquisição da mesma, são também uma fonte de dados valiosa, que pode ser usada para melhorar os sistemas existentes ou para levar a cabo estudos de investigação sobre diferentes temas ligados à mobilidade e à forma como os cidadãos se deslocam e interagem com o meio envolvente.

Neste trabalho, exploram-se os dados abertos do sistema público de partilha de bicicletas de Bruxelas (Villo), com vista a elaborar um modelo de regressão espacial que permita estimar o número de viagens em bicicleta. Procura-se entender os fatores que influenciam a mobilidade ao longo do dia e são analisados os limites e potencialidades do modelo, nomeadamente no que diz respeito à aplicação a outros contextos e à generalização para todas as viagens em bicicleta.

Neste estudo começa-se por realizar um modelo de regressão linear exploratório, que permite identificar um conjunto de variáveis potencialmente explicativas do número de viagens em bicicleta nas estações. No final parametriza-se um modelo de regressão geograficamente ponderado com distribuição de Poisson, com as variáveis identificadas previamente.

Os resultados do modelo final sugerem que as relações entre as variáveis analisadas apresentam uma grande variação espacial e temporal. O modelo permite além disso a revelação de relações locais escondidas, que se contextualizadas com outros estudos, permitem um melhor conhecimento dos fatores que afetam as viagens em bicicleta.

Abstract

In recent years there has been a renewed interest in utilitarian cycling due to its recognized potential in the reduction of energy consumption and pollution in the cities. In addition to the above-mentioned factors, the attached label of sustainable means of transport is justified by the little space occupied by bicycles compared to cars, as well as by the smaller ecological impact of the entire production chain. Within this context, several cities around the world have been launching different initiatives of bike promotion. Among these initiatives, bike-sharing systems are perhaps the best example given their universality. Bike-sharing systems allow users to ride a bicycle between two places (not necessarily different places) without the need of having a bike of one's own. These systems generate a big amount of data which can then be used to improve the systems themselves or to carry out research studies on different subjects related to mobility and the way how citizens move around and interact with the surrounding environment.

In this paper, the open data automatically generated by the Brussel's bike-sharing system (Villo) is explored, aiming to elaborate a spatial regression model to estimate the number of bicycle trips at stations. The main goal of the modelling process is to understand the underlying factors influencing mobility patterns throughout the day. The weaknesses and strengths of the model approach are analyzed, especially in what regards its implementation in other geographic contexts and the possibilities of generalization for all bicycle trips.

The first steps of the modelling process consist in setting up an exploratory Ordinary Least Squares model (OLS) in order to identify the potential explanatory variables among those available in the database. In the end, a Geographically Weighted Poisson Regression model (GWRP) is parametrized using the previous identified variables.

The results suggest the relationships between the dependent and independent variables are complex and spatially varying. Furthermore, the results show hidden patterns that enable further local investigation on these relationships.

Palavras Chave

Bike-sharing; Ciclismo utilitário; Mobilidade; Estatística espacial; Modelação; Regressão espacial

Keywords

Bike-sharing; Utility cycling; Mobility; Spatial statistics; Modelling; Spatial regression

Abreviaturas / Acrónimos

OLS: Regressão linear ordinária ¹

GWR: Regressão geograficamente ponderada clássica²

GWRP: Regressão geograficamente ponderada com distribuição de Poisson ³

SGWRP: Regressão geograficamente ponderada semi-paramétrica com distribuição de Poisson ⁴

AIC: Akaike Information Criterion

¹*Ordinary Least Squares*

²*Geographically Weighted Regression*

³*Geographically Weighted Regression - Poisson*

⁴*Semiparametric Geographically Weighted Regression - Poisson*

Índice do texto

1	Introdução	1
2	Revisão da Literatura	4
3	O sistema de bicicletas partilhadas de Bruxelas	8
4	Dados	10
4.1	Recolha e tratamento dos dados	10
4.2	Dados Villo	11
4.3	Dados sócio-económicos	14
4.4	Outros dados (mobilidade, infraestrutura, etc)	15
4.5	Base de dados final	15
5	Metodologia	16
5.1	Análise exploratória dos dados	17
5.1.1	Análise de <i>clusters</i> e <i>outliers</i> espaciais	17
5.1.2	Regressão linear exploratória	19
5.2	Análise de Regressão Geograficamente Ponderada (GWR)	20
5.3	Modelos com distribuição de Poisson	21
6	Resultados	23
6.1	Resultados da análise exploratória	23
6.2	Análise de regressão linear exploratória (OLS exploratório)	24
6.2.1	OLS exploratório - Dias úteis / Fins de semana e feriados	25
6.2.2	OLS exploratório - Períodos de ponta da manhã e da tarde	26
6.3	Seleção das variáveis explicativas	27
6.4	Resultados da modelação com distribuição de Poisson	29
6.5	Interpretação das variações geográficas dos parâmetros	34
7	Conclusão	40

Índice de Figuras

1	Orografia da região	9
2	Processo de recolha e tratamento dos dados	10
3	Número de viagens registadas por hora no conjunto das estações Villo em Maio de 2017	12
4	Estações Villo, zonas de influência (polígonos de Voronoi) e limites administrativos	13
5	Metodologia	16
6	Histograma matriz de pesos	18
7	Análise da auto-correlação local da variável dependente	23
8	Análise da auto-correlação local - Geração e atração de viagens no período de ponta da manhã	24
9	Análise da auto-correlação local - Geração e atração de viagens no período de ponta da tarde	24
10	Resultados GWRP - Variável <i>Densidade de paragens de transporte público</i>	35
11	Resultados GWRP - Variável <i>Densidade de escritórios</i>	36
12	Resultados GWRP - Variável <i>Densidade de estacionamento para bicicletas</i>	37
13	Resultados GWRP - Variável <i>Distância ao centro</i>	38
14	Resultados GWRP - Variável <i>Nº de docas Villo</i>	39
A.1	Resultados SGWRP - [G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)]: Variável <i>Densidade de paragens de transporte público</i>	51
A.2	Resultados SGWRP - [G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)]: Variável <i>Densidade de escritórios</i>	52
A.3	Resultados SGWRP - [G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)] (Dias úteis / Fins de semana e feriados): Variável <i>Densidade de estacionamento para bicicletas</i>	53
A.4	Resultados SGWRP - [G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)] (Dias úteis / Fins de semana e feriados): Variável <i>Número de docas Villo</i>	54
A.5	Mapa de resíduos: Modelo SGWRP - [G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)]	55
A.6	Mapa de resíduos: Modelo GWRP - (bst; dens_ptsto; dens_bpark; dtcc; dens_offic)	55

Índice de Tabelas

1	Resultados da regressão linear exploratória - Dias úteis / Fins de semana e feriados	25
2	Resultados da regressão linear exploratória para os períodos de ponta da manhã e da tarde	26
3	Comparação entre o modelo OLS alternativo e o modelo de regressão linear exploratório OLS*	28
4	Estatísticas descritivas das variáveis em análise	29
5	Resultados do modelo de regressão de Poisson global	31
6	Comparação estatística entre os modelos	33
A.1	Tarifas em vigor em 2018	46
A.2	Dados	46

1 Introdução

A premissa que o sistema de transportes é serviço e não uma mercadoria, e que por isso tem de ser utilizado quando é "produzido" [Ortúzar and Willumsen, 2011] condiciona a forma como se planeia e dimensiona um determinado sistema de transportes. Uma consequência desta característica dos sistemas de transportes é que no processo de planeamento e conceção de um dado sistema de transportes é fundamental quantificar as viagens geradas e atraídas nas diferentes zonas do território, assim como identificar quando é que essas viagens ocorrem.

Se no passado o conhecimento da procura de transportes contribuía essencialmente para fundamentar as tomadas de decisão ao nível das infraestruturas viárias e de transportes públicos, nas últimas décadas um novo paradigma de políticas urbanas tem dado destaque aos modos de transporte ditos ativos (ou suaves), como são a marcha a pé e o ciclismo utilitário. Esta mudança de paradigma é resultado da necessidade de assegurar a segurança das pessoas, assim como de procurar formas de transporte mais limpas e sustentáveis. Neste contexto, novas estratégias, que têm por objetivo reduzir o tráfego automóvel e a dependência do automóvel, têm sido adotadas. [Randriamanamihaga et al., 2014]

Dentro dos modos ativos, o ciclismo utilitário tem tido especial destaque, não só pelas melhorias das infraestruturas cicláveis operadas nos últimos anos em várias cidades europeias, mas também pelas medidas que visam favorecer a multi-modalidade ou o acesso a este modo de transporte através de sistemas coletivos de partilha de bicicletas ou *bike-sharing*. Os sistemas coletivos de partilha de bicicletas contemporâneos correspondem à oferta de bicicletas para alugueres de curta duração entre diferentes estações. Estas bicicletas são normalmente dotadas de tecnologias que permitem ao operador monitorizar os movimentos entre estações e, nos casos em que as bicicletas são dotadas de GPS, registar os movimentos ao longo da rede [Fishman et al., 2013].

A implementação destes sistemas tem contribuído para aumentar o número de ciclistas, a utilização de transportes públicos e como consequência diminuir os gases com efeito de estufa e melhorar a saúde da população. [DeMaio, 2009], [Otero et al., 2018]. Os sistemas de bicicletas partilhadas inscrevem-se assim numa perspetiva de desenvolvimento sustentável, contribuindo em alguns casos como uma alternativa em si mesmo aos modos de transporte tradicionais, constituindo-se noutros casos como complementares a estes [Migdley, 2011].

Neste contexto de crescente interesse pelo ciclismo utilitário, a falta de informação e investigação é um problema com que se têm confrontado os técnicos ao tentar integrar este modo de transporte nos diferentes modelos e previsões. Nalguns casos, o ciclismo utilitário é calculado como uma constante a partir de matrizes origem-destino, que apenas contemplam o automóvel e o transporte público. Os dados de origem e destino são muitas vezes obtidos através de inquéritos que apenas contemplam o automóvel e o transporte público como categorias separadas, sendo que os modos não motorizados (marcha a pé, ciclismo utilitário e outros) são agregados na mesma categoria [Levy et al., 2017].

A falta de dados sobre o volume de viagens em bicicleta para estimar a procura em estudos de planeamento pode ser parcialmente remediada pelas novas fontes de dados. Estas são potencialmente relevantes no planeamento de sistemas de transportes dada a capacidade de fornecer informações valiosas sobre a localização, horário e frequência de atividades geradoras de viagens ou fornecem evidências diretas do volume, concentração e direção dos movimentos de pessoas e veículos [Milne and Watling, 2018].

Por este motivo, nos últimos anos, tem havido um número crescente de estudos de investigação baseados nos dados de viagens e dos utilizadores de sistemas coletivos de partilha de bicicletas. Desta forma, os dados produzidos a partir de sistemas coletivos de partilha de bicicleta podem ser vistos como um exemplo de *big-data* aplicado aos transportes [Levy et al., 2017].

Os sistemas atuais de partilha de bicicletas beneficiam dos recentes avanços tecnológicos e permitem uma utilização automática 24 horas por dia. Uma das características destes sistemas é a possibilidade de oferecer um serviço e ao mesmo tempo armazenar dados importantes sobre a utilização do mesmo. Uma parte dos sistemas de partilha de bicicletas permite aos utilizadores a consulta em linha em quase tempo real da disponibilidade de bicicletas nas estações. Em muitos casos, estes dados são tornados públicos através de um interface de programação de aplicações, vulgo API, de maneira a que programadores ou investigadores os possam partilhar ou estudar. Outros casos há em que a informação sobre as viagens individuais entre estações ou de um ponto a outro é disponibilizada [O'Brien et al., 2014].

Em ambos os casos a disponibilização de dados dos sistemas de bicicletas partilhadas abre novas possibilidades no âmbito do planeamento de transportes, uma vez que permite entender o comportamento dos utilizadores complementando ou substituindo as entrevistas, contagens ou outros métodos clássicos de recolha de dados, mais custosos e consumidores de tempo e recursos humanos.

De entre os sistemas estudados por [O'Brien et al., 2014], o sistema público de bicicletas partilhadas de Bruxelas (Villo) é um dos maiores implementados em território europeu em termos de número de bicicletas e principalmente em área de cobertura. Contudo, à parte dos meta-estudos realizados por [O'Brien et al., 2014] e [Otero et al., 2018], em que se comparam diversos indicadores de diferentes sistemas, não se conhecem, à data, estudos específicos sobre o sistema público de bicicletas partilhadas de Bruxelas, ao contrário do que acontece em outras cidade europeias, como são os casos de Barcelona ([Kaltenbrunner et al., 2010] e [Faghieh-Imani et al., 2017]), Londres ([Lathia et al., 2012]), Lyon ([Tran et al., 2015]), Madrid ([García-Palomares et al., 2012]), Paris ([Randriamanamihaga et al., 2014] e [Feng et al., 2017]), Sevilha ([Faghieh-Imani et al., 2017]) ou Viena ([Vogel et al., 2011]).

Este trabalho pretende assim suprir uma lacuna existente no que diz respeito à investigação do ciclismo quotidiano em Bruxelas e integra-se num contexto de interesse crescente pelos modos suaves, fazendo uso das novas tecnologias e fontes de dados, para entender os fatores que estão na origem das viagens em bicicleta.

Através da utilização de métodos de regressão, pretende-se analisar os fatores sócio-económicos, de infraestrutura e de uso do solo que influenciam as viagens nas diferentes estações de bicicletas partilhadas de

Bruxelas, assim como perceber de que forma esses fatores são afetados consoante o período de análise.

A opção pela utilização de modelos de regressão é justificada pela existência de vasta literatura científica sobre a aplicação de modelos de regressão em diversos domínios do conhecimento, de entre os quais o domínio dos transportes. No âmbito deste último, é de destacar a existência de vários estudos no domínio da segurança rodoviária e modelação de acidentes de viação através de modelos de regressão espacial ([Dereli and Erdogan, 2017], [Hadayeghi et al., 2010], [Li et al., 2013], [Rhee et al., 2016] e [Huang et al., 2018]). Já no que diz respeito aos modos suaves e aos sistemas coletivos de bicicletas partilhadas são de destacar os trabalhos de [Zhou et al., 2018], [Levy et al., 2017], [Noland et al., 2016], [Griffin and Jiao, 2015] e [Yang et al., 2017]. Pela sua importância e paralelismo com este trabalho, alguns destes estudos serão abordados com mais detalhe no capítulo seguinte.

O objetivo último deste trabalho é entender os limites da metodologia utilizada na generalização do modelo a outros contextos geográficos e às viagens em bicicleta como um todo.

No contexto deste trabalho, termos como "número de viagens", "procura de viagens" ou "atividade das estações", são usados como sinónimos, referindo-se indiferentemente ao número de viagens com origem ou destino nas diferentes estações, ou seja, à soma das entradas e das saídas de bicicletas. Já os termos "atração" e "geração" dizem respeito às viagens "de entrada" ou de "saída" respetivamente.

2 Revisão da Literatura

Embora a literatura no âmbito dos sistemas de bicicletas partilhadas seja ainda relativamente escassa, vários estudos científicos visando entender os fatores que influenciam a utilização da bicicleta enquanto meio de transporte ou com o intuito de contribuir para um melhor planeamento dos sistemas de bicicletas partilhadas têm sido publicados. De entre esses estudos, este capítulo faz referência a várias publicações científicas que utilizam métodos e modelos de análise espacial, como é o caso dos modelos de regressão espacial, de data-mining ou outros e que têm como objetivo entender, explicar ou prever os fenómenos em análise.

No âmbito mais geral da utilização de ferramentas de sistemas de informação geográfica na otimização da localização de estações de bicicletas partilhadas, [García-Palomares et al., 2012] propõem um modelo que define a localização das estações de bicicletas partilhadas que melhor respondem às condições de procura potencial em função de dois critérios: a maximização da cobertura e a minimização dos custos (distância entre a oferta e a procura potencial). No cálculo da procura potencial, os autores têm em conta a população e o número de empregos associado a cada edifício assim como os dados de um inquérito de mobilidade correspondentes ao número de viagens atraídas e geradas em cada zona. A partir do mapa de distribuição da procura potencial e da rede de estradas, com informação sobre a velocidade teórica das bicicletas em cada troço, os autores determinam a localização das estações tendo em conta os critérios de otimização definidos previamente e cinco diferentes cenários no que respeita ao número de estações. Importa referir, que neste estudo a localização de estações de bicicletas partilhadas nas imediações das principais estações de metro e de comboios suburbanos de Madrid foi considerada fundamental, pelo que 52 das estações foram previamente localizadas nesses locais, não entrando como *input* no modelo.

Noutro estudo, [O'Brien et al., 2014] analisam os dados de 38 sistemas de bicicletas partilhadas localizados na América, Ásia, Europa, Médio-Oriente e Oceânia obtidos durante um período de dois anos e armazenados numa base de dados. A partir desses dados, os autores analisam as diferentes estações de acordo com as suas características operacionais, espaciais e temporais. Na análise dessas características são usados vários indicadores, como é o caso do rácio entre o número de docas com bicicletas e o total de docas ⁵, apenas para citar aquele que é o indicador descrito como sendo o mais importante. Este indicador é analisado sob diferentes declinações. Os vários sistemas são assim comparados para cada um dos indicadores individualmente, sendo agrupados no final usando um método de agrupamento hierárquico. O sistema Villo de Bruxelas, é um dos 38 sistemas analisados no âmbito deste estudo, sendo caracterizado, desde uma perspetiva temporal, como um sistema orientado essencialmente para o lazer, com maior utilização durante os fim de semana e usualmente como tendo um elevado rácio entre o número de docas com bicicletas e total de docas. De um ponto de vista das características espaciais, o sistema é caracterizado como sendo relativamente compacto (aproximando-se a uma forma circular) e como sendo o segundo, de entre os sistemas analisados, com a distribuição mais uniforme, ou seja em que a variação da distância entre

⁵load factor

estações varia menos em função da localização de cada estação.

Com um enfoque numa análise temporal dos dados com o objetivo de descobrir padrões de utilização, vários estudos utilizam técnicas de agrupamento ou análise de *clusters*. [Feng et al., 2017] analisam a evolução da taxa de ocupação das estações do sistema parisiense de bicicletas compartilhadas, "Vélib", através de dois métodos de análise de *clusters*: K-means e Agrupamento hierárquico. Os padrões temporais são avaliados de forma diferente consoante se pretende analisar o padrão de utilização ao longo do dia ou especificamente durante os períodos de ponta da manhã e da tarde. Os autores obtêm através deste método diferentes categorias, que correspondem a padrões de utilização e que encontram relação com o tipo de atividades predominantes nas diferentes zonas da cidade. Também [Vogel et al., 2011] recorrem a métodos de *clusters* para descobrir padrões de utilização desde uma perspetiva espaço-temporal. Neste estudo são utilizados três métodos diferentes para analisar a atividade das estações ao longo do dia: K-means, algoritmo EM - *Expectation-Maximization* e *sequential Information Bottleneck* (sIB).

Também recorrendo a modelos de *data-mining* de análise de *clusters* [Randriamanamihaga et al., 2014], propõem a utilização de modelos de mistura de Poisson e de um algoritmo EM - *Expectation-Maximization* para estimação dos parâmetros, com o objetivo encontrar padrões em vetores de origem-destino referentes ao sistema de bicicletas compartilhadas de Paris, "Vélib". Os autores desenvolvem três modelos, com diferente grau de complexidade, sendo que o mais complexo apresenta três declinações que permitem identificar diferentes tipos de padrões: diferença de utilização em semana e durante o fim de semana, utilização específica conforme o dia da semana e influência das condições meteorológicas.

Com um enfoque diferente dos anteriores, [Tran et al., 2015] recorrem a um modelo de regressão linear robusto para modelar as viagens geradas e atraídas nas diferentes estações do sistema de bicicletas compartilhadas de Lyon "Vélo'v". Na construção do modelo os autores testam diferentes variáveis explicativas: rede de transportes públicos, variáveis sócio-económicas, variáveis que têm em conta a topografia, variáveis que incorporam diferentes indicadores do próprio sistema de bicicletas compartilhadas e variáveis ligadas às atividades de lazer. Dois modelos de regressão são realizados para o período da manhã e para o período da tarde. O estudo conclui que existe uma correlação possível positiva entre a densidade do sistema, a capacidade das estações e a utilização das mesmas. Da mesma forma, a proximidade de estações de comboio afeta positivamente a utilização das estações assim como a população e o número de empregos. No que diz respeito aos utilizadores ocasionais ⁶, a atividade nas estações durante o período de ponta da tarde é também explicada pelas variáveis relacionadas com o lazer. Por outro lado, a altitude parece influenciar negativamente as viagens geradas e atraídas pelas estações.

No artigo "*Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus*", [Griffin and Jiao, 2015] analisa de que forma variáveis sócio-económicas e topográficas, assim como as variáveis que dizem respeito à rede viária e à existência ou não de infraestruturas cicláveis, afetam o ciclismo desportivo em Travis County, no Texas. Os autores analisam os dados correspondentes

⁶short term subscribers

aos quilómetros percorridos em bicicleta obtidos a partir de uma plataforma online que por sua vez são agregados, junto com as restantes variáveis por setor estatístico para posterior modelação através de um modelo de regressão linear. Dada a auto-correlação espacial da variável dependente os autores utilizam um modelo de regressão geograficamente ponderada (GWR - (Geographically Weighted Regression)), com as mesmas variáveis independentes utilizadas no modelo de regressão linear. O estudo conclui que os ciclistas desportivos parecem ser pouco sensíveis à presença ou não de infraestruturas dedicadas aos ciclistas e preferem zonas com forte inclinação e baixa densidade populacional ou de empresas.

Procedimento idêntico ao anterior foi usado por [Yang et al., 2017] para estimar o volume de peões e ciclistas em cruzamentos, em Chittenden County, no estado do Vermont, Estados Unidos. Tendo como base contagens de peões e ciclistas em cruzamentos, variáveis sócio-económicas e de uso do solo, características dos cruzamentos e da rede viária assim como da acessibilidade aos transportes públicos, os autores constroem numa primeira fase um modelo de regressão linear e numa segunda fase dois modelos de regressão geograficamente ponderada diferentes: um tradicional, em que todas as variáveis independentes são locais e um misto, com variáveis dependentes locais e globais. As variáveis são agregadas tendo em conta uma área de serviço de 300 m para cada uma das interseções viárias. A capacidade de generalização do modelo é avaliada através do método *k-fold*.

Uma abordagem diferente das anteriores é adotada por [Noland et al., 2016], num estudo sobre a procura em estações de bicicletas partilhadas em Nova Iorque. O objetivo deste trabalho de investigação é não só compreender os fatores que influenciam a procura, mas também criar um modelo de previsão do número de viagens. Uma vez que o tipo de dados corresponde a contagens (ou seja observações com números inteiros não negativos), a normalidade dos erros não é assegurada. Por outro lado o modelo apresenta heterocedasticidade. Como consequência um modelo de regressão binomial negativa com heterocedasticidade condicional auto-regressiva é utilizado. Na elaboração do modelo a área de serviço das estações é calculada através de polígonos de Thiessen, também designados polígonos de Voronoi, truncados a 1/4 de milha de forma integrar as variáveis explicativas. Os autores concluem que a abordagem de inferência dos fatores que afetam a procura é um método que garante bons resultados, mas que a estimação da procura para novas estações não consegue ser validada. A dificuldade de criar um modelo preditivo prende-se com o facto do crescimento da rede afetar a procura nas estações vizinhas, assim como da grande variabilidade da procura em função da meteorologia.

Já [Levy et al., 2017], utilizam diferentes metodologias na exploração dos dados dos sistema de bicicletas partilhadas de Tel Aviv, em Israel. Os autores começam por comparar o número de viagens com origem ou com destino nas diferentes estações. Numa segunda fase, com base em dados dos pares origem destino, os autores analisam os padrões das viagens para quatro categorias de distância diferentes: menos de 2km, 2km - 4km, 4km - 6km e mais de 6 km. De forma a estimar os itinerários dos diferentes pares origem-destino, os autores partem da hipótese que os ciclistas tendem a utilizar o caminho mais curto e recorrem ao método de afetação de tráfico "Tudo ou nada", de forma a ter em conta o efeito do congestionamento na escolha dos itinerários. Conclui-se desta análise que as viagens mais curtas tendem a ser concêntricas

ao centro da cidade, sendo que as viagens mais longas estão localizadas ao longo da costa, entre o centro da cidade e o norte. Este estudo analisa também a relação entre o transporte público e os padrões de utilização da bicicleta partilhada. Na modelação desta relação é utilizado um modelo de regressão espacialmente ajustado (SAR). O estudo conclui que existe uma relação entre a atividade das estações de bicicleta partilhada e as viagens em autocarro, sendo essa relação positiva para as viagens em bicicletas mais curtas (menos de 4 km) e negativa para as mais longas. Os autores propõem que esse padrão de utilização é fruto da existência de ciclovias separadas do tráfego na periferia do centro, que possibilitam viagens em bicicleta mais rápidas que em transporte público e que como consequência concorrem com este nos trajetos mais longos.

3 O sistema de bicicletas partilhadas de Bruxelas

A área de estudo corresponde à região de Bruxelas-Capital, na Bélgica. A região tem uma área de 161,38 km² e uma população de 1.198.726 habitantes a 1 de Janeiro de 2018⁷.

Fazem parte da região de Bruxelas, 19 municípios, com densidades populacionais que variavam em 2016 entre 1093,58 habitantes / km², em Watermael-Boitsfort ao sul, e 23.988,51 habitantes / km² em Saint-Josse-ten-Noode, no centro. A densidade populacional média de toda a região era neste ano de 7.360,76 habitantes / km².

Não obstante a altitude relativamente baixa, Bruxelas não beneficia de um território plano como é o caso da quase totalidade da região da Flandres a Norte. O vale do rio Senne, que atravessa a cidade e a região num eixo sudoeste-noroeste, e as bacias afluentes moldam o relevo, o que, em algumas partes da cidade, pode condicionar a generalização da bicicleta enquanto modo de transporte quotidiano. A Figura 1 ilustra a orografia da Região de Bruxelas-Capital.

A utilização da bicicleta enquanto modo de transporte tem vindo a crescer de forma constante nos últimos anos, facto que confirma a crescente atratividade deste modo de transporte. Apesar da crescente popularidade da bicicleta, o sistema de bicicletas partilhadas de Bruxelas parece ter perdido alguma popularidade desde o ano de 2012. Há seis anos a percentagem de bicicletas Villo correspondia a mais de 7% do total, no ano de 2018 essa percentagem baixou para pouco mais de 4 % [Kesteloot et al., 2017], isto apesar do aumento do número de estações e de bicicletas.

O sistema Villo, conta hoje, em 2018, com cerca de 360 estações e mais de 5000 bicicletas⁸ distribuídas pelos 19 municípios da Região de Bruxelas-Capital. O sistema funciona através de um passe válido por um ano que pode ser integrado com um passe normal de transporte público. Contudo, qualquer pessoa não detentora de passe anual pode adquirir um bilhete válido por um dia ou por uma semana (7 dias consecutivos).

Os primeiros 30 minutos de utilização são gratuitos. Esta medida é válida tanto para os detentores de um passe ou para os utilizadores esporádicos (bilhete válido por um dia ou por uma semana). A partir dos 30 minutos de utilização, o consumidor paga conforme o tempo de utilização.

A progressividade das tarifas e a gratuitidade nos primeiros 30 minutos tem como objetivo favorecer as utilizações curtas de modo a incentivar a rotação das bicicletas.

As tarifas em vigor no ano de 2018⁹ são apresentadas em anexo (ver Tabela A.1).

⁷<http://ibsa.brussels/chiffres/chiffres-cles-de-la-region>, acedido a 15 de Agosto de 2018

⁸<http://www.villo.be/Villo!-C-est-12-millions-de-trajets-realises>

⁹<http://www.villo.be/Tarifs/Les-tarifs/Consultez-les-tarifs>, acedido a 15 de Agosto de 2018

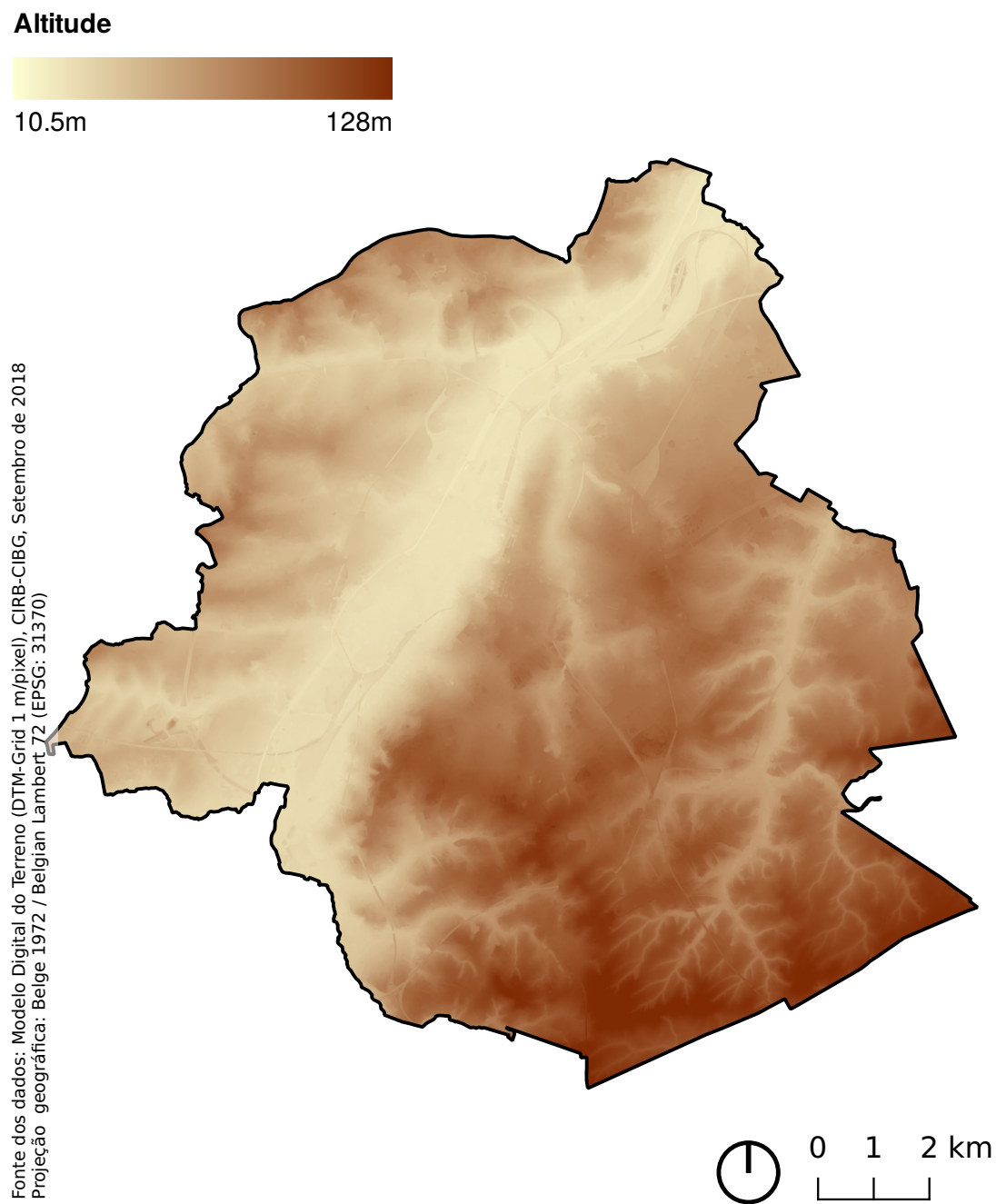


Figura 1: Orografia da região

4 Dados

Neste capítulo descreve-se separadamente o processo de recolha e tratamento dos dados correspondentes ao sistema Villo, aos indicadores sócio-económicos assim como aos dados referentes à infraestrutura e mobilidade.

4.1 Recolha e tratamento dos dados

O processo de recolha e tratamento de dados é detalhado na Figura 2, que sintetiza as diferentes fases do processo. Todo o processo de análise e tratamento de dados foi realizado utilizando Python e os programas QGIS, Postgresql / Postgis.

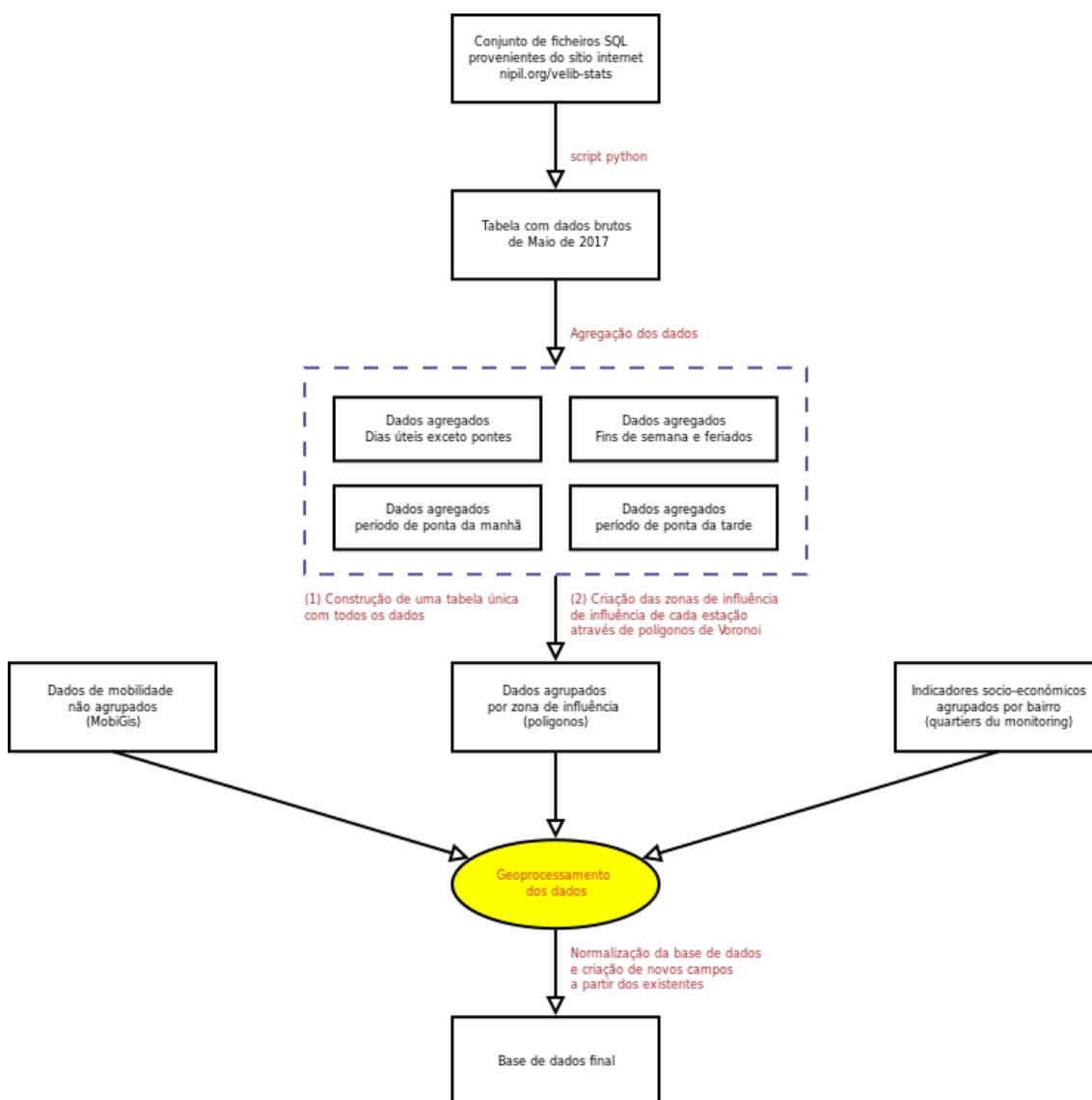


Figura 2: Processo de recolha e tratamento dos dados

4.2 Dados Villo

O operador do sistema Villo (JCDecaux), disponibiliza os dados de utilização das viagens em bicicleta em tempo real através de uma API¹⁰. Contudo, o histórico dos dados não é disponibilizado. Como alternativa, os dados foram obtidos através dos dados descarregados por um utilizador anónimo, obtidos entre Dezembro de 2015 e Abril de 2018 e disponibilizados no sítio da Internet seguinte: <https://nihil.org/jcdecaux.history.api.exports>.

Para além dos dados referentes ao sistema Villo, este sítio disponibiliza ainda dados de outros sistemas de bicicletas partilhadas operadas pela empresa JCDecaux.

A partir deste sítio da Internet foram descarregados 280.555 registos de entradas ou saídas de bicicletas nas diferentes estações Villo em Maio de 2017.

Uma vez descarregados os dados, foi efetuado um trabalho de tratamento e correção de dados redundantes ou incorretos. Desta forma, foram eliminadas todas as estações cujos dados apresentavam anomalias. É o caso de estações com atividade intermitente, ou sem atividade. Como resultado, apenas foram consideradas 342 estações.

O mês de Maio foi escolhido como período de referência por diferentes motivos:

Por um lado, uma campanha de contagens de bicicletas é realizada anualmente nos meses de Janeiro, Maio, Setembro e Novembro [asbl ProVelo,]. Este estudo anual disponibiliza dados sobre o volume de bicicletas observadas em diferentes pontos chave da cidade e distingue entre os ciclistas "normais" e os ciclistas Villo.

Por outro lado, pretende-se captar a procura num período do ano em que as viagens em bicicleta sejam o menos possível afetadas por condições meteorológicas excecionais, assumindo como verdadeiras as hipóteses que os meses de inverno e de outono são desfavoráveis para o ciclismo quotidiano e que episódios meteorológicos extremos afetam mais os modos suaves que outros modos de transporte. De fato, dos dados de [Kesteloot et al., 2017], entre 2010 e 2017, observam-se sistematicamente menos ciclistas nos meses de Janeiro e Novembro, que nos meses de Maio e Setembro. [Thomas et al., 2013], num estudo sobre a influência das condições meteorológicas na prática do ciclismo, em Gouda, Países Baixos, conclui, com base num modelo de regressão, que a temperatura média do ar em 24 horas, a precipitação, as horas de luz e a velocidade do vento, são por esta ordem, os fatores que mais influenciam a procura.

Por último, exclui-se o mês de Setembro, dado que é um mês excepcional no que diz respeito à utilização da bicicleta. A realização da semana da mobilidade que tem lugar nesta altura do ano implica a existência de um grande número de atividades de promoção dos modos suaves que leva a um "aumento artificial" do número de viagens em bicicleta neste período do ano.

A agregação dos dados em quatro períodos tem como objetivo conhecer de que forma o fator temporal influencia a maior ou menor procura nas diferentes estações. Além da agregação dos dados por dias úteis

¹⁰<https://developer.jcdecaux.com>

e fins de semana e feriados, foram ainda criadas duas subdivisões: período de ponta da manhã, entre 07h00 e 10h00 e período de ponta da tarde, entre 16h00 e 19h00. Os períodos de ponta foram definidos tendo em conta as horas de maior atividade no conjunto das estações do sistema Villo como mostra a Figura 3.

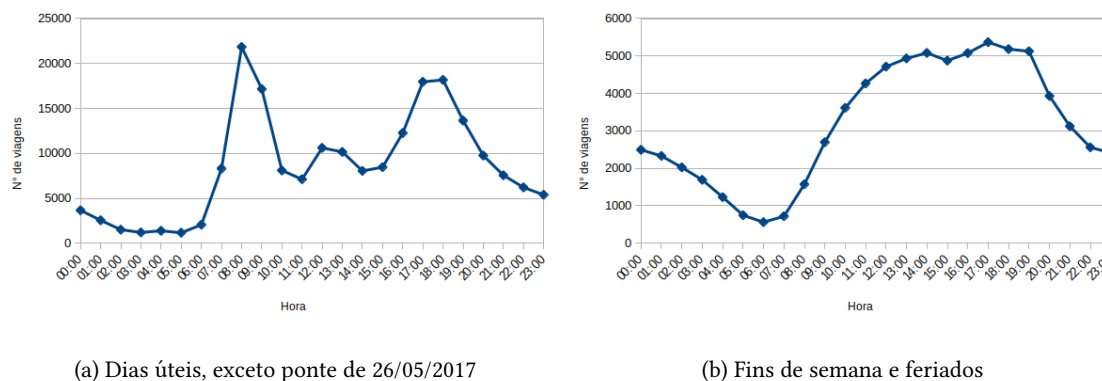


Figura 3: Número de viagens registadas por hora no conjunto das estações Villo em Maio de 2017

De um ponto de vista metodológico é razoável considerar uma distância limite que os utilizadores estão dispostos a percorrer entre o ponto de origem ou destino e a estação. [O'Brien et al., 2014], por exemplo utilizam um raio de 1 km. Não deixando de considerar essa distância generosa, os autores tentam evitar raios de influência mais pequenos, que poderiam criar "buracos" no interior das zonas em análise. Por sua vez, [Tran et al., 2015], consideram um raio de 300 m, como uma distância razoável que qualquer pessoa está disposta a percorrer a pé.

Analogamente aos estudos de [Noland et al., 2016] e [Levy et al., 2017] optou-se por utilizar polígonos de Voronoi na definição das zonas de influência das estações partindo da hipótese que os utilizadores procuram a estação mais próxima. Esta metodologia é uma simplificação uma vez que os polígonos de Voronoi são construídos tendo em conta a distância euclidiana. Uma alternativa seria calcular a zona de influência tendo em conta a rede viária. Contudo, tendo em conta a densa rede viária de Bruxelas, é razoável considerar que a distância a pé em linha reta é diretamente proporcional à distância real e que o fator de proporcionalidade é mais ou menos constante em todo o território.

Numa primeira fase, as zonas de influência das estações foram ainda delimitadas tendo em conta um raio de 400 m, que corresponde a um tempo de percurso a pé de cerca de 5 min. Desta análise concluiu-se que as zonas de influência corresponderiam a 58 % do território da região, ou seja mais de metade da região encontra-se a menos de 400 m de uma estação Villo.

Apesar da razoabilidade de considerar uma tal distância, haveria um conjunto de ilhas que não seria analisado, além de que é razoável pensar que alguns utilizadores estão dispostos a percorrer distâncias superiores a 400 m entre a origem ou o destino final e a estação. Por estas razões, uma distância limite de 1200 m foi considerada, sendo que nas zonas onde as estações estão muito próximas do limite regional, é este que delimita as zonas de influência, como se pode observar na Figura 4.

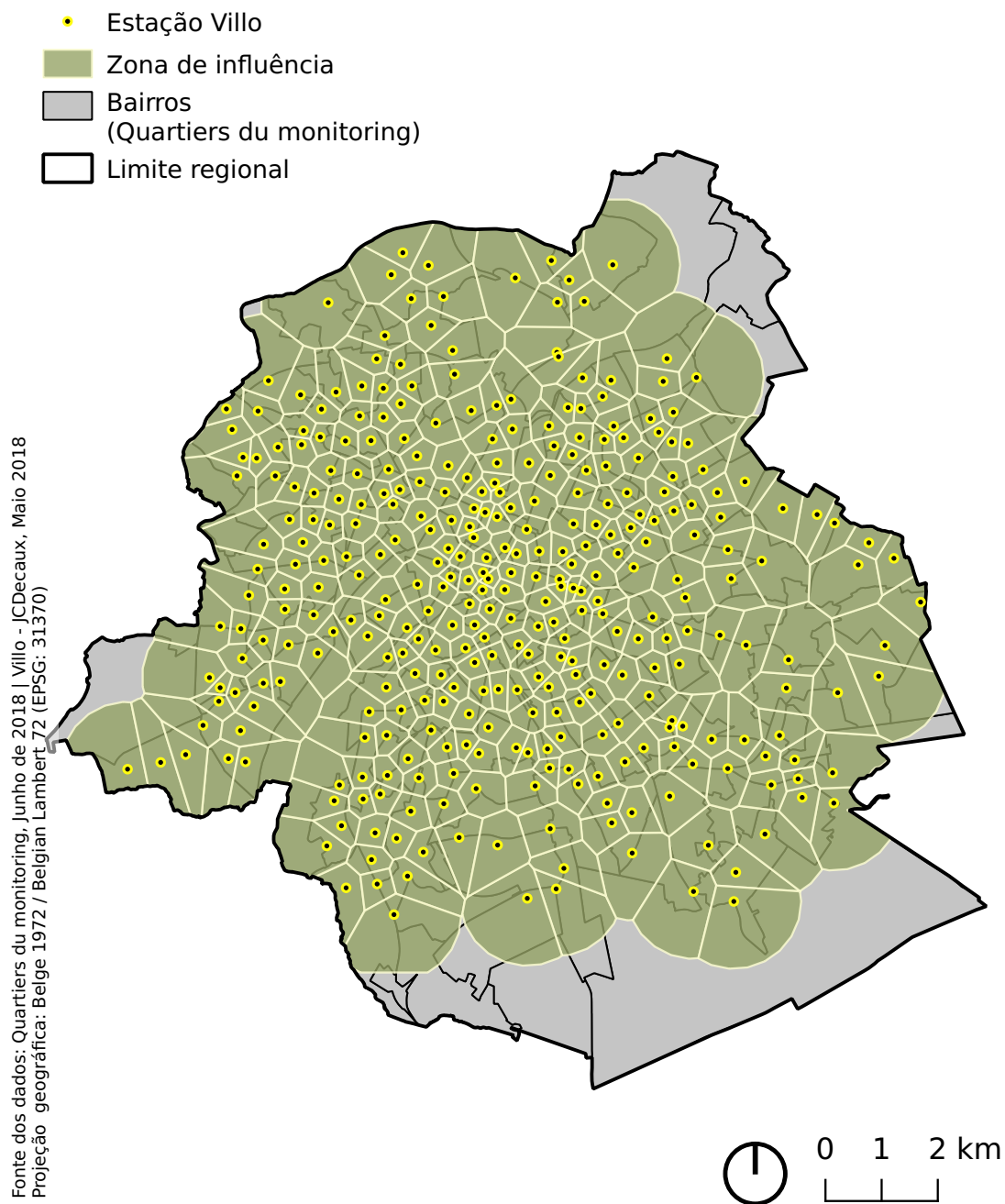


Figura 4: Estações Villo, zonas de influência (polígonos de Voronoi) e limites administrativos

Com base nas considerações do parágrafo anterior, calculou-se o rácio entre a área das zonas de influência pela área total da região, tendo-se concluído que as zonas de influência do conjunto das estações correspondem a cerca 86 % do território da região de Bruxelas-Capital.

As zonas de influência correspondem assim à quase totalidade do território, sendo que as zonas situadas longe das estações Villo são todas situadas na periferia, sendo que a maior parte dessas áreas corresponde a zonas com uma densidade de população e de serviços muito baixa, ocupadas maioritariamente por floresta

ou por zonas desabitadas e sem atividade comercial. Dois exemplos disso, são a extensa zona florestal situada a sul e a zona de ocupada por estaleiros ferroviários, por indústria pesada e por um estabelecimento prisional a nordeste.

A utilização de polígonos de Voronoi para a criação das zonas de influência das estações está descrita na literatura, tendo sido o método usado por [Levy et al., 2017] na subdivisão de Tel Aviv ou por [Noland et al., 2016], em Nova Iorque. No caso deste último, a zona de influência é ainda limitada a um raio de aproximadamente 385m ¹¹.

4.3 Dados sócio-económicos

Os dados sócio-económicos foram obtidos a partir dos dados do Instituto Bruxelense de Estatística e Análise¹², que disponibiliza dados para diferentes secções estatísticas: bairros ou municípios. No âmbito deste trabalho, de maneira a obter informação detalhada foram utilizados dados correspondentes aos 145 bairros de Bruxelas. Estes bairros foram concebidos como entidades de observação urbana e por esse motivo uma boa parte dos indicadores estatísticos estão disponíveis para esta porção do território. Dentro das 145 unidades estatísticas: 118 correspondem a bairros residenciais (onde reside 99,7 % da população), 6 zonas industriais ou ferroviárias, 18 espaços verdes e 3 cemitérios¹³.

Alguns indicadores dos bairros não residenciais apresentam valores omissos por terem uma população muito baixa. É o caso dos indicadores de 21 dos 27 bairros em questão. Os valores omissos são fruto da existência de um valor de população limite, que varia consoante o indicador, para o qual os dados são disponibilizados. Partindo deste pressuposto, optou-se por considerar como 0 as variáveis dos bairros com menos de 200 habitantes. No fundo, estes bairros correspondem a zonas sem população e com atividade económica irrelevante.

Uma vez tratados os indicadores sócio-económicos, obtiveram-se novos indicadores ($Indicador_{seZI}$), determinados a partir da proporção da área de um determinado bairro no polígono de Voronoi intersectado. Assim, para uma determinada zona de influencia (ZI), o indicador sócio-económico ($Indicador_{seZI}$) é calculado de acordo com a expressão seguinte:

$$Indicador_{seZI} = \frac{\sum_{bairro} Area_{Inters.} \times Indicador_{bairro}}{Area_{ZI}} \quad (1)$$

em que $Area_{Inters.}$ corresponde à área da interseção geométrica entre o polígono de Voronoi, ou zona de influência (ZI) e os polígonos dos bairros, de acordo com a expressão (2).

O parâmetro $Indicador_{bairro}$ corresponde ao indicador sócio-económico de cada um dos bairros analisa-

¹¹1/4 de milha

¹²<http://ibsa.brussels/>, acedido a 18 de Abril de 2018

¹³<https://monitoringdesquartiers.brussels/a-propos>, acedido a 18 de Abril 2018

dos e a $Area_{ZI}$ corresponde à área total da zona de influência.

$$Area_{Inters.} = Area(Geom_{bairro} \cap Geom_{ZI}) \quad (2)$$

Note-se que sempre que o polígono de um bairro não intersecta a zona de influência, a área da interseção é igual a zero e por conseguinte o indicador de um bairro que não intersecta a zona de influência não é contabilizado.

4.4 Outros dados (mobilidade, infraestrutura, etc)

Os dados referentes à infraestrutura e mobilidade foram obtidos através do geo-portal do serviço de mobilidade do governo regional de Bruxelas¹⁴. Este sítio na Internet disponibiliza dados vetoriais e vários temas relacionados com a mobilidade.

Os dados obtidos foram agregados espacialmente em polígonos de Voronoi (ZI) correspondentes às zonas de influência das estações Villo através da interseção geométrica entre os diferentes indicadores de mobilidade e os polígonos de Voronoi (ZI).

4.5 Base de dados final

O resultado final deste processo é uma base de dados espacial com as diferentes variáveis descritas agregadas por polígonos de Voronoi, que aqui se designam por zonas de influência (ZI).

A base de dados é composta por 49 variáveis:

- uma chave primária
- 19 variáveis relacionadas com o sistema Villo, de entre as quais 4 variáveis dependentes;
- 12 variáveis sócio-económicas;
- 15 variáveis relativas à infraestrutura viária, ciclável e a diversos aspetos relacionados com a mobilidade;
- 2 variáveis que correspondem às coordenadas, no sistema de projeção Lambert 72 (EPSG: 31370), das estações;
- 1 variável que corresponde à área dos polígonos;

Cada uma destas variáveis é descrita na Tabela A.2, em Anexo.

¹⁴<http://data-mobility.irisnet.be/mobigis/fr/>, acedido a 18 de Abril de 2018

5 Metodologia

Neste capítulo descrevem-se as diferentes etapas de análise e modelação. As opções da abordagem metodológica estão sintetizadas na Figura 5 e são abordadas detalhadamente nas diferentes secções deste capítulo.

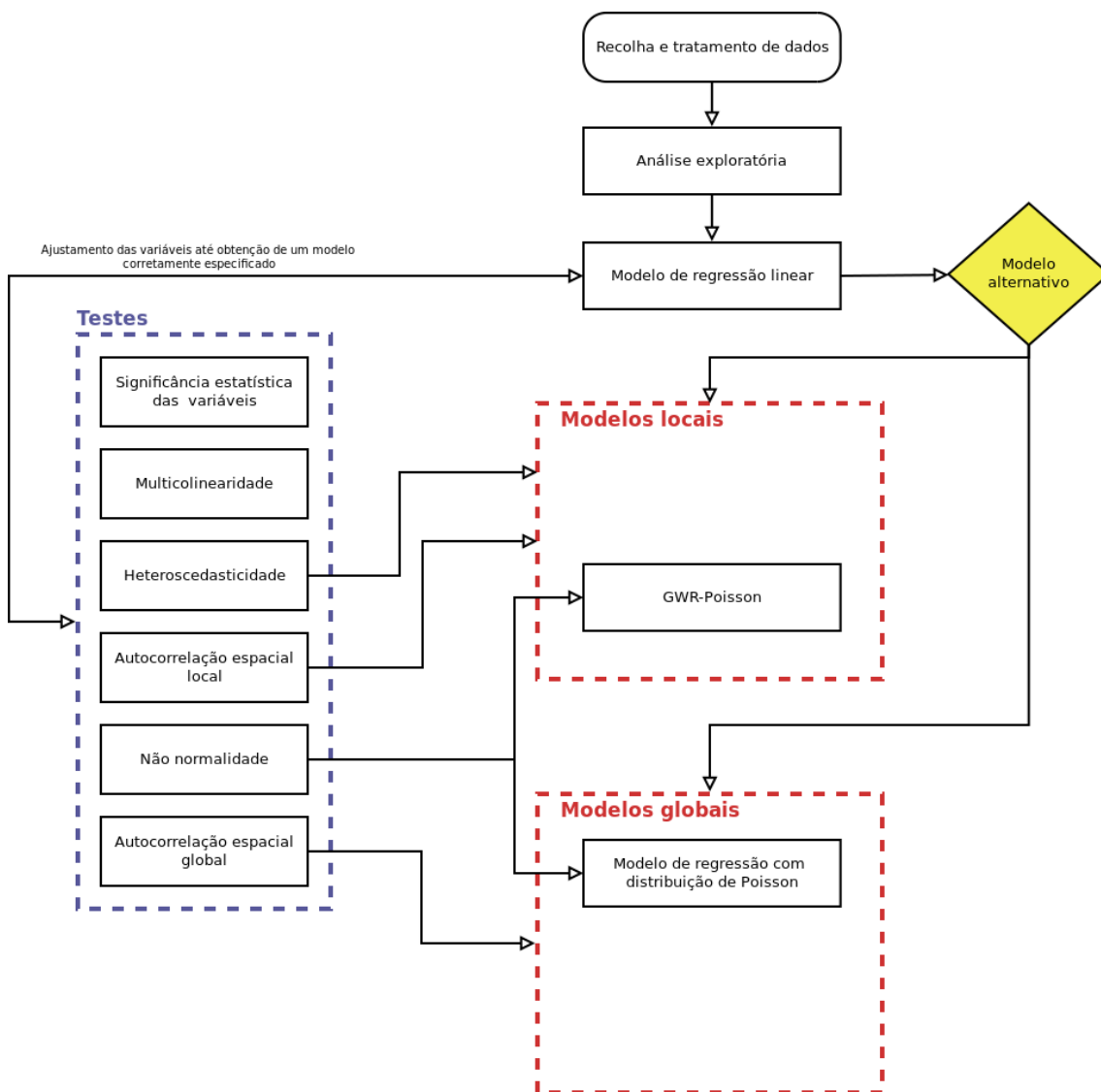


Figura 5: Metodologia

Partindo do objetivo de analisar os fatores sócio-económicos, de infraestrutura e de uso do solo que influenciam as viagens nas diferentes estações de bicicletas partilhadas de Bruxelas, assim como perceber de que forma esses fatores são afetados consoante o período de análise, optou-se por considerar as variáveis dependentes seguintes:

- N° de viagens realizadas durante os dias úteis (exceto pontes), ou seja todos os dias úteis do mês

24/24h exceto o dia 26 de Maio. O período de análise corresponde a 480 horas;

- N° de viagens realizadas durante o período de ponta da manhã, entre 7h00 e 10h00 de todos os dias úteis. O período de análise corresponde a 60 horas;
- N° de viagens realizadas durante o período de ponta da tarde, entre 16h00 e 19h00 de todos os dias úteis. O período de análise corresponde a 60 horas;
- N° de viagens realizadas durante os fins de semana e feriados 24/24h período de análise corresponde a 240 horas;

Sendo o enfoque deste trabalho o ciclismo quotidiano, uma análise mais cuidada será reservada aos resultados que dizem respeito aos dias úteis, nomeadamente na análise dos resultados dos períodos de ponta da manhã e da tarde.

Esta abordagem toma como hipótese que nos fins de semana, as bicicletas partilhadas são geralmente utilizados no âmbito de lazer ou de turismo e que durante a semana existe uma utilização mista, sendo que uma parte significativa corresponde a viagens pendulares domicílio-trabalho ou domicílio-escola.

Uma análise exploratória dos dados referentes à atividade das estações nos fins-de-semana é igualmente realizada de forma a identificar diferenças com os dias úteis.

As variáveis independentes ou explicativas no modelo de regressão foram escolhidas, primeiramente com base na literatura científica abordada no segundo capítulo e numa segunda fase recorrendo a um modelo de regressão linear exploratório.

Nas diferentes secções deste capítulo abordam-se detalhadamente as principais etapas da metodologia utilizada e justificam-se as opções tomadas no que respeita à escolha das variáveis, dos modelos e da parametrização dos mesmos.

5.1 Análise exploratória dos dados

O objetivo da análise exploratória de dados espaciais é identificar padrões globais e locais de associação espacial que indiquem a existência de não-estacionaridade local, ilhas de heterogeneidade espacial, etc. Um aspeto crucial na definição da auto-correlação espacial é a determinação das relações de vizinhança relevantes de uma dada zona, isto é, as zonas próximas de um dado ponto que são suscetíveis de influenciar as observações nesse ponto [M. Fischer and Wang, 2011].

5.1.1 Análise de *clusters* e *outliers* espaciais

A análise de *clusters* e *outliers* espaciais é realizada com base em indicadores locais de associação espacial¹⁵ (LISA) propostos por [Anselin, 1995], utilizando o indicador Moran's I (local) como forma de identificar

¹⁵Local Indicators of Spatial Association

clusters na variável dependente, ou seja zonas onde existe auto-correlação espacial local positiva, e *outliers*, que correspondem a zonas onde valores dissemelhantes surgem agrupados no mapa (auto-correlação espacial local negativa).

A criação de uma matriz espacial constitui uma etapa prévia necessária para a realização desta análise.

Neste contexto, definiu-se uma matriz de pesos espaciais que corresponde a uma matriz de vizinhança baseada nas relações de contiguidade. Optou-se por relações de contiguidade do tipo *queen* de primeiro grau. Ou seja, para um dado polígono, são considerados vizinhos os polígonos que partilham os lados ou os vértices.

Optou-se por uma matriz de contiguidade de primeiro grau, tendo em conta que a alternativa a uma dada estação é a estação vizinha mais próxima. Este pressuposto, só não é completamente válido, nos casos em que a estação não apresente bicicletas disponíveis nos casos de saída ou se encontre completamente ocupada nos casos de chegada. Também, nos casos em que as estações estão muito próximas, poder-se-ia considerar uma relação de contiguidade de ordem superior.

Optou-se por descartar a matriz de distâncias, uma vez que existe uma grande diferença entre as distâncias das estações no centro da cidade e na periferia, sendo que a distância mínima para que todas as estações tenham pelo menos um vizinho é de 1135 m.

O histograma apresentado na Figura 6 representa a distribuição do número de vizinhos. A média corresponde a 5,7 vizinhos e a mediana a 6 vizinhos. Note-se que todos os polígonos têm pelo menos um vizinho. As zonas periféricas são obviamente aquelas com menos vizinhos, dado o maior tamanho e a situação geográfica. As zonas com maior densidade de estações Villo são mais pequenas e como consequência apresentam um maior número de vizinhos.

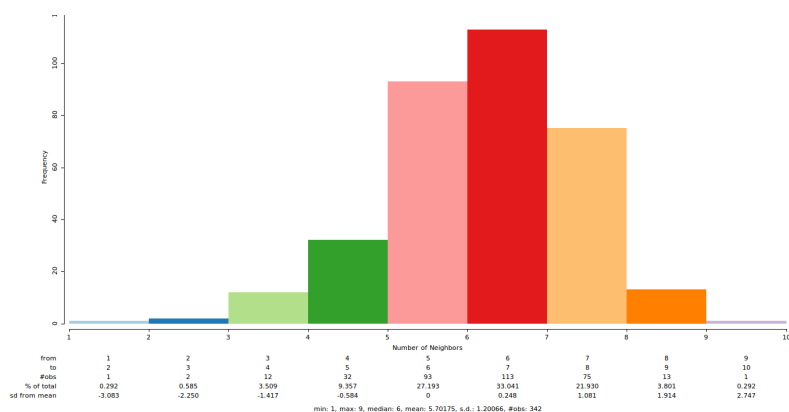


Figura 6: Histograma matriz de pesos

A obtenção da matriz de vizinhança é uma etapa fundamental não só no âmbito da análise exploratória, mas também para avaliar a auto-correlação espacial dos resíduos, no caso de modelos baseados numa distribuição normal dos resíduos.

Uma vez obtida a matriz de contiguidade e com base nos dados referentes à procura de viagens nas estações

é possível avaliar as zonas do território com maior ou menor procura, e de que forma a procura é influenciada pelas relações de vizinhança.

5.1.2 Regressão linear exploratória

A realização da análise de regressão linear exploratória tem como principal objetivo descartar um certo número de variáveis incluídas inicialmente na base de dados, com base no facto de estas não serem significativas na explicação das variáveis dependentes, ou que apresentem multicolinearidade com outras variáveis. Note-se que a base de dados foi elaborada tendo em conta as variáveis explicativas utilizadas em estudos no mesmo âmbito, com diferentes declinações dependendo da disponibilidade ou não de dados. Neste contexto, uma análise de regressão exploratória foi realizada para cada uma das variáveis dependentes referentes aos dias úteis (dia inteiro, período de ponta da manhã, período de ponta da tarde), assim como fins de semana e feriados, a partir de diferentes combinações de 30 variáveis variáveis potencialmente explicativas (variáveis independentes).

A partir da realização da análise exploratória foi possível identificar as variáveis independentes que melhor explicam cada uma das variáveis dependentes em análise. A seleção das variáveis independentes foi realizada tendo em conta os resultados dos *testes-t*, sendo consideradas significativas aquelas que apresentam um *p-value* inferior a 0.05.

A existência ou não de multicolinearidade foi outro elemento tido em conta na escolha das variáveis explicativas, uma vez que a escolha de variáveis relacionadas entre si pode afetar negativamente os resultados do modelo.

A existência de variáveis com um VIF (*Variation Inflation Factor*) superior a 10 é apontada como um indicador da redundância de uma dada variável, contudo outros autores sugerem valores inferiores a 10 [Zuur et al., 2010]. Um valor superior a 30 de outro indicador, o *multicollinearity condition number*, sugere a existência de problemas de multicolinearidade [Anselin, 2005]. No âmbito deste trabalho, foram tidos em conta ambos indicadores, uma vez que existem casos em que todas as variáveis independentes apresentam um VIF inferior a 10 e um *multicollinearity condition number* superior a 30. Assim, a abordagem usada neste trabalho foi a de reter as combinações das variáveis independentes que configuram os modelos de regressão linear melhor ajustados. Isto, mesmo se o modelo apresenta valores de *multicollinearity condition number* superior a 30, desde que o VIF de cada uma das variáveis alternativas seja inferior a 10. Ao mesmo tempo, escolheu-se uma outra combinação de variáveis explicativas que melhor ajuste o modelo, mas que respeite a condição de multicolinearidade (*multicollinearity condition number* inferior a 30).

De acordo com [Levine et al., 2010], os pressupostos do modelo de regressão linear são a independência das observações, a normalidade e homocedasticidade dos resíduos, assim como a não existência de autocorrelação espacial nos resíduos. De forma a testar estes pressupostos foram utilizados os seguintes testes, disponíveis no *software* Geoda [Anselin, 2005]:

- Teste de Jarque-Bera que toma como hipótese nula a distribuição normal dos resíduos;
- Testes de Breusch-Pagan e de Koenker-Bassett. Ambos testam os pressupostos de homocedasticidade dos resíduos, assumindo como hipótese nula a estacionaridade espacial destes, sendo que o teste de Koenker-Basset assume especial relevância nos casos em que a distribuição normal dos resíduos não é observada;
- Testes de Moran's I global e local, que testam respetivamente se existe auto-correlação espacial nos resíduos e as zonas onde esta auto-correlação é estatisticamente significativa;

Por sua vez o ajuste do modelo é avaliado tendo em conta os seguintes indicadores (ver [Anselin, 2010]):

- R^2 , que corresponde à percentagem de variância da variável dependente explicada pelo modelo. Este indicador varia entre 0 e 1. Quanto mais elevado for o valor, mais bem ajustado é o modelo;
- R^2 ajustado. Este indicador que corresponde a uma declinação do primeiro, ajustado aos graus de liberdade, permite avaliar o ajustamento do modelo, compensando o aumento artificial do primeiro sempre que são adicionadas variáveis explicativas ao modelo;
- Akaike Information Criterion (AIC) e Schwartz Criterion (SC), ambos construídos a partir do indicador *maximized log-likelihood*, que permitem comparar diferentes especificações do modelo. O modelo melhor ajustado é considerado aquele que apresenta o valor mais baixo destes indicadores;

Como mostrado na Figura 5, a adoção de um modelo alternativo é considerada a partir da verificação dos pressupostos do modelo de regressão linear. A violação destes pressupostos serve assim de guia na escolha dos modelos a adotar.

5.2 Análise de Regressão Geograficamente Ponderada (GWR)

Uma violação dos pressupostos do modelo de regressão linear antecipada é a não estacionaridade espacial das variáveis dependentes. De acordo com a primeira lei da geografia proposta por [Tobler, 1970], as relações de proximidade têm influência no comportamento dos fenómenos, violando por isso os pressupostos estatísticos de independência das variáveis.

A regressão geograficamente ponderada apresenta-se, neste contexto, como uma técnica alternativa aos modelos de regressão globais, que se baseia nestes últimos, mas que tem em conta as relações espaciais locais [Fotheringham et al., 2002].

No caso dos modelos de regressão linear global, a variável independente é calculada pela equação seguinte :

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i \quad (3)$$

em que y_i representa a variável dependente no ponto i , x_{ik} uma dada variável explicativa ou independente k , no ponto i , β_0 , um coeficiente constante, β_k o coeficiente de uma dada variável explicativa k e ε_i o erro associado a cada ponto, ou seja a variação de y que não é explicada pelo modelo.

De acordo com [Fotheringham et al., 2002], a GWR integra na equação geral as coordenadas do ponto i da seguinte forma:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \quad (4)$$

Note-se que os parâmetros $\beta_k(u_i, v_i)$ são estimados a partir de pontos próximos de i , de maneira a que os pontos mais próximos de i tenham mais peso na estimação de $\beta_k(u_i, v_i)$, que aqueles que estão situados mais longe.

A maneira como os parâmetros $\beta_k(u_i, v_i)$ são estimados é definida através de uma função de ponderação (*spatial weighting function*) que pode ser fixa ou variável (*kernel* fixo ou adaptativo). Neste estudo utilizou-se uma função de ponderação adaptativa do tipo *bi-square kernel*, que se ajusta em função da maior ou menor densidade de pontos no espaço, conforme a equação seguinte:

$$w_{ij} = \begin{cases} (1 - d_{ij}^2/\theta^2)^2 d_{ij} < \theta_{i(k)} \\ 0 d_{ij} < \theta_{i(k)} \end{cases} \quad (5)$$

em que w_{ij} é o valor do peso da observação no local j para estimar o coeficiente na localização i , d_{ij} é a distância euclidiana entre i e j e $\theta_{i(k)}$ corresponde à largura de banda¹⁶ adaptativa definida a partir da distância do k^{th} *nearest neighbour* [Nakaya et al., 2016]. Ou seja, ao contrário do *kernel* fixo, em que a distância é definida pelo utilizador, no caso de um *kernel* adaptativo a largura deste varia em função do número de vizinhos.

A seleção do número de vizinhos a considerar no cálculo da largura de banda é realizada automaticamente através do *software* GWR 4.0 tendo em conta o número de vizinhos que minimiza o AICc.

O AICc é uma variante do AIC, corrigido de forma a ter em conta graus de liberdade pequenos, como é o caso dos modelos de regressão local [Nakaya et al., 2005]. Este indicador servirá também de comparação entre as diferentes especificações do modelo GWR.

5.3 Modelos com distribuição de Poisson

A não normalidade dos resíduos é antecipada uma vez que as variáveis dependentes utilizadas correspondem a dados de contagens, ou seja, valores inteiros maiores que zero.

Nestes casos a utilização de modelos com distribuição de Poisson é mais adequada, uma vez que a utilização da função exponencial permite obter valores superiores a zero [Anselin, 2010].

¹⁶bandwidth

Vários estudos têm procurado introduzir os efeitos da componente espacial na modelação de dados de contagens (ver [Simões and Natário, 2016], [Glaser, 2017]). Neste trabalho, e tendo em conta a utilização de modelos GWR no âmbito dos transportes, utiliza-se um modelo GWR com distribuição de Poisson proposto por [Fotheringham et al., 2002].

Tal como nos modelos de regressão linear os resultados do modelo local são comparados com os resultados do modelo global (Regressão de Poisson). De acordo com [Nakaya et al., 2005] os modelos global e local apresentam-se respetivamente de acordo com as fórmulas (6) e (7):

$$\hat{y}_i \sim Poisson[\exp(\sum_k \beta_k x_{k,i})] \quad (6)$$

onde \hat{y}_i é o valor calculado da variável dependente, $x_{k,i}$ é a variável explicativa k no ponto i e β_k o coeficiente k .

O segundo corresponde a um modelo de regressão geograficamente ponderada com distribuição de Poisson (GWRP) que toma a forma:

$$\hat{y}_i \sim Poisson[\exp(\sum_k \beta_k(u_{xi}, u_{yi})x_{k,i})] \quad (7)$$

onde (u_{xi}, u_{yi}) é um vetor bidimensional que corresponde às coordenadas do ponto i .

Como extensão desta última equação e tendo em conta o fato que algumas variáveis explicativas poderão ser espacialmente estacionárias, ou seja, globais, realizou-se o GWRP semi-paramétrico com variáveis locais e globais. A especificação do modelo proposta por [Nakaya et al., 2009] é a seguinte:

$$\hat{y}_i \sim Poisson[\exp(\sum_k \beta_k(u_{xi}, u_{yi})x_{k,i} + \sum_l \gamma_l z_{l,i})] \quad (8)$$

onde, γ_l é um coeficiente que permanece constante no espaço associado à variável explicativa $z_{l,i}$

Note-se que ao contrário dos modelos clássicos de regressão linear, o termo correspondente ao erro aleatório não existe nos modelos de regressão com distribuição de Poisson [Anselin, 2010].

A comparação entre os modelos de Poisson global e local é realizada tendo em conta o indicador AICc, cuja formulação para o modelos de Poisson é dada por [Nakaya et al., 2005]. O modelo melhor ajustado é aquele que apresenta o AICc mais baixo.

6 Resultados

6.1 Resultados da análise exploratória

Numa primeira fase realizou-se uma análise de auto-correlação espacial para cada uma das variáveis dependentes em análise, incluindo o número de viagens Villo nos fins de semana e feriados. Utilizou-se uma significância do indicador ao nível de 99 %.

Os resultados apresentados na Figura 7 permitem concluir que, com a exceção de dois *outliers* localizados perto do centro da cidade, a variável dependente apresenta auto-correlação espacial positiva em todos os períodos analisados, com *clusters* de valores elevados no centro da cidade e *clusters* de valores baixos na periferia.

O número de *clusters* é maior nos dias úteis que nos fins de semana, principalmente a sudeste do centro. A grande atividade nesta zona onde estão instalados uma grande parte dos escritórios das instituições europeias, contrasta com a pouca procura durante o fim-de-semana.

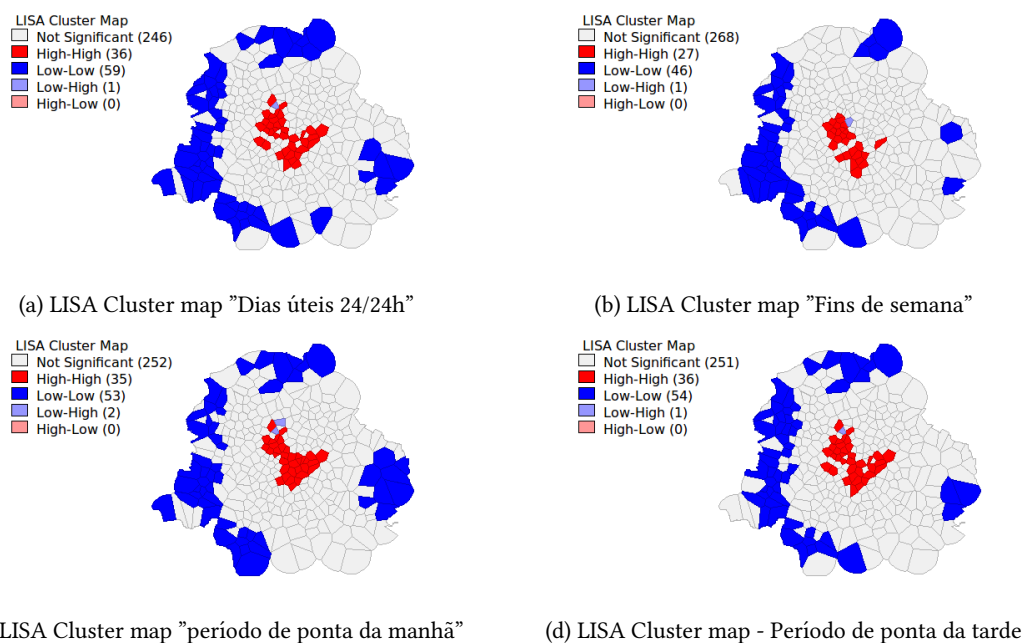


Figura 7: Análise da auto-correlação local da variável dependente

De forma a analisar a influência do período do dia na procura de viagens, realizou-se uma análise semelhante à anterior fazendo a distinção entre as viagens com origem (geração) ou destino (atração) nas diferentes estações. A Figura 8 apresenta os resultados dessa análise para o período de ponta da manhã.

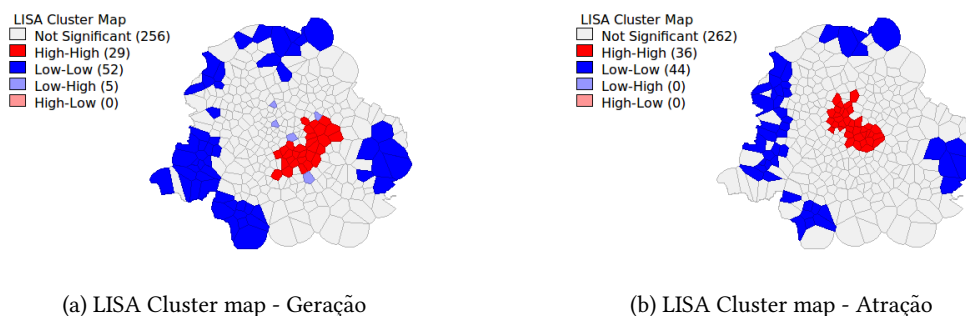


Figura 8: Análise da auto-correlação local - Geração e atração de viagens no período de ponta da manhã

A análise da auto-correlação local no período de ponta da manhã permite perceber que certas zonas do centro são ao mesmo tempo polos geradores e de atração de viagens. Contudo, observa-se um maior número de *clusters* no que respeita à geração de viagens numa zona ligeiramente periférica a sudeste do centro da cidade, enquanto que os *clusters* referentes à atração de viagens são situados no centro da cidade, com maior concentração de escritórios, serviços e atrações turísticas. Por sua vez, a dimensão temporal parece ter pouca ou nenhuma influência no que respeita à atração ou geração de viagens nas zonas mais periféricas, próximas dos limites administrativos da região.

Da mesma forma, a Figura 9 apresenta os resultados que dizem respeito ao período de ponta da tarde.

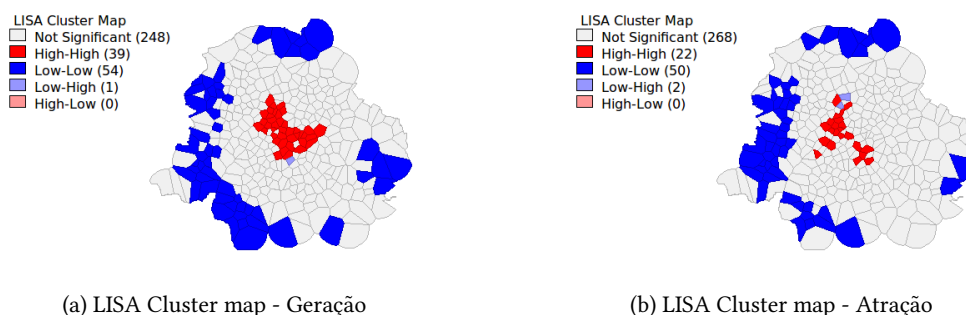


Figura 9: Análise da auto-correlação local - Geração e atração de viagens no período de ponta da tarde

Ao comparar-se os gráficos das Figuras 9a e 8b, observa-se que ambos são muito semelhantes, o que pode ser um indicador de movimentos pendulares casa-trabalho, dada a simetria das viagens. Por outro lado, entre as Figuras 8a e 9b, não há grande semelhança, o que, pode ser facilmente explicado pelo carácter multifuncional do centro da cidade, que permite a existência de fluxos importantes ao longo de todo o dia.

6.2 Análise de regressão linear exploratória (OLS exploratório)

Nesta secção apresentam-se os resultados da análise de regressão exploratória separadamente para as variáveis dependentes *Dias úteis* e *Fins de semana e feriados*, assim como para as variáveis dependentes *Período de ponta da manhã* e *Período de ponta da tarde*.

6.2.1 OLS exploratório - Dias úteis / Fins de semana e feriados

Os resultados da regressão linear exploratória, tendo como base as variáveis dependentes *Dias úteis* e *Fins de semana e feriados* são apresentados na Tabela 1:

Tabela 1: Resultados da regressão linear exploratória - Dias úteis / Fins de semana e feriados

Variável	Dias úteis		Fins de semana e feriados	
	Coefficiente	testes-t	Coefficiente	testes-t
Constante	-1825,17	-8,81963	-973,323	-13,7265
Número de docas	47,3934	11,7409	18,0889	11,1506
Porcentagem de adultos (18 - 64 anos)	19,8562	9,22022	9,18705	3,76719
Distância ao centro	-0,071137	-5,66985	-	-
Densidade de paragens de transporte público	1,52547	7,32487	0,705403	8,44156
Densidade estacionamentos para bicicletas	4,26419	5,34798	-	-
Densidade de população	-	-	0,00713072	6,72664
Porcentagem da área com calçada	-	-	15,695	7,48618
R²	0,768306		0,728014	
R² Ajustado	0,764858		0,723966	
Log Likelihood	-2427,82		-2123,51	
Akaike info criterion (AIC)	4867,64		4259,03	
Schwarz criterion	4890,65		4282,04	

Os resultados dos testes-t são aqui apresentados a título meramente informativo, uma vez que os resíduos do modelo não apresentam uma distribuição normal ¹⁷.

Por outro lado, existe o risco de multicolinearidade ¹⁸, provavelmente devido à presença das variáveis *densidade populacional* (dens_pop) e *porcentagem da população com idades compreendidas entres os 18 e 64 anos* (pc_1864).

Apesar do modelo não ser confiável, é interessante verificar que as variáveis se comportam conforme esperado, nomeadamente no que diz respeito ao sinal. Como seria de esperar, a variável *distância ao centro* (dtcc) apresenta sinal negativo, ao contrário de todas as outras.

A variável *porcentagem de zonas com calçada* (pc_bpave) parece ser uma variável redundante com a *distância ao centro* (dtcc) e substitui esta nos fins de semana. Mesmo que pareça contra-intuitivo o sinal positivo desta variável, uma vez que os ciclistas tendem a preferir revestimentos lisos, o fato é que estas zonas se encontram em geral situadas nas zonas nobres da cidade, próximas de atrações turísticas e de zonas de lazer, muito procuradas nos fins de semana e feriados.

¹⁷Teste de Jarque-Bera (p-value <0.05)

¹⁸Multicollinearity condition number >30

6.2.2 OLS exploratório - Períodos de ponta da manhã e da tarde

Da mesma forma que para os dias úteis, fins de semana e feriados, foi realizada uma análise de regressão linear exploratória para os períodos de ponta da manhã e da tarde. Os resultados são apresentados na Tabela 2:

Tabela 2: Resultados da regressão linear exploratória para os períodos de ponta da manhã e da tarde

Variável	Manhã		Tarde	
	Coefficiente	testes-t	Coefficiente	testes-t
Constante	-574,424	-10,7992	-458,347	-8,36933
Número de docas	11,9531	10,625	12,161	11,3842
Densidade de escritórios	0,000132368	8,06922	-	-
Densidade da população	0,00438645	5,9395	-	-
Densidade de paragens de transporte público	0,285928	4,85715	0,401165	7,27894
Percentagem de adultos (18-64 anos)	4,97478	6,61534	4,62875	6,65493
Densidade de estacionamento para bicicletas	-	-	0,916025	4,34119
Distância ao centro	-	-	-0,0148071	-4,45957
R²	0,702221		0,731232	
R² Ajustado	0,697790		0,727233	
Log Likelihood	-1995,94		-1973,17	
Akaike info criterion (AIC)	4003,87		3958,34	
Schwarz criterion	4026,88		3981,35	

Como no caso da análise realizada para os dias úteis e fins-de-semana, os testes-t não são confiáveis e existe a possibilidade de multicolinearidade entre as variáveis.

É interessante salientar que o parâmetro *densidade de escritórios* (*dens_offic*) embora seja próximo de zero, deverá ser significativo, dada a unidade de medição do mesmo em (m^2/km^2). Como se pode observar na Tabela 4, este indicador apresenta uma amplitude importante, havendo diferenças de mais de 1.000.000 (m^2/km^2) entre os bairros administrativos situados próximos do centro da cidade e os bairros residenciais da periferia. Note-se que 1.000.000 m^2/km^2 de diferença na densidade de escritórios corresponde a uma procura de 132 bicicletas no período de ponta da manhã.

O mesmo fenómeno, ainda que com menor amplitude que o anterior, é observado no parâmetro *densidade da população* (*dens_pop*), uma vez que existem diferenças notáveis entre os polígonos da região.

6.3 Seleção das variáveis explicativas

As combinações das variáveis explicativas resultantes da análise de regressão anterior, apesar de serem aquelas que permitem o melhor ajustamento do modelo, em termos de R^2 ajustado e AIC, são potencialmente redundantes entre si. Neste contexto, levou-se a cabo uma análise OLS com variáveis explicativas alternativas, tendo como objetivo encontrar um conjunto de variáveis significativas, para cada um dos períodos analisados e *multicollinearity condition number* inferior a 30. Nos modelos de regressão linear alternativos foram consideradas as variáveis seguintes:

- Número de docas (bst);
- Densidade de escritórios(dens_offic);
- Distância ao centro (dtcc);
- Densidade de paragens de transporte público (dens_ptsto);
- Densidade de estacionamentos para bicicletas (dens_bpark);

O *número de docas* (bst), podendo ser uma variável endógena, uma vez que é provavelmente fruto de um processo de dimensionamento que tem em conta a procura [Noland et al., 2016], é significativa em todos os períodos analisados. O mesmo acontece com a variável *densidade de paragens de transporte público* (dens_ptsto). [DeMaio, 2009] dá como exemplo o aumento do transporte público como um dos impactos positivos da implementação de sistemas de bicicletas partilhadas, deixando implícita a existência de complementaridade entre ambos. Neste particular, [Levy et al., 2017], [Tran et al., 2015] e [Noland et al., 2016] incluem variáveis relacionadas com o transporte público nos seus modelos.

Estes mesmos autores incluem variáveis relacionadas com a infraestrutura ciclável e com o emprego, às quais correspondem as variáveis *densidade de estacionamentos para bicicletas* (dens_bpark) e *densidade de escritórios* (dens_offic).

Por sua vez, a distância ao centro parece ser uma variável importante no que diz respeito à atividade das estações, como se pode verificar pela análise de *clusters* realizada anteriormente. Através desta análise é possível observar a existência de *clusters* de valores elevados localizados no centro da cidade e de *clusters* de valores baixos na periferia.

A Tabela 3 apresenta uma comparação entre os indicadores de ajustamento do modelo de regressão linear exploratória e do modelo de regressão linear alternativo para cada um dos períodos analisados. Como referido no capítulo anterior, o modelo melhor ajustado é aquele que apresenta menor valor de AIC e maior valor de R^2 ajustado.

Tabela 3: Comparação entre o modelo OLS alternativo e o modelo de regressão linear exploratório OLS*

Modelo	R²	R² aj.	LL	AIC
Dias úteis				
OLS (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	0,735727	0,731794	-2450,32	4912,64
OLS* (bst; dtcc; pc_1864; dens_ptsto; dens_bpark)	0,768306	0,764858	-2427,82	4867,64
Fins de semana e feriados				
OLS (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	0,696940	0,692430	-2142,01	4296,02
OLS* (bst; pc_1864; dens_pop; dens_ptsto; pc_bpaves)	0,728014	0,723966	-2123,51	4259,03
Período de ponta da manhã				
OLS (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	0,683734	0,679027	-2006,24	4024,47
OLS* (bst; pc_1864; dens_pop; dens_ptsto; dens_offic)	0,728014	0,723966	-2123,51	4259,03
Período de ponta da tarde				
OLS (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	0,708134	0,703790	-1987,27	3986,54
OLS* (bst; pc_1864; dens_pop; dens_ptsto; dens_bpark)	0,731232	0,727233	-1973,17	3958,34

Na tabela 4 apresentam-se as estatísticas descritivas de cada uma das variáveis dependentes e explicativas.

Note-se que, no que diz respeito às variáveis dependentes, a comparação entre períodos distintos, só é possível entre os períodos de ponta, uma vez que o número de viagens é mais elevado nos períodos de análise maiores.

Tabela 4: Estatísticas descritivas das variáveis em análise

	Mínimo	Média	Máximo	Desvio padrão
Variáveis dependentes				
Nº de viagens (dias úteis)	26	597,04	3681	608,58
Nº de viagens (ponta da manhã)	1	138,19	1030	151,84
Nº de viagens (ponta da tarde)	4	141,46	966	149,53
Nº de viagens (fins de semana e feriados)	9	228,47	1843	230,71
Variáveis explicativas - Villo				
Nº de docas Villo	14	25	44	4,56
Variáveis explicativas - Sócio-económicas				
Densidade da população (hab/km ²)	40,51	11.373,25	31.280,60	6.252,95
Densidade de escritórios (m ² /km ²)	625,26	180.492,21	2.589.272,57	358.561,675
% da população entre 18 e 64 anos	34,24	65,32	83,95	7,37
Variáveis explicativas - Mobilidade e infraestrutura				
Paragens de transporte público / km ²	1,13	79,34	1026,91	95,22
% de área de calçada	0	1,39	23,45	3,38
Estacionamentos de bicicletas / km ²	0,82	23,72	193,78	25,83
Distância ao centro (m)	181,51	3535,82	7984,85	1693,51

6.4 Resultados da modelação com distribuição de Poisson

Nesta secção analisam-se os resultados da modelação com distribuição de Poisson para os diferentes períodos de análise separadamente. As combinações das variáveis explicativas que decorrem da análise de regressão linear são modeladas através de um modelo global, um modelo GWRP local, e dois modelos semi-paramétricos (SGWRP):

- o primeiro onde apenas a variável *distância ao centro* (dtcc) é considerada variável global e todas as outras são consideradas variáveis locais;
- o segundo com duas variáveis globais, a *distância ao centro* (dtcc) e o *número de docas* (bst);

Antes de analisar os resultados do modelo melhor ajustado, que corresponde a um modelo GWRP nos períodos de ponta e a um modelo SGWRP nos dias úteis e fins de semana, é interessante verificar os resultados do modelo global com distribuição de Poisson nos diferentes períodos. O modelo global analisado é o seguinte: Poisson G. (bst; dtcc; dens_ptsto; dens_bpark; dens_offic).

Os resultados do modelo global são apresentados na Tabela 5, estandardizados de modo a comparar a contribuição de cada uma das variáveis explicativas no resultado da variável dependente. Os valores de z

ou *pseudo-t values* indicam que todas as variáveis são significativas ao nível de 99 %.

Tal como na modelação OLS exploratória, o sinal das variáveis está de acordo com o esperado. A *distância ao centro* (dtcc) comporta-se em todos os períodos analisados da mesma forma, ou seja, quanto maior é a distância ao centro, menor é a procura de viagens em bicicleta partilhada. Esta é também a variável com maior poder preditivo em todos os períodos analisados.

A variável *densidade de escritórios* (dens_offic) apresenta sinal positivo durante a semana, em parte devido aos movimentos pendulares casa-trabalho. No entanto, durante o fim de semana, esta mesma variável apresenta sinal negativo, dada a pouca atratividade destas zonas em contraste com os dias úteis. O peso desta variável é particularmente importante durante o período de ponta da manhã, talvez devido ao fato de ser durante este período do dia que há uma maior e mais concentrada procura de movimentos pendulares casa-trabalho. À tarde estes movimentos são menos concentrados. Se for analisado um período de 24 horas esta é a variável com menor poder explicativo.

Excluindo a análise dos dias úteis, a variável *densidade de paragens de transporte público* (dens_ptsto) é a que contribui menos para o resultado da variável dependente, parecendo sugerir que esta ou não é uma variável significativa ou o seu poder explicativo varia no espaço, ou seja é não-estacionário espacialmente.

O modelo global, embora permita obter uma imagem das relações entre as variáveis explicativas e as variáveis dependentes, não é apropriado para modelar os fenómenos uma vez que não existe estacionariedade espacial. Note-se a este propósito, que não só os indicadores de ajuste do modelo AICc e *Deviance* são mais baixos nos casos dos modelos locais GWRP e SGWRP, como o valor do parâmetro *constante* é particularmente elevado, indicando que uma boa parte do valor da variável dependente não é explicada pelas variáveis explicativas.

Tabela 5: Resultados do modelo de regressão de Poisson global

Variável	Coef.	Erro padrão (SE)	z (Coef./SE)	exp(Coef.)
Dias úteis				
Constante	6.1307	0.0027	2234.6086	459.7561
Número de docas	0.2046	0.0019	105.7319	1.2270
Dens. de escritórios	0.0392	0.0018	22.0937	1.0400
Dens. de estacionamentos para bicicletas	0.0874	0.0020	42.9450	1.0914
Dens. de paragens de transporte público	0.0405	0.0016	25.8549	1.0414
Distância ao centro	-0.4871	0.0034	-142.9702	0.6144
Fins de semana e feriados				
Constante	5.1853	0.0044	1182.0577	178.6273
Número de docas	0.1796	0.0032	55.6456	1.1967
Dens. de escritórios	-0.0756	0.0035	-21.7516	0.9272
Dens. de estacionamentos para bicicletas	0.1069	0.0032	33.1681	1.1129
Dens. de paragens de transporte público	0.0700	0.0025	27.7256	1.0725
Distância ao centro	-0.5207	0.0056	-92.8838	0.5941
Período de ponta da manhã				
Constante	4.6550	0.0058	808.5998	105.1102
Número de docas	0.2032	0.0040	51.3545	1.2254
Dens. de escritórios	0.1196	0.0032	36.5506	1.1271
Dens. de estacionamentos para bicicletas	0.0598	0.0044	13.6694	1.0616
Dens. de paragens de transporte público	0.0194	0.0033	5.8211	1.0195
Distância ao centro	-0.4784	0.0070	-68.1242	0.6198
Período de ponta da tarde				
Constante	4.6979	0.0056	840.7181	109.7200
Número de docas	0.2174	0.0040	54.9960	1.2429
Dens. de escritórios	0.0574	0.0035	16.2166	1.0591
Dens. de estacionamentos para bicicletas	0.0820	0.0042	19.4147	1.0855
Dens. de paragens de transporte público	0.0460	0.0032	14.4795	1.0471
Distância ao centro	-0.4457	0.0069	-64.1863	0.6404

Na Tabela 6 são apresentados os resultados de cada um dos modelos analisados, estando representado a azul o modelo que melhor se ajusta para cada um dos períodos analisados. As variáveis explicativas são apresentadas segundo a abreviatura. Os modelos analisados são o modelo de regressão com distribuição de Poisson global (Poisson G.), o modelo de regressão geograficamente ponderada com distribuição de

Poisson (GWRP), onde todas as variáveis explicativas são consideradas locais, e o modelo de regressão geograficamente ponderada semi-paramétrico com distribuição de Poisson (SGWRP). Neste último, as variáveis globais entre parêntesis são precedidas pela letra *G* e as variáveis locais, pela letra *L*.

É interessante constatar, que as combinações de variáveis resultantes da análise de regressão linear exploratória não são as que melhor ajustam o modelo. Para isso deverá contribuir as suspeitas de multicolinearidade já analisadas.

Pela análise da Tabela 6 pode-se constatar que tanto para o período de ponta da manhã como da tarde o modelo GWRP (bst; dtcc; dens_ptsto; dens_bpark; dens_offic) é aquele que apresenta melhor performance. Já no caso dos "Dias úteis" e "Fins de semana e feriado" o modelo semi-paramétrico SGWRP [G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)] é aquele que apresenta melhores resultados.

A hipótese de que a *distância ao centro* (dtcc) contribui negativamente para o número de viagens, independentemente da zona, ou seja, que esta é uma variável global, apenas se confirma para os períodos de análise maiores, dias úteis e fins de semana, onde o modelo semi-paramétrico (SGWRP) com quatro variáveis explicativas locais e a *distância ao centro* (dtcc) como variável global tem uma melhor performance que o GWRP.

Tabela 6: Comparação estatística entre os modelos

	NN ^a	AICc ^(*)	Dev. ^b	DOF ^c
Dias úteis				
Poisson G. (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	-	52145,99	52133,73	336
GWRP (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	12	4720,16	2594,90	52,94
SGWRP G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)	10	4684,68	2041,77	43,90
SGWRP G(dtcc;bst); L(dens_ptsto; dens_bpark; dens_offic)	10	5546,87	4093,00	78,47
Período de ponta da manhã				
Poisson G. (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	-	16258,70	16246,45	336
GWRP (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	15	2639,18	1413,47	88,70
SGWRP G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)	15	2875,52	1884,47	123,28
SGWRP G(dtcc;bst); L(dens_ptsto; dens_bpark; dens_offic)	15	3362,12	2724,58	143,59
Poisson G. (bst; pc_1864; dens_pop; dens_ptsto; dens_offic)	-	14613,32	14601,07	336
GWRP (bst; pc_1864; dens_pop; dens_ptsto; dens_offic)	15	2959,67	1726,59	19,64
Período de ponta da tarde				
Poisson G. (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	-	14024,9	14012,24	336
GWRP (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	15	2581,483	1361,84	88,01
SGWRP G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)	13	2683,28	1355,83	82,59
SGWRP G(dtcc; bst); L(dens_ptsto; dens_bpark; dens_offic)	11	2602,82	1437,51	93,59
Poisson G. (bst; pc_1864; dens_pop; dens_ptsto; dens_bpark)	-	12336,26	12324,01	336
GWRP (bst; pc_1864; dens_pop; dens_ptsto; dens_bpark)	14	2798,13	1350,78	76,39
Fins de semana e feriados				
Poisson G. (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	-	20452,69	20440,44	336
GWRP (bst; dtcc; dens_ptsto; dens_bpark; dens_offic)	14	3293,53	1859,61	76,19
SGWRP G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)	12	3219,14	2110,38	70,66
SGWRP G(dtcc; bst); L(dens_ptsto; pc_bpark; dens_offic)	9	3489,21	1576,14	62,53
Poisson G. (bst; pc_1864; dens_pop; dens_ptsto; pc_bpaves)	-	17115,05	17102,80	336
GWRP (bst; pc_1864; dens_pop; dens_ptsto; pc_bpaves)	22	4909,33	4293,31	146,17

(*) O valor mais baixo de AICc, que corresponde ao modelo considerado melhor ajustado, é representado a azul;

(^a) Número de vizinhos;

(^b) Deviance;

(^c) Graus de liberdade

6.5 Interpretação das variações geográficas dos parâmetros

Como referido no capítulo anterior e de acordo com os objetivos fixados, interessa sobretudo perceber de que forma os fatores que afetam as viagens em bicicleta variam no tempo. Por este motivo, nesta secção analisam-se apenas as variações geográficas dos diferentes parâmetros do modelo GWRP (bst; dtcc; dens_ptsto; dens_bpark; dens_offic) para os períodos da manhã e da tarde.

Os resultados referentes aos períodos *Dias úteis*, assim como *Fins de semana e feriados* são apresentados em anexo, uma vez que dizem respeito a períodos de 24 horas, razão pela qual eventuais variações ao longo do dia não são observáveis.

Para as diferentes variáveis explicativas, apresenta-se sistematicamente à esquerda o valor do coeficiente e à direita o *pseudo-t value*, que permite ter uma melhor perceção do sinal das variáveis e onde é que estas são ou não significativas. As variáveis não significativas ao nível de 90 % são as que correspondem a um *pseudo-t value* entre -1.96 e 1.96.

A complementaridade entre o transporte público e os sistemas de bicicletas partilhadas é amplamente descrita na literatura (ver [Fishman et al., 2013]). Esperar-se-ia portanto obter valores positivos elevados nas zonas onde existe maior densidade de paragens de transporte público.

Como se pode observar na Figura 10, se a existência de uma relação positiva entre a procura nas estações Villo e a densidade de transportes públicos é uma realidade na maior parte do território, também existem pequenos núcleos onde a relação é de sinal contrário, o que parece contra-intuitivo. Uma possível explicação para este facto é que em certas zonas do território a bicicleta partilhada é complementar com o transporte público, enquanto noutras zonas a bicicleta constitui uma verdadeira alternativa a este. Este fenómeno é igualmente descrito por [Levy et al., 2017].

Numa vasta área da periferia a sul, parece não haver uma relação significativa entre esta variável explicativa e a variável dependente (*pseudo-t values* entre -1.96 e 1.96).

Optou-se por incluir a localização das estações de comboio e de metro nos mapas dos coeficientes, uma vez que é com estes meios de transporte que se esperaria que a complementaridade bicicleta partilhada e transporte público fosse maior. No entanto, por simples observação dos mapas não é possível tirar qualquer conclusão a esse respeito. Note-se a este respeito, que a variável *distância em linha reta entre as estações Villo e a estação de metro ou comboio mais próxima* foi testada na análise de regressão linear exploratória, não se tendo revelado significativa.

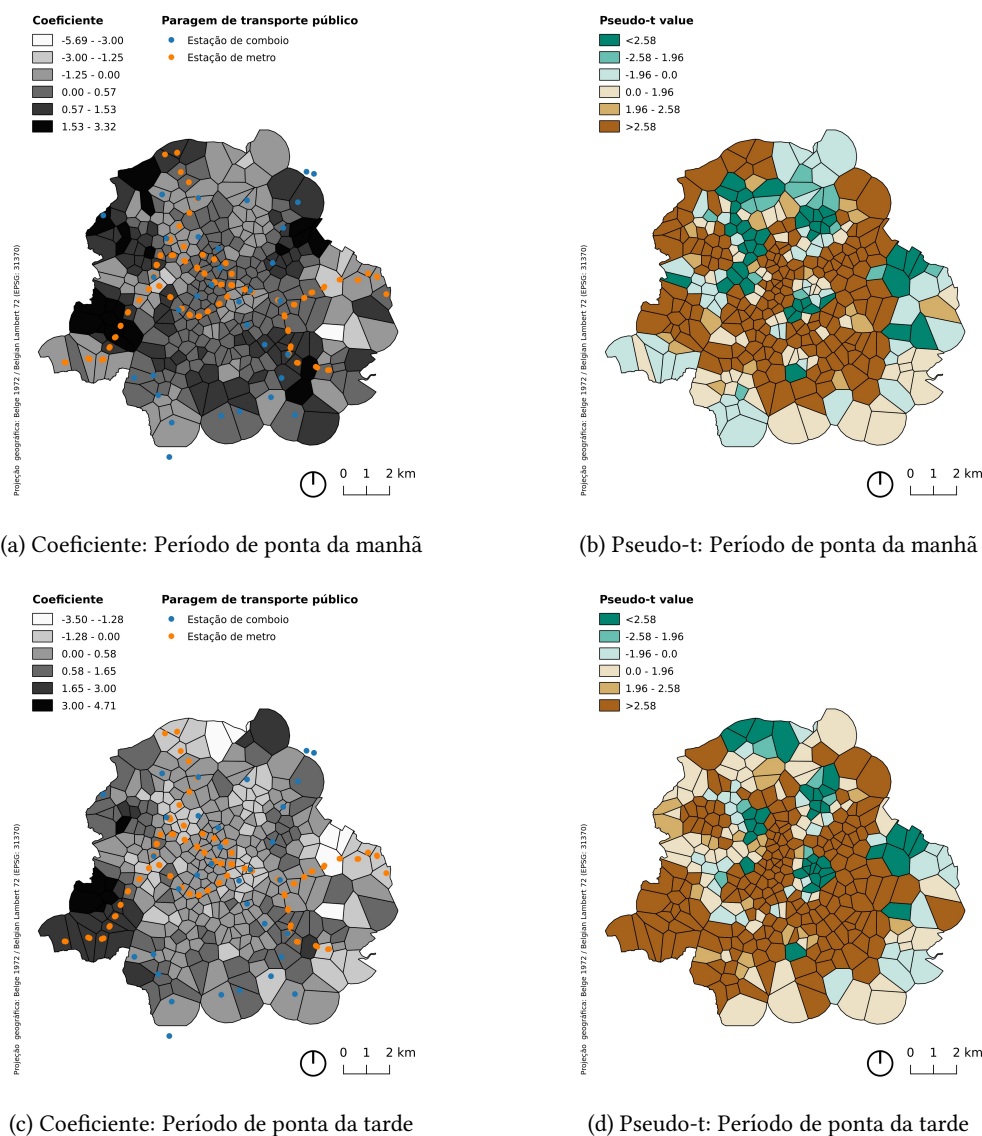


Figura 10: Resultados GWRP - Variável *Densidade de paragens de transporte público*

A Figura 11 apresenta a variação geográfica dos parâmetros da variável *densidade de escritórios* (*dens_offic*). No caso desta variável, observa-se uma grande variação geográfica dos parâmetros.

Da análise dos *pseudo-t values* verifica-se que há mais zonas onde o parâmetro não é significativo durante o período de ponta da tarde.

As zonas com maior densidade de escritórios, não são necessariamente aquelas em que o parâmetro é positivo e elevado, como se esperaria para o período de ponta da manhã, dados os resultados do modelo global. Ao contrário, nessas zonas, parâmetros negativos intercalam com parâmetros positivos, sendo em alguns casos significativos e noutros não. Uma possível explicação para este fenómeno é o facto de algumas estações ficarem cheias rapidamente, uma vez que os movimentos pendulares ocorrem maioritariamente no sentido casa-trabalho durante o período de ponta da manhã e no sentido inverso durante a tarde. Não

havendo redistribuição das bicicletas pelo operador estações próximas com capacidade disponível são procuradas.

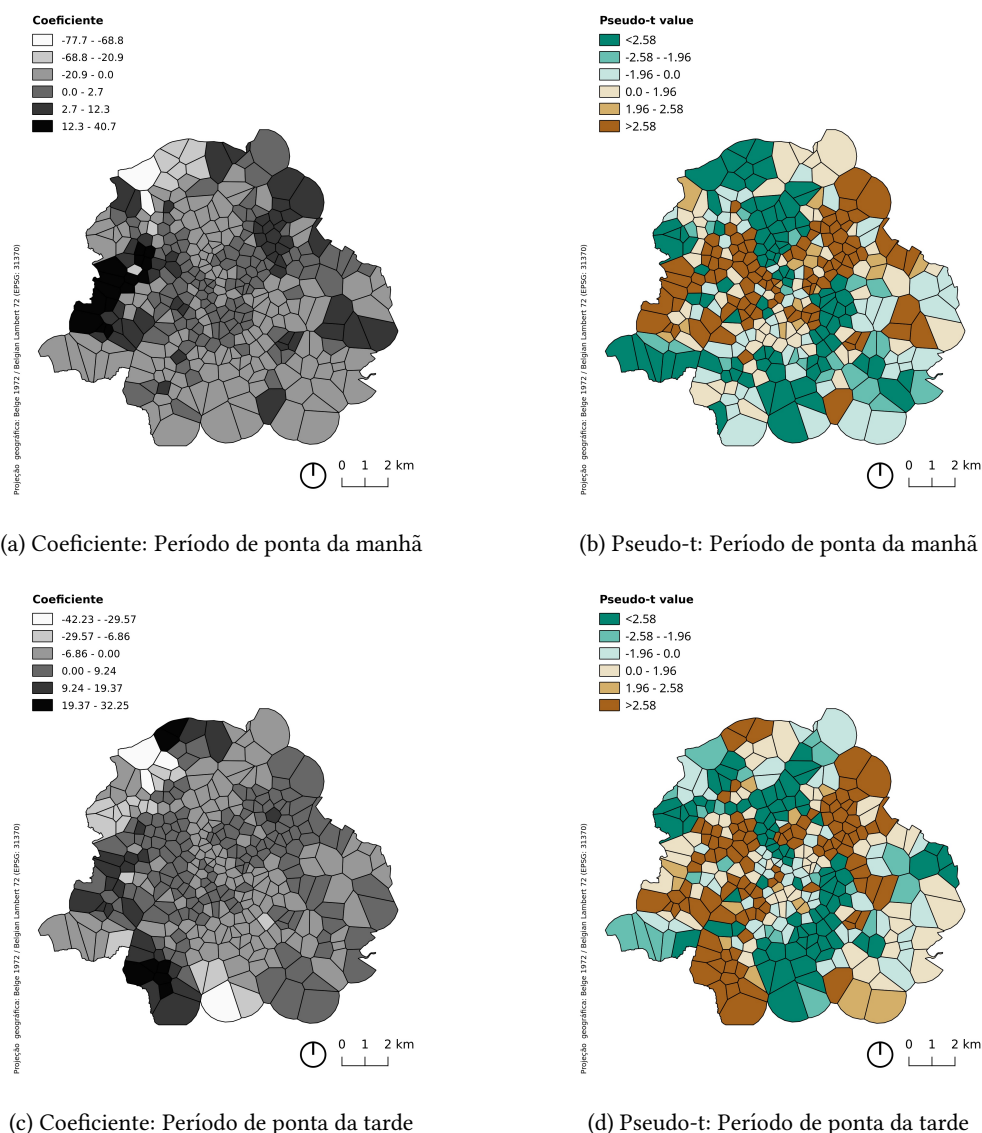


Figura 11: Resultados GWRP - Variável *Densidade de escritórios*

A existência de estacionamentos para bicicletas embora não influencie a procura dos utilizadores do sistema de bicicletas partilhadas, constitui indiretamente um indicador das zonas onde existe maior procura de ciclistas.

Da análise da Figura 12 apenas se pode concluir que a hipótese formulada no parágrafo anterior é apenas verdadeira numa parte do território.

Da análise dos *pseudo-t values* pode-se constatar que existe uma vasta zona onde o parâmetro é negativo tanto durante o período de ponta da manhã como da tarde. Em três núcleos a Oeste, Sudeste e Nordeste o *pseudo t-value* é em paralelo inferior a 2.58 e o coeficiente tem um valor negativo elevado em ambos

os períodos. Estas zonas têm em comum o fato de se localizarem na periferia da região, relativamente afastadas do centro da cidade. Nos núcleos a Oeste e a Sudeste é de salientar que existem dois Campus universitários, sendo que a zona a Nordeste coincide com a sede da NATO.

Conclui-se desta análise que as motivações dos ciclistas não são necessariamente as mesmas dos usuários de Villo, dependendo muito da localização geográfica. Uma hipótese que poderá explicar este fenómeno é o facto de em algumas zonas do território haver mais preponderância para a utilização da bicicleta particular, ou seja, a proporção entre o número de utilizadores de Villo e aqueles que utilizam bicicleta própria varia de zona para zona.

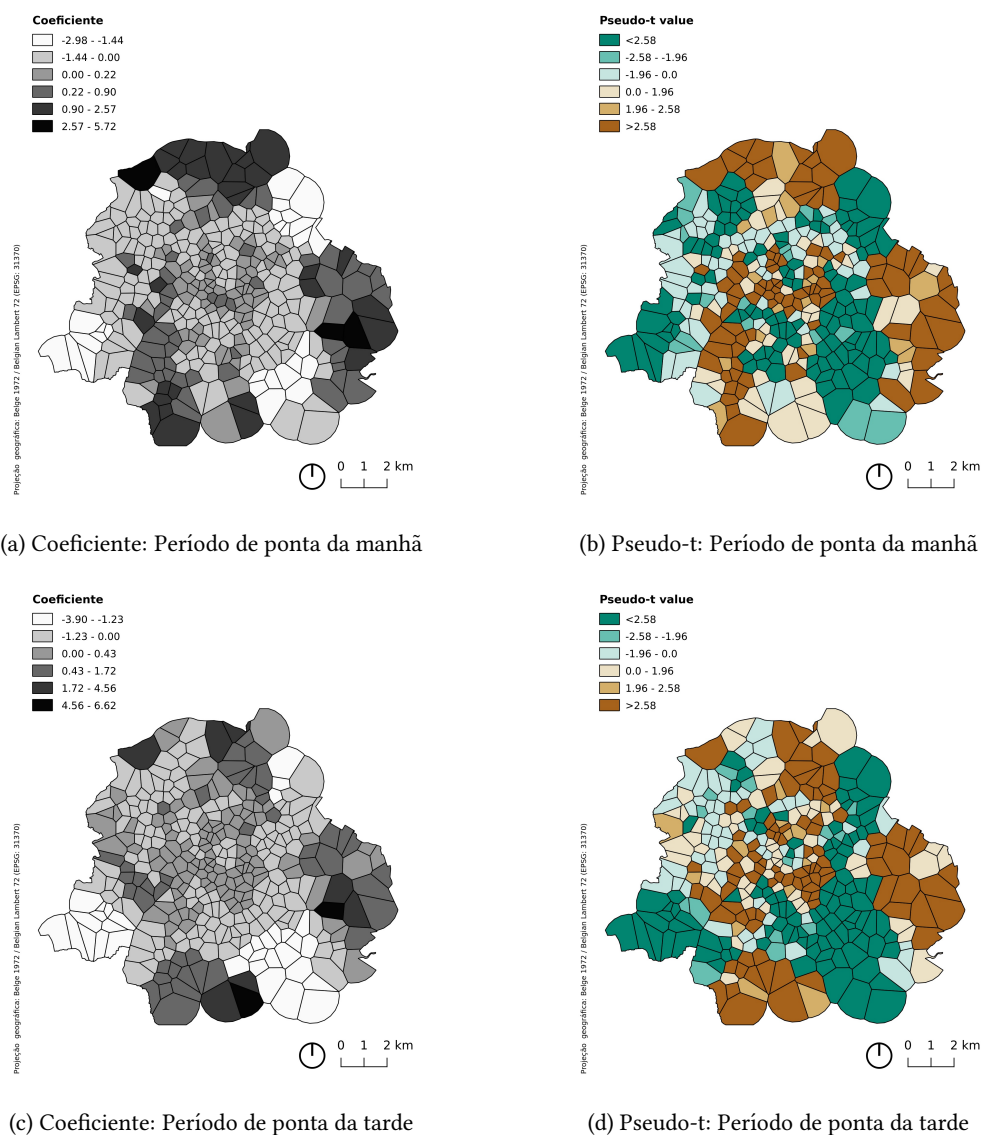


Figura 12: Resultados GWRP - Variável *Densidade de estacionamentos para bicicletas*

No que diz respeito à variável *distância ao centro* (dtcc), a Figura 13 mostra que, em geral, quanto mais longe uma zona está do centro, menor é a procura. Partindo deste pressuposto, esta variável foi testada

num modelo semi-paramétrico enquanto variável global. O facto de para os períodos de ponta o modelo melhor ajustado corresponda ao modelo GWRP, é provavelmente devido à existência de algumas exceções em que o coeficiente é positivo.

É de realçar que o centro corresponde, no contexto deste trabalho, à Grand'Place, a principal atração turística da cidade, na proximidade da qual estão implementados inúmeros comércios, serviços e atrações noturnas. Contudo, a região possui uma grande heterogeneidade, havendo várias outras zonas centrais geradoras de viagens.

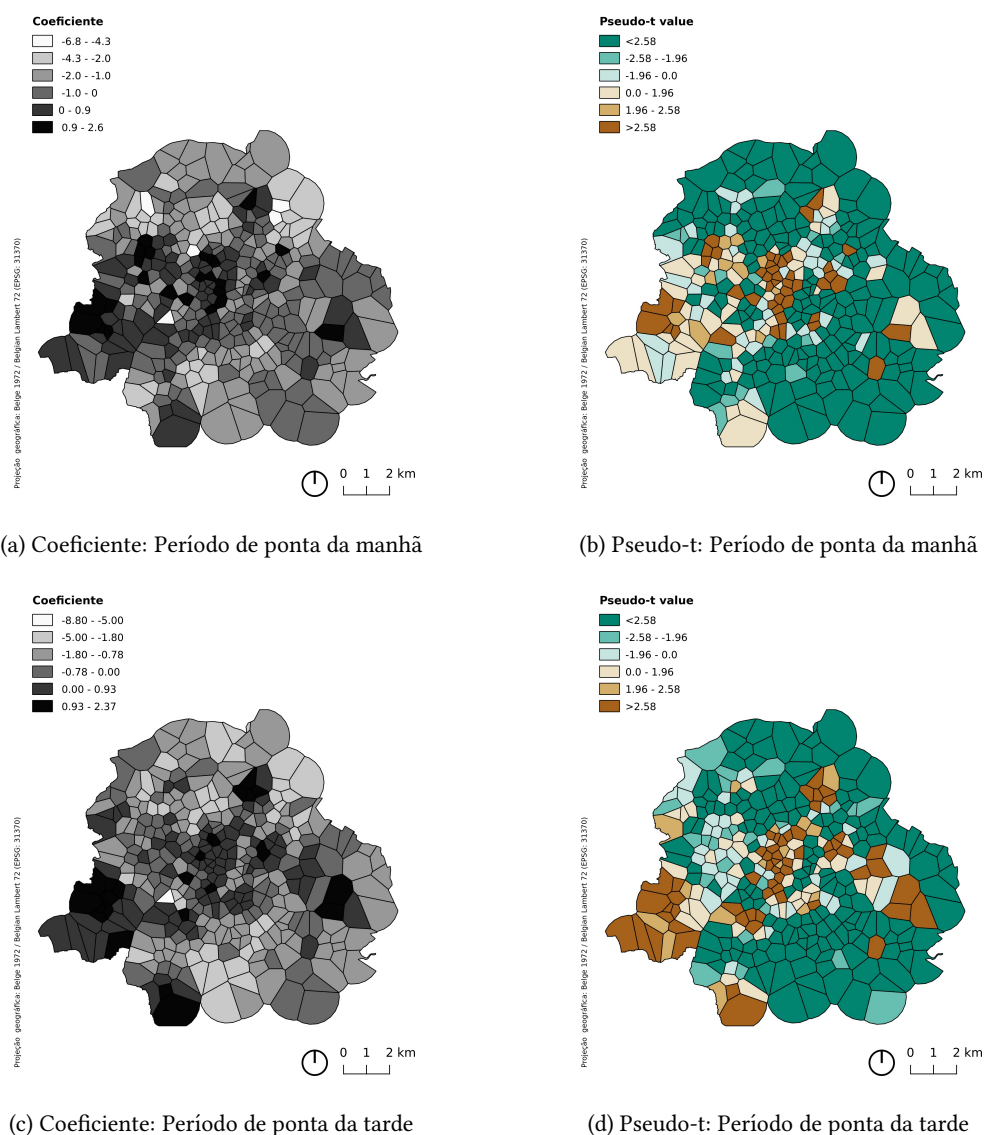


Figura 13: Resultados GWRP - Variável *Distância ao centro*

O número de docas Viljo parece ser um fator diretamente relacionado com a procura. Como se pode observar na Figura 14, este coeficiente é positivo e significativo na maioria do território. Além disso, este fator parece ser mais importante na periferia a sul e a norte, provavelmente resultado da distância

entre estações ser geralmente maior. Porém, também existem zonas periféricas onde o número de docas influencia negativamente a procura.

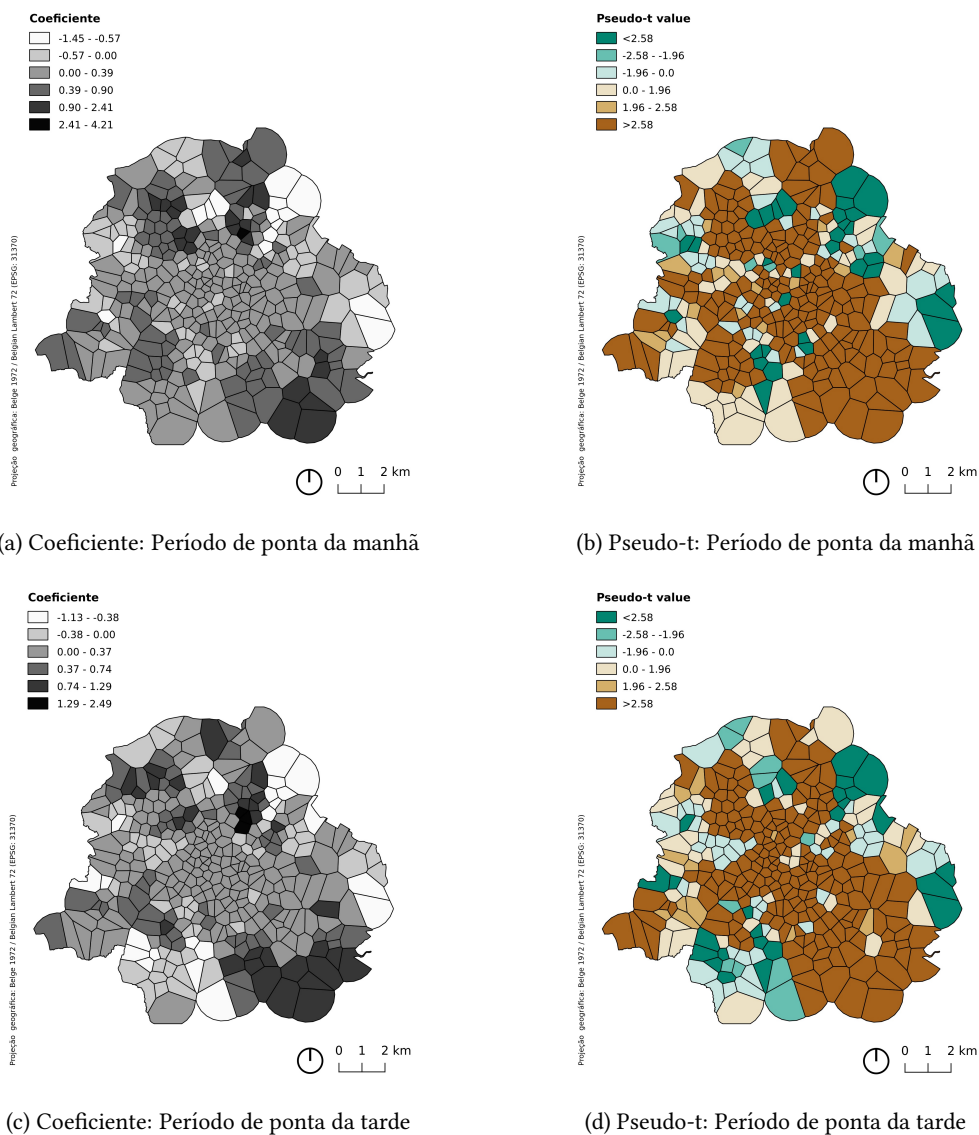


Figura 14: Resultados GWRP - Variável *Nº de docas Vilvoorde*

7 Conclusão

Na elaboração deste trabalho usaram-se diferentes técnicas para a modelação da procura de viagens nas diferentes estações do sistema público de bicicletas partilhadas de Bruxelas.

Os modelos foram desenvolvidos para quatro períodos distintos de forma a entender de que forma o período do dia afeta os diferentes parâmetros analisados, tendo-se dado especial atenção aos períodos de ponta por serem aqueles que, segundo a hipótese colocada inicialmente, são fundamentalmente usados com um propósito outro que turismo ou lazer.

Os resultados permitem concluir que de todas as variáveis analisadas, aquelas que melhor explicam a variável dependente em cada um dos períodos de análise são a *densidade de transportes públicos*, a *distância ao centro*, a *densidade de estacionamentos para bicicletas*, a *densidade de escritórios* e o *n° de docas Villo*.

Se o modelo local GWRP é aquele que melhor se ajusta nos períodos de ponta, um modelo semi-paramétrico (SGWRP), com a *distância ao centro* enquanto variável global e todas as outras como variáveis locais apresenta melhor performance para os períodos referentes aos "dias úteis" e "fins de semana e feriados".

Se a análise exploratória e a análise dos modelos globais permitem compreender globalmente a relação entre as variáveis independentes e dependentes, os modelos locais utilizados permitem descobrir relações menos evidentes entre estas, sem a utilização dos quais permaneceriam ocultas. O GWRP pode assim ser visto como um método de análise exploratória, que contextualizado com outros dados e informações permite apreender os fenómenos locais.

Contudo os resultados são por vezes contra-intuitivos e de difícil interpretação e não é de excluir que haja problemas de especificação do modelo. Como sugere [Fotheringham et al., 2002], a contextualização dos parâmetros só deve ser realizada uma vez resolvidos todos os possíveis problemas de especificação do modelo.

É de salientar que a disponibilidade de dados para a área geográfica pretendida constitui um fator determinante para a elaboração dos modelos, pelo que a não integração de alguma variável não identificada como pertinente ou não disponível poderá resultar num modelo mal especificado. Futuros estudos, deverão procurar integrar dados relacionados com o consumo, com o turismo e eventualmente outros dados da atividade económica e do uso do solo.

Uma alternativa possível ao modelo apresentado, seria a realização de um modelo de Regressão Binomial Negativa que permitiria corrigir eventuais problemas de sobre-dispersão.

Uma outra possível extensão deste estudo, seria a modelação de pares origem-destino, tal como no estudo levado a cabo por [Levy et al., 2017], que permitiria, em certa medida, diferenciar padrões de utilização e conhecer a relação entre as estações. Uma análise deste tipo em paralelo com a elaboração de um modelo de afetação de tráfego contribuiria para um conhecimento fino da mobilidade suave na região de Bruxelas,

que poderia ser calibrada em certas zonas a partir de dados de contagens.

A utilização de um modelo de regressão espacial enquanto modelo preditivo é outra possível extensão deste estudo. Para a realização de um tal modelo, ter-se-iam que integrar variáveis meteorológicas e outras, que permitissem a estimação para outros períodos do ano, dada a grande variação sazonal do número de ciclistas. Por outro lado, o número de ciclistas é em constante evolução, pelo que um modelo preditivo, teria que ter este aspeto em consideração.

Por último, e de acordo com os objetivos fixados de entender os limites da metodologia utilizada na generalização do modelo a outros contextos geográficos e às viagens em bicicleta como um todo, conclui-se que ambas generalizações poderiam ser abusivas.

Veja-se a título de exemplo os resultados, apresentados na secção 6.5, da variável *densidade de estacionamentos para bicicletas*. A grande variação espacial do parâmetro e o facto de este apresentar valores positivos em determinadas zonas e negativos noutras sugere que as motivações dos ciclistas não são necessariamente as mesmas dos usuários do sistema de bicicletas partilhadas.

Tal como sugere [Li et al., 2013], o modelo GWRP não é diretamente aplicável a outros contextos geográficos. No âmbito deste estudo, as variáveis locais do modelo foram calibradas especificamente para a região de Bruxelas-Capital e obedecem portanto às especificidades locais desta.

Bibliografia

- [Anselin, 1995] Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2):93–115.
- [Anselin, 2005] Anselin, L. (2005). *Exploring Spatial Data with GeoDa: A Workbook*. Spatial Analysis Laboratory, Department of Geography - University of Illinois, Urbana-Champaign.
- [Anselin, 2010] Anselin, L. (2010). Ordinary Least Squares and Poisson Regression Models. In *Crimestat Documentation*.
- [asbl ProVelo,] asbl ProVelo. Observatoire du vélo en Région de Bruxelles-Capitale 2017.
- [DeMaio, 2009] DeMaio, P. (2009). Bike-sharing: History, Impacts, Models of Provision, and Future. *Journal of Public Transportation*, 12(4):41–56.
- [Dereli and Erdogan, 2017] Dereli, M. A. and Erdogan, S. (2017). A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods. *Transportation Research Part A: Policy and Practice*, 103:106–117.
- [Faghih-Imani et al., 2017] Faghih-Imani, A., Hampshire, R., Marla, L., and Eluru, N. (2017). An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville. *Transportation Research Part A: Policy and Practice*, 97:177–191.
- [Feng et al., 2017] Feng, Y., Affonso, R. C., and Zolghadri, M. (2017). Analysis of bike sharing system by clustering: the Vélib' case. *IFAC-PapersOnLine*, 50(1):12422–12427.
- [Fishman et al., 2013] Fishman, E., Washington, S., and Haworth, N. (2013). Bike Share: A Synthesis of the Literature. *Transport Reviews*, 33(2):148–165.
- [Fotheringham et al., 2002] Fotheringham, A. S., Brunson, C., and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons, Chichester, 1st edition edition.
- [García-Palomares et al., 2012] García-Palomares, J. C., Gutiérrez, J., and Latorre, M. (2012). Optimizing the location of stations in bike-sharing programs: A GIS approach. *Applied Geography*, 35(1):235–246.
- [Glaser, 2017] Glaser, S. (2017). A review of spatial econometric models for count data. Working Paper 19-2017, Hohenheim Discussion Papers in Business, Economics and Social Sciences.
- [Griffin and Jiao, 2015] Griffin, G. P. and Jiao, J. (2015). Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. *Journal of Transport & Health*, 2(2):238–247.

- [Hadayeghi et al., 2010] Hadayeghi, A., Shalaby, A. S., and Persaud, B. N. (2010). Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. *Accident Analysis & Prevention*, 42(2):676–688.
- [Huang et al., 2018] Huang, Y., Wang, X., and Patton, D. (2018). Examining spatial relationships between crashes and the built environment: A geographically weighted regression approach. *Journal of Transport Geography*, 69:221–233.
- [Kaltenbrunner et al., 2010] Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., and Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, 6(4):455–466.
- [Kesteloot et al., 2017] Kesteloot, L., Verstraeten, F., and Humbert, E. (2017). Comptages et analyse des données. Technical Report Rapport 2017, Provelo, Bruxelles.
- [Lathia et al., 2012] Lathia, N., Ahmed, S., and Capra, L. (2012). Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research Part C: Emerging Technologies*, 22:88–102.
- [Levine et al., 2010] Levine, N., Lord, D., and Park, B.-J. (2010). CrimeStat Version 3.3 Update Notes: Part II: Regression. In *CrimeStat Documentation*.
- [Levy et al., 2017] Levy, N., Golani, C., and Ben-Elia, E. (2017). An exploratory study of spatial patterns of cycling in Tel Aviv using passively generated bike-sharing data. *Journal of Transport Geography*.
- [Li et al., 2013] Li, Z., Wang, W., Liu, P., Bigham, J. M., and Ragland, D. R. (2013). Using Geographically Weighted Poisson Regression for county-level crash modeling in California. *Safety Science*, 58:89–97.
- [M. Fischer and Wang, 2011] M. Fischer, M. and Wang, J. (2011). Spatial Data Analysis Models, Methods and Techniques. In *Spatial Data Analysis Models, Methods and Techniques*, page 19. Springer, first edition.
- [Migdley, 2011] Migdley, P. (2011). Bicycle sharing schemes: Enhancing Sustainable Mobility in Urban Areas. *Commission on Sustainable Development, 19th Session, Background Paper No. 8 - CSD19/2011/BP 8*, pages 1–26.
- [Milne and Watling, 2018] Milne, D. and Watling, D. (2018). Big data and understanding change in the context of planning transport systems. *Journal of Transport Geography*.
- [Nakaya et al., 2005] Nakaya, T., Fotheringham, A. S., Brunson, C., and Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine*, 24(17):2695–2717.

- [Nakaya et al., 2009] Nakaya, T., Fotheringham, S., Charlton, M., and Brunsdon, C. (2009). Semiparametric geographically weighted generalised linear modelling in GWR 4.0. In Lees, editor, *10th International Conference on GeoComputation, UNSW, Sydney, November-December, 2009.*, pages 1–5, UNSW, Sydney.
- [Nakaya et al., 2016] Nakaya, T., Martin Charlton, Brunsdon, C., Paul Lewis, Jing Yao, and A Stewart Fotheringham (2016). *GWR4.09 User Manual*.
- [Noland et al., 2016] Noland, R. B., Smart, M. J., and Guo, Z. (2016). Bikeshare trip generation in New York City. *Transportation Research Part A: Policy and Practice*, 94:164–181.
- [Ortúzar and Willumsen, 2011] Ortúzar, J. d. D. and Willumsen, L. G. (2011). *Modelling Transport, 4th Edition*. John Wiley & Sons, UK, 4th edition edition.
- [Otero et al., 2018] Otero, I., Nieuwenhuijsen, M. J., and Rojas-Rueda, D. (2018). Health impacts of bike sharing systems in Europe. *Environment International*, 115:387–394.
- [O’Brien et al., 2014] O’Brien, O., Cheshire, J., and Batty, M. (2014). Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, 34:262–273.
- [Randriamanamihaga et al., 2014] Randriamanamihaga, A. N., Côme, E., Oukhellou, L., and Govaert, G. (2014). Clustering the Vélib dynamic Origin/Destination flows using a family of Poisson mixture models. *Neurocomputing*, 141:124–138.
- [Rhee et al., 2016] Rhee, K.-A., Kim, J.-K., Lee, Y.-i., and Ulfarsson, G. F. (2016). Spatial regression analysis of traffic crashes in Seoul. *Accident Analysis & Prevention*, 91:190–199.
- [Simões and Natário, 2016] Simões, P. and Natário, I. (2016). Spatial Econometric Approaches for Count Data: An Overview and New Directions. *International Journal of Economics and Management Engineering*, 10(1):10.
- [Thomas et al., 2013] Thomas, T., Jaarsma, R., and Tutert, B. (2013). Exploring temporal fluctuations of daily cycling demand on Dutch cycle paths: the influence of weather on cycling. *Transportation*, 40(1):1–22.
- [Tobler, 1970] Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46:234–240.
- [Tran et al., 2015] Tran, T. D., Ovtracht, N., and d’Arcier, B. F. (2015). Modeling Bike Sharing System using Built Environment Factors. *Procedia CIRP*, 30:293–298.
- [Vogel et al., 2011] Vogel, P., Greiser, T., and Mattfeld, D. C. (2011). Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns. *Procedia - Social and Behavioral Sciences*, 20:514–523.
- [Yang et al., 2017] Yang, H., Lu, X., Cherry, C., Liu, X., and Li, Y. (2017). Spatial variations in active mode trip volume at intersections: a local analysis utilizing geographically weighted regression. *Journal of Transport Geography*, 64:184–194.

- [Zhou et al., 2018] Zhou, Y., Wang, L., Zhong, R., and Tan, Y. (2018). A Markov Chain Based Demand Prediction Model for Stations in Bike Sharing Systems. *Mathematical Problems in Engineering*.
- [Zuur et al., 2010] Zuur, A. F., Ieno, E. N., and Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1):3–14.

Anexos

Tabela A.1: Tarifas em vigor em 2018

	Passe de 1 ano	Passe de 1 ano - cartão Mobib	Bilhete de 1 dia	Bilhete de 7 dias
Preço de base	34,70 €	33,70 €	1,60 €	8,20 €
Nº de viagens	Ilimitado	Ilimitado	Ilimitado	Ilimitado
0 - 30 min	Gratuito	Gratuito	Gratuito	Gratuito
30 min - 1 h	0,50 €	0,50 €	0,50 €	0,50 €
1 h - 1^{1/2} h	1 €	1 €	1 €	1 €
1^{1/2} h - 2 h	2 €	2 €	2 €	2 €
>2 h	2 € / 30 min	2 € / 30 min	2 € / 30 min	2 € / 30 min

Tabela A.2: Dados

id	Identificador Chave primária	-	-
uid	Código estação	-	Villo - JCDecaux
st_name	Nome da estação	-	Villo - JCDecaux
st_address	Morada da estação	-	Villo - JCDecaux
bst	Número de docas	Docas	Villo - JCDecaux
trvtot_ad	Total de viagens com origem ou destino nas estações durante os dias úteis (Obtido a partir da variação do número de bicicletas disponíveis nas docas)	Viagens	calculado de: Villo - JCDecaux
trvtot_mp	Total de viagens com origem ou destino nas estações no período de ponta da manhã (Obtido a partir da variação do número de bicicletas disponíveis nas docas)	Viagens	calculado de: Villo - JCDecaux

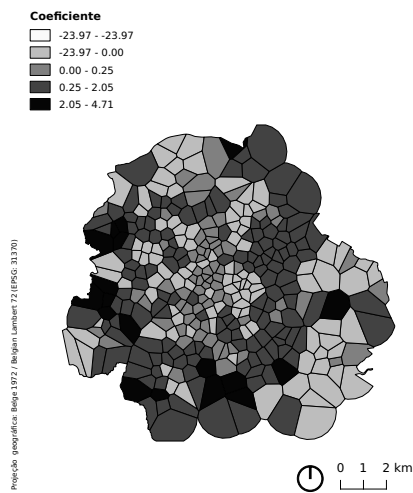
Dados	Descrição	Unidade	Fonte
trvtot_ep	Total de viagens com origem ou destino nas estações no período de ponta da tarde (Obtido a partir da variação do número de bicicletas disponíveis nas docas)	Viagens	calculado de: Villo - JCDecaux
trvout_ad	Total de viagens com origem nas estações - dias úteis (Obtido a partir da variação negativa do número de bicicletas disponíveis nas docas)	Viagens	calculado de: Villo - JCDecaux
trvout_mp	Total de viagens com origem nas estações - Período de ponta da manhã (Obtido a partir da variação negativa do número de bicicletas disponíveis nas docas)	Viagens	calculado de: Villo - JCDecaux
trvout_ep	Total de viagens com origem nas estações - Período de ponta da tarde (Obtido a partir da variação negativa do número de bicicletas disponíveis nas docas)	Viagens	calculado de: Villo - JCDecaux
trvout_we	Total de viagens com origem nas estações - Fins de semana e feriados (Obtido a partir da variação negativa do número de bicicletas disponíveis nas docas)	Viagens	calculado de: Villo - JCDecaux
trvin_ad	Total de viagens com destino nas estações - dias úteis (Obtido a partir da variação positiva do número de bicicletas disponíveis nas docas)	Viagens	calculado de: Villo - JCDecaux
trvin_mp	Total de viagens com destino nas estações - Período de ponta da manhã (Obtido a partir da variação positiva do número de bicicletas disponíveis nas docas)	Viagens	calculado de: Villo - JCDecaux

Dados	Descrição	Unidade	Fonte
trvin_ep	Total de viagens com destino nas estações - Período de ponta da tarde (Obtido a partir da variação positiva do número de bicicletas disponíveis nas docas)	Viagens	calculado de: Villo - JCDecaux
trvin_we	Total de viagens com destino nas estações - Fins de semana e feriados (Obtido a partir da variação positiva do número de bicicletas disponíveis nas docas)	Viagens	calculado de: Villo - JCDecaux
loadfct_ad	Ocupação das estações - dias úteis	%	Villo - JCDecaux
loadfct_mp	Ocupação das estações - período de ponta da manhã (dias úteis)	%	Calculado de: Villo - JCDecaux
loadfct_ep	Ocupação das estações - período de ponta da tarde (dias úteis)	%	calculado de: Villo - JCDecaux
loadfct_we	Ocupação das estações - fins de semana e feriados	%	calculado de: Villo - JCDecaux
tot_pop	População total em 2016	Habitantes	IBSA
dens_pop	Densidade da população em 2016	$\frac{Habitantes}{km^2}$	IBSA
pc_pop1864	Percentagem da população com idade compreendida entre os 18 e os 65 anos em 2016	%	IBSA
pc_pop65	Percentagem da população com mais de 65 anos em 2016	%	IBSA
pc_unemp	Taxa de procura de emprego em 2014 Rácio entre as pessoas à procura de emprego e as pessoas em idade de trabalhar (com idades compreendidas entre os 18 e os 65 anos)	%	IBSA
med_inc	Mediana dos rendimentos brutos por declaração fiscal em 2015	€	IBSA

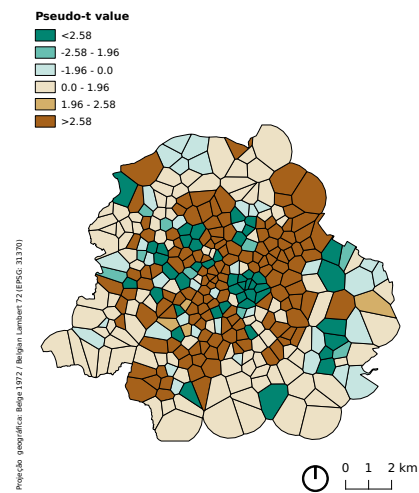
Dados	Descrição	Unidade	Fonte
pc_prim	Percentagem de alunos do ensino primário residentes no bairro ou nos bairros vizinhos inscritos numa escola primária do bairro, no ano lectivo de 2016-2017	%	IBSA
pc_sec	Percentagem de alunos do ensino secundário residentes no bairro ou nos bairros vizinhos inscritos numa escola primária do bairro, no ano letivo de 2016-2017	%	IBSA
dens_offic	Densidade de escritórios em 2016	$\frac{m^2}{km^2}$	IBSA
pc_nohab	Percentagem da área construída não reservada à habitação Este indicador permite apreender a função do bairro distinguindo os bairros mono-funcionais de escritórios industria ou residenciais dos bairros multi-funcionais	%	IBSA
nbpc_com	Número de comércios per capita	$\frac{Com.}{1000hab.}$	IBSA
pc_poppark	Percentagem da população com residência a menos de 300 metros em linha reta de um parque ou zona verde com mais 500m ²	%	IBSA
pc_nearpt	Percentagem da população que vive próxima de uma paragem ou estação de transporte público: metro, tram ou autocarros urbanos (dados de 2015)	%	IBSA
pc_nearmetro	Percentagem da população que vive próxima de uma estação de metro(dados de 2015)	%	IBSA
nb_ptstops	Número de paragens de transporte público (metro, tram, autocarros e comboio)	Paragens	Calculado de: Portal MobiGIS / SNCB

Dados	Descrição	Unidade	Fonte
dens_ptstops	Densidade de paragens de transporte público (metro, tram, autocarros e comboio)	$\frac{paragens}{km^2}$	Calculado de: Portal MobiGIS / SNCB
nb_bpark	Número de estacionamentos para bicicletas	Estacionamentos	Calculado de: Portal MobiGIS
dens_bpark	Densidade de estacionamentos para bicicletas	$\frac{Estacionamentos}{km^2}$	Calculado de: Portal MobiGIS
len_binf	Comprimento das infraestrutura cicláveis	m	Calculado de: Portal MobiGIS
dens_binf	Densidade das infraestrutura cicláveis	$\frac{m}{km^2}$	Calculado de: Portal MobiGIS
pc_surfimp	Percentagem da área impermeabilizada	%	Calculado de: Portal MobiGIS
area_bpav	Área com calçada	m ²	Calculado de: Portal MobiGIS
pc_bpav	Percentagem de área com calçada	%	Calculado de: Portal MobiGIS
slp_o5	Comprimento das ruas com mais de 5% de inclinação	m	Calculado de: Portal MobiGIS
dens_slpo5	Densidade das ruas com mais de 5% de inclinação	$\frac{m}{km^2}$	Calculado de: Portal MobiGIS
dtcc	Distância ao centro (Distância em linha reta entre as estações e o centróide do bairro "Grand-Place", centro turístico e administrativo de Bruxelas)	m	Calculado de: Portal MobiGIS
nndist_mst	Distância em linha reta entre as estações Villo e a estação de metro ou comboio mais próxima	m	Calculado de: Portal MobiGIS
area	Área das zonas de influência	km ²	Calculado de: Villo - JCDecaux
lon	Longitude na projeção "Belgian Lambert 72"	m	Calculado de: Villo - JCDecaux

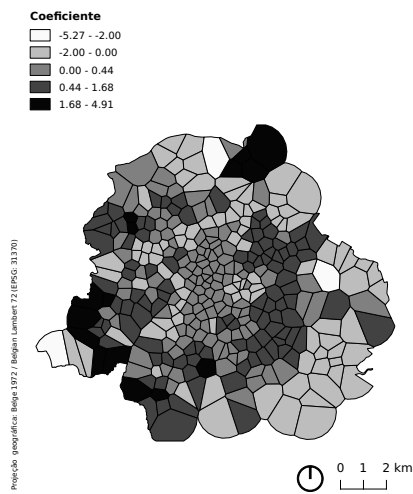
Dados	Descrição	Unidade	Fonte
lat	Latitude na projeção "Belgian Lambert 72"	m	Calculado de: Villo - JCDecaux



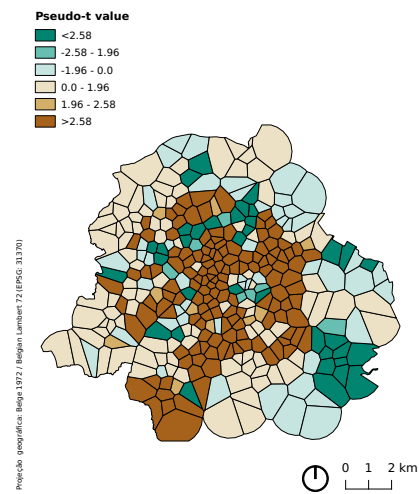
(a) Coeficiente: Dias úteis



(b) Pseudo-t: Dias úteis

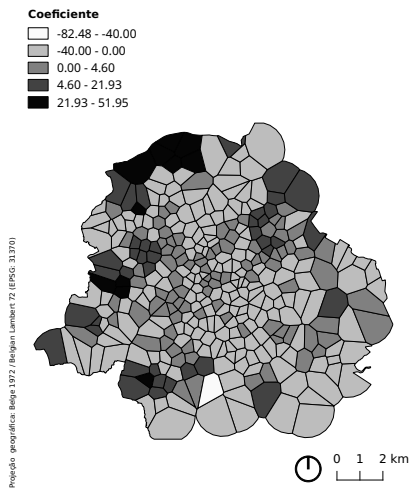


(c) Coeficiente: Fins de semana e feriados

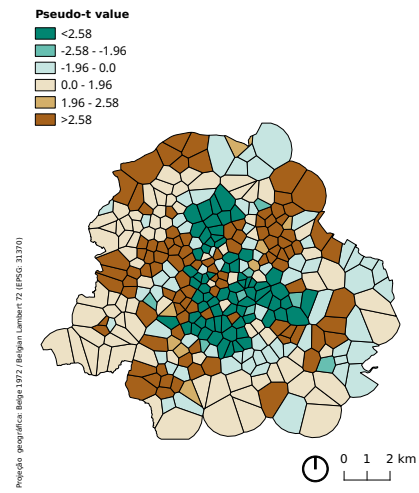


(d) Pseudo-t: Fins de semana e feriados

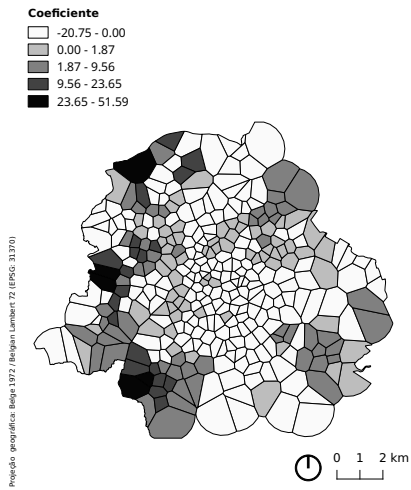
Figura A.1: Resultados SGWRP - [G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)]: Variável *Densidade de paragens de transporte público*



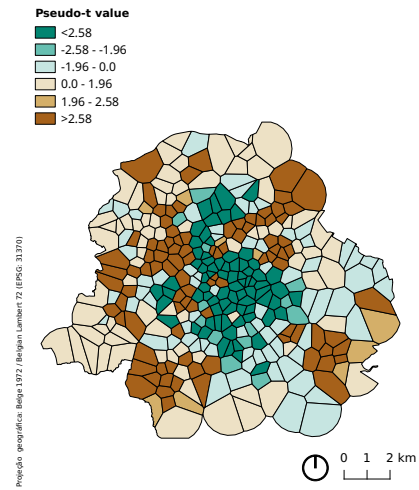
(a) Coeficiente: Dias úteis



(b) Pseudo-t: Dias úteis

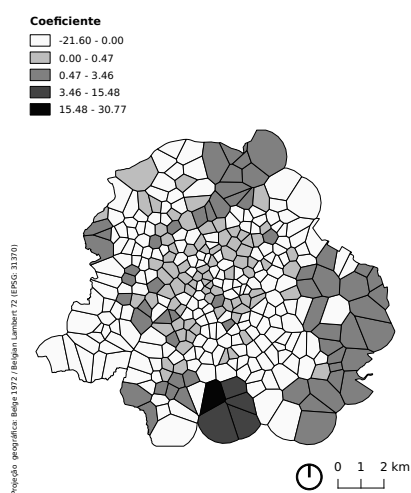


(c) Coeficiente: Fins de semana e feriados

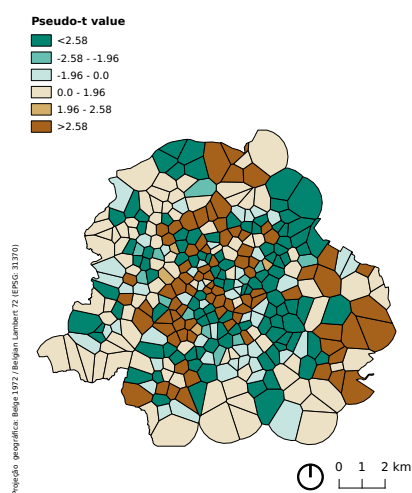


(d) Pseudo-t: Fins de semana e feriados

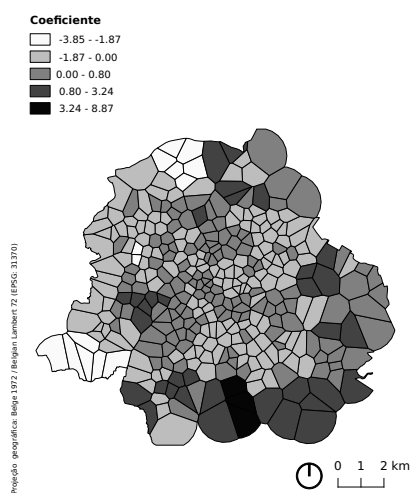
Figura A.2: Resultados SGWRP - [G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)]: Variável *Densidade de escritórios*



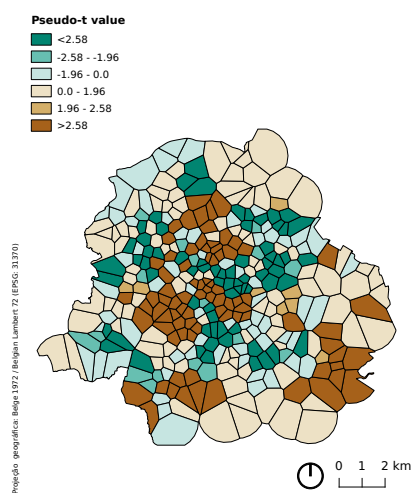
(a) Coeficiente: Dias úteis



(b) Pseudo-t: Dias úteis

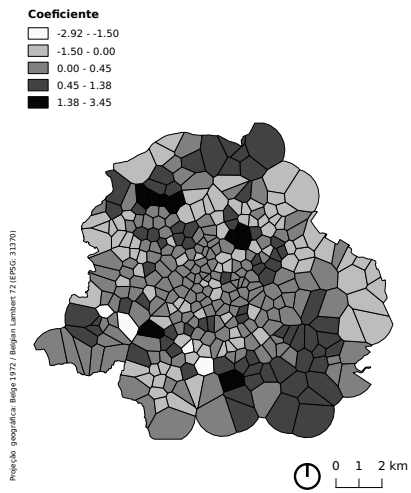


(c) Coeficiente: Fins de semana e feriados

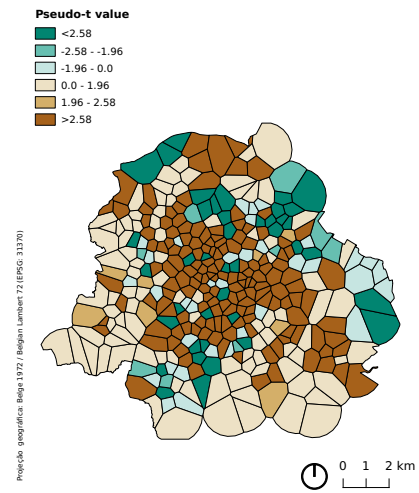


(d) Pseudo-t: Fins de semana e feriados

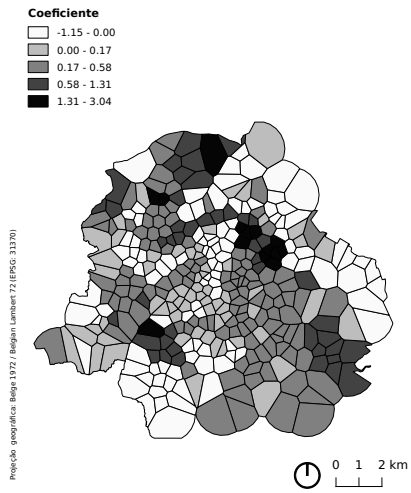
Figura A.3: Resultados SGWRP - [G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)] (Dias úteis / Fins de semana e feriados): Variável *Densidade de estacionamentos para bicicletas*



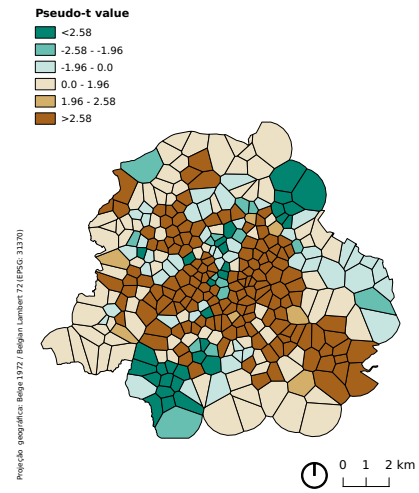
(a) Coeficiente: Dias úteis



(b) Pseudo-t: Dias úteis

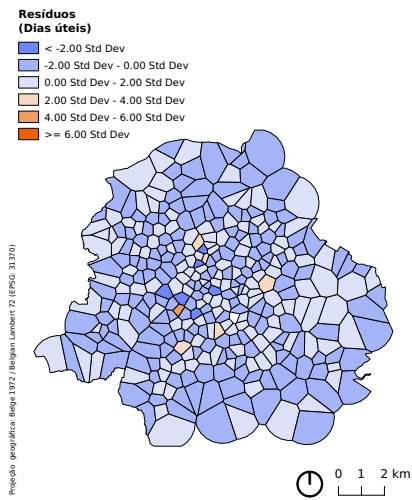


(c) Coeficiente: Fins de semana e feriados

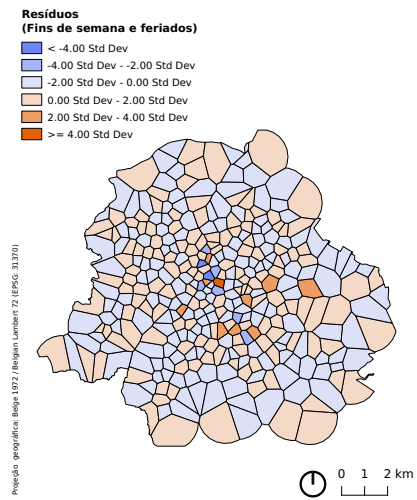


(d) Pseudo-t: Fins de semana e feriados

Figura A.4: Resultados SGWRP - [G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)] (Dias úteis / Fins de semana e feriados): Variável *Número de docas Villo*

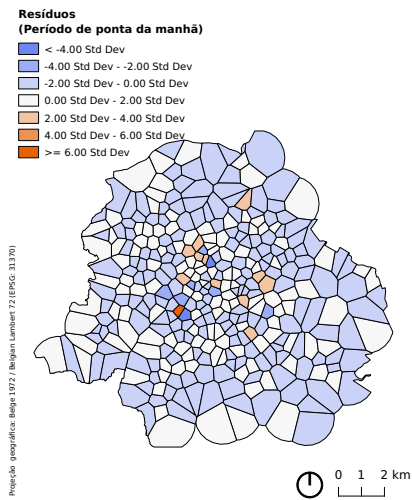


(a) Dias úteis

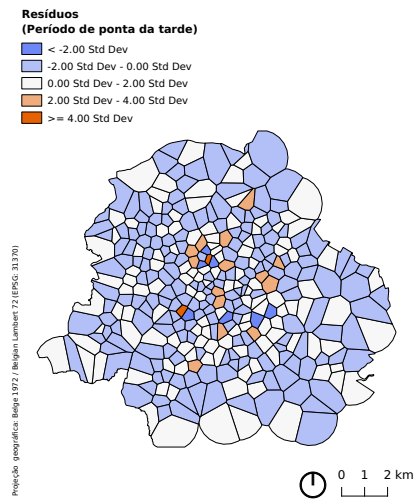


(b) Fins de semana e feriados

Figura A.5: Mapa de resíduos: Modelo SGWRP - [G(dtcc); L(bst; dens_ptsto; dens_bpark; dens_offic)]



(a) Período de ponta da manhã



(b) Período de ponta da tarde

Figura A.6: Mapa de resíduos: Modelo GWRP - (bst; dens_ptsto; dens_bpark; dtcc; dens_offic)

C& SIG



UNIGIS PT

