Nova School of Business and Economics

Universidade Nova de Lisboa

Dissertation, presented as part of the requirements for the Degree of
Doctor of Philosophy in Economics and Finance

Leveraging Geospatial Statistics for Measuring and Valuing the Urban Environment

Jacob L. Macdonald
- 8 -

A Dissertation carried out on the PhD in Economics and Finance, under the supervision of
Professor Sofia F. Franco

Lisbon, October 2019

For mom

“The bear went over the mountain
to see what he could see”

ABSTRACT

This thesis looks at emerging uses of geospatial data for analysing the urban environment. As high-dimensional data becomes increasingly available, sophisticated spatial and temporal statistical estimation strategies can assess the minutia of environmental processes in a dynamic urban context. Each essay focuses on the improved measurement of high-resolution non-market environmental amenities and evaluating them using observed impacts on house prices or transportation networks. While valuation techniques for each amenity vary depending on context, these works all highlight a set of spatial methodologies for detailed urban analytics with a particular focus on urban greenery, seismic and flood risk, and pollution mitigation.

ACKNOWLEDGEMENTS

It should need no acknowledgement that this work is the product of support and mentorship from a large collective of friends and family. For the many different ways in which people have contributed their part – Thank you.

First and foremost to my family. To my Dad for the constant motivation to strive and pursue great things. Without the push to accept this move and the help during and after the transition, none of this would have been possible.

To Katie, Julie and Hugo – always there and always providing the best encouragement. Life is different now than when I started this but as I powered through, knowing that family is constant has easily guided me through it all. Given the amount of attention and distraction coming from writing this thesis, this work is in a sense a child of my own that I have been raising with lots of help. Going forward, it is my turn now to return the favours and support.

To my Mom who gave me absolutely everything I need and more. The most inspiration woman I know demonstrating that kindness and humanity are the most important things above all else.

A significant amount of gratitude needs to be given to the entire community at Nova SBE. The entire network of course instructors, academic faculty, administrative support workers and management have been a large part of my life over the past six years and each have had great impact on me. Having come from away, I couldn't have asked for a more supportive network of people helping me along this process. These years spent within the university walls, from *Campolide* to *Carcavelos*, have been the most formative period of my life. The motivation and drive for success was planted early and will continue with me throughout my career path.

To my colleagues turned closest friends. Julia S., Stefan L., Rute C., Robbie H., Felix B., Matilde G., Inês V., Joana C., Sara A., Ana G., Mattia F., Ana L. This is as much a team effort as anything, and I couldn't have asked for a better team. The collaborations, discussions, support and good times have contributed as much to this work as anything that can be taught in a classroom.

For the friends who supported this entire process in their many ways from Canada, Amy G., Ashley P., Michelle T., Kyla B., Lesley H., Lydia OC, Lauren F., Jackie W., Anna N., Kristi F., from Lisbon, Hugo P., Daniella M., Maria M., Francesca F., Marta C., Tom W., Manuel M., Joao N., Manuel H., and all those abroad. Having such a network of motivation has literally opened up the world to me. Thank you for following through with me.

It would be incomplete not to acknowledge to profound impact that Lisbon itself has had on me. What started as culture shock very quickly turned into my second home. It should not be lost on anyone that Lisbon's aesthetics, culture, dynamics and history are a large part of what shaped my academic interests and the topic of this thesis. So much of this work stems from simple walks along the tree lined streets, coffee and *pasteis* in front of *Neptuno*, meandering the medieval streets of *Mouraria*, watching the sunrise from a *miradouro*, afternoons in *Jardim da Estrela*, *tascas*, *fado* and *vinho da casa*. These essays are my attempt to capture a fraction of these experiences down on paper.

Finally, to Sofia Franco. This work, and the PhD as a whole, would not be what it is without the mentorship and support provided. The detailed attention and guidance, through all parts of this process, were paramount to its success.

CONTENTS

LIST OF TABLES

Chapter 1:

Chapter 1 Appendix:

Chapter 2:

Chapter 2 Appendix:

Chapter 3:

LIST OF FIGURES

Chapter 1:

Chapter 1 Appendix:

Chapter 2:

Chapter 2 Appendix:

Chapter 3:

Chapter 3 Appendix:

VOCABULARY AND TERMINOLOGY

AIC	Akaike Information Criteria
CBD	Central Business District
CML	*Câmara Municipal de Lisboa*
CO	Carbon Monoxide
CV	Cross Validation
GDR	Geographic Regression Discontinuity
IDW	Inverse Distance Weight
IV	Instrumental Variable
LEZ	Low Emission Zone
LL	Log-Likelihood
LM	(Spatial) Lagrange Multiplier
LOOCV	Leave One Out Cross Validation
NDIV	Normalized Difference Vegetation Index
NIR	Near-Infrared
NN	Nearest Neighbour
NO	Nitrogen Oxide
NO_2	Nitrogen Dioxide
NO_X	Nitric Oxides
O_3	Ozone
OLS	Ordinary Least Squares (Regression)
PM_{10}	Particulate Matter
RMSE	Root Mean Square Error
SAR	Spatial Autoregressive Model
SARAR	Spatial Autoregressive Model with Autoregressive Disturbances
SEM	Spatial Error Model
SO_2	Sulfur Dioxide
SSE	Sum of Squared Errors
SVM	Support Vector Machine (Learning Algorithm)
SW	Spatial Weight (Matrix)
VIF	Variance Inflation Factor

PREFACE

The overarching theme of this dissertation is the use of geospatial data and techniques to measure, value and explore the impacts and spill overs of urban environmental amenities and dis-amenities. In exploring these concepts each chapter uses frontier spatial and temporal statistical methods applied to high dimensional, multi-tiered big-data in a dense urban context.

Advances in geographic data collection, storage and analysis has yielded significant improvements in the scale of detail available for sophisticated urban analytics. While techniques such as remote sensing and spatial interpolation improve the data and coverage itself, geographic discontinuities and spatial identification strategies can be used to estimate non-market benefits and costs related to environmental outcomes.

As municipalities are at the forefront of managing environmental concerns from flooding or pollution, for example, better understanding the value and spill over impact of these types of urban amenities are crucial for tailoring best practices for local policy implementation. Spatial data advances thus have huge potential in the future of empirical urban research, especially as applied to the collection and use of novel environmental data.

The first chapter of the dissertation focuses on the valuation of urban green amenities in Lisbon. Using a machine learning remote sensing algorithm, aerial images of Lisbon are classified to identify the distribution and sparsity of tree canopy, vegetation, and levels of greenness across the city. This measure more accurately portrays how residents perceive the ecology and greenness of their neighborhood. Hedonic regressions then value the classifications of greenness via the real estate market. Results show positive values for healthy vegetation and variation in the value of open space amenities conditional on size and greenery.

The second chapter focuses on the impact that urban geohazard risks in the form of seismic or flooding susceptibility have on the property market. When persistent risks of an urban hazard exist, residents may significantly undervalue their properties accordingly. Here, the research makes use of a spatial regression discontinuity to identify these effects and

measure the dis-amenity value that geohazard risks transmit, and whether this effect can be mitigated or exacerbated by locational amenities such as greenery or urban topography.

The final chapter focuses on optimizing the interpolation of spatial-temporal air pollution to construct a space-time longitudinal database for the municipality of Lisbon tracking the distribution of air quality since 2000. A significant emphasis is put on determining the most appropriate means to conduct the spatial interpolations and aggregation of pollution point data and is based a series of generalizable algorithms and diagnostics which can be applied in varying urban contexts.

Transit interventions in the form of metro station openings and low emission zones are then studied to estimate their pollution abatement influence across the city in the short and long-run. The application uses spatial-temporal difference-in-differencing, borrowing from the spatial regression discontinuity design framework to identify thresholds and decay effects. Results suggest that local pollution is significantly abated when transportation initiatives, in the form of metro stations or low emission zones, are opened. Further, these impacts are not equal across the city suggesting changes to the overall transport patterns.

These applications highlight the advancement in how geographic data can be used for detailed urban analysis. Levering the spatial or temporal nature of high-dimensional data allows for a wide variety of cost and benefit valuations.

Jacob L. Macdonald

October 2019

CHAPTER 1:
Machine Learning for Measuring and Valuing Urban Greenery

This chapter explores the role of remote sensing in capturing urban environmental data in the form of tree canopy coverage and measures of greenery. A machine learning classification model is applied to high-resolution aerial photography of Lisbon, Portugal. Aggregating measures to a neighbourhood scale allows the exploration of the impact of greenness and vegetation on the residential property market, capturing the heterogeneity and complementarity relative to open spaces and other local environmental attributes.

1 – 1. Introduction

Urban green amenities range from planted street trees, manicured parks and gardens, natural forests, and green infrastructure such as green walls and roofs. The amenity value of these local public goods not only service residents and visitors but further have important interactions with the local ecology. In addition to contributing to neighbourhood aesthetics and appeal, benefits include the mitigation of storm water runoff and pollution, balancing the urban heat island effect and increased biodiversity, all of which spillover and influence human quality of life and health (Zupancic et al. 2015). The value of open spaces and greenness is tied to ongoing discussions within many municipalities regarding city-wide greening policies.[1]

While there is extensive empirical work on the valuation of open spaces and green amenities, research in the context of the urban environmental literature is challenged by the necessity of highly detailed spatial data. Moreover, previous research on the amenity effects of open space on the real estate market have typically used distance to different categories of open space, permanency of open space, view of open space or proportion of open space within a dwellings neighbourhood as the primary variables to evaluate these land uses.

These studies have helped in understanding how households value accessibility to alternative urban green amenities however there is still a need to fully understand how

[1] Since 2012 Lisbon has prioritized greening in their municipal guidelines with specific objectives to create new green spaces, transform existing spaces and infrastructure by planting street trees, linking the city through *green corridors* and the creation of community allotment gardens (*Câmara Municipal de Lisboa* 2015b). Further the municipality is targetting a 20% increase in biodiversity by 2020 (*Câmara Municipal de Lisboa* 2015a).

community greening efforts, especially those which do not require significant land commitments, may impact property values and subsequent tax revenue. This is particularly important in central areas of a city with scarce and expensive vacant land. In fact, community greening through planting trees, tailoring lawns, flower arrangements and other types of sparser visible vegetation, have the potential to increase the attractiveness of neighbourhoods through its aesthetics without the acquisition of large amounts of land.

If households value the overall greenness of their residential neighbourhood, then there should be a positive capitalization through housing values of these landscape attributes. Yet, a valuation of the latter green amenities requires the development of measures of neighbourhood greenness that incorporate the various dimensions of green cover and also capture the extent of green coverage for residential properties in an urban neighbourhood context.

This chapter examines the role of machine learning remote sensing techniques to create coverage measures for neighbourhood greenery from images that serve as explanatory variables in hedonic valuation models. Additionally, the work explore how such data can be interacted with other geospatial variables for enhanced measures of the urban environment.

In particular, the results of a one-class support vector machine (SVM) supervised learning algorithm are used to capture urban tree canopy coverage and sparser vegetation, and a normalized difference vegetation index (NDVI) to measure green density and land-use diversity in close surrounding of properties using high resolution aerial photography.[2] The technique is applied to Lisbon, Portugal, with the results incorporated into a hedonic framework that includes not only the traditional measurements of proximity to urban open space as explanatory variables of the impacts of urban greenness on property values, but also includes greenness variables that capture the extent and quality of greenspace coverage in a

[2] Tree canopies defined as the above ground biomass attributed to trees in the form of branches and foliage, forming the crown of the tree which covers a ground area when viewed from above.

neighbourhood.[3] We test the hypothesis that the contribution of traditional open space measures to residential property values is conditional on the overall pleasantness of a neighbourhood by tree canopy cover or type of vegetation, and further explore the substitutability and complementarity of these types of green attributes with other urban features.

Traditional methods of collecting data on urban tree canopies involve in situ sampling or the usage of municipal inventories that can be costly, time consuming or subject to data availability. Thus, machine learning algorithms applied to remotely sensed aerial photography is a valuable complementary tool to accurately classify urban tree canopy coverage and vegetation, increasing the dimensionality though which greening can be analyzed. Previous classifications for specific urban areas in the United States estimate that, on average from the cities sampled, approximately 27% of a city is covered by urban tree canopy (Watt and Gunther 2010). Different studies use varying scales and resolutions which may overestimate coverage of tree canopy by classifying sparse contiguous tracts of trees as complete coverage.

The remote sensing of urban tree canopy indicates that approximately 8% of Lisbon is covered directly by canopy with an overall accuracy rating of approximately 90%. These results are based on very high-resolution images, and can thus accurately capture the density and sparsity of tree canopy coverage. Moreover, the residential real estate market values the relative size of neighbourhood tree canopy coverage. We estimate the effect of neighbourhood tree coverage to be approximately 0.20% of a dwelling price, however there are positive ecological interaction effects between greenness and broader measures of the urban environment. Results indicate positive amenity value for vegetation in mitigating storm water runoff, complementary effects between historic conservation areas and lush vegetation, and yields some substitutable benefits to properties which are not located near the Tagus River.

[3] Urban open space is a term used generally to define areas of vacant lots, natural landscapes and public green space in cities. However, the definition of urban open space has evolved in time embracing various types of urban open space (such as playgrounds, parks and urban forests) and green elements.

Proximity to urban forests are positively valued at 0.03% per kilometer decrease in distance with stronger effects from proximity to larger urban forests which provide the most recreational services. Alternatively, while residents may not value living near parks, there is a marginal premium for living closer to smaller parks compared to larger ones, potentially due to the heterogeneity of these parks and the congestion of visitors at the largest of them. Results suggest that residents value more the size of nearby open space rather than the greenness of the space, yet value overall greenness for their neighbourhood and surrounding areas.

The remainder of the chapter is structured as follows. Section 2 positions this research in the economic literature, highlighting studies that have incorporated remote sensing to assess the impact of open space amenities and greenness on housing markets. Section 3 describes the study region and section 4 presents the data and describes the measurements of neighbourhood greenness. The results and accuracy assessment from the one-class SVM procedure are then presented in section 5 while section 6 applies these results within a hedonic model testing for spatial dependence and including interaction effects to value how open space and urban greenness influences property values. Finally, section 7 provides overall conclusions.

1 – 2. Literature Review

There is an extensive body of work on the urban environment and interactions with broader socioeconomic and ecological factors. Research however is challenged by limitations in accurately measuring environmental variables. With increasingly powerful data capturing techniques, such as remote sensing, the measurement of detailed environmental data is feasible and allows for increased work on the impacts of a range of environmental and green variables.

The value of green and open spaces has been subject to a number of studies using hedonic and stated preference methods, as surveyed in McConnell and Wall (2005) and Waltert and Schläpfer (2010). Within these studies, there are a range of different methodologies and scopes in how open spaces have been measured and categorized. In general, the results are

mixed as in the case with the valuation of urban forests, with positive, negative and insignificant effects found for similar amenities. Spatial dependence and interaction effects are important component of these studies and are addressed by modelling this spatial dependence or incorporating interaction effects to capture locational heterogeneity.

Much of the early work focuses on evaluating the accessibility of open spaces including its permanency (Geoghegan 2002, Irwin 2002), the view of an open space (Tyrväinen and Miettinen 2000) or accessibility to different types of urban open spaces through distance to the amenity or a dummy variable indicating proximity within a buffer (Tyrväinen and Miettinen 2000, Morancho 2003, Kaufman and Cloutier 2006, Voicu and Been 2008). Capturing measures of urban greening however is an ongoing challenge and a smaller subset of studies have evaluated different measures of urban greenness via the hedonic pricing methodology including tree cover (Anderson and Cordell 1988, Dombrow et al. 2000), landscaping attributes (Des Rosiers et al. 2002) and NDVI measurement of green density (Kestens et al. 2004).

On the other hand, most recent work has focused on capturing the heterogeneity of different categories of open spaces, each with different amenities and benefits (or costs) to residents, while controlling for spatial autocorrelation. Fewer studies have explored the impacts of overall neighbourhood greenness. Sander and Polasky (2009) consider proximity to parks, trails, lakes and rivers, testing and rejecting for spatial autocorrelation in their data. Czembrowski and Kronenberg (2016) group open and green spaces into 9 different categories: small, medium and large parks and forests, the single largest forest, cemeteries and allotment gardens in Lodz, Poland. The authors correct for spatial dependence with a spatial autoregressive model with autoregressive disturbances and find that proximity to large parks and the Lagiewniki forest, which are well known in the city, have the strongest positive impacts on housing prices with property values increasing approximately 1.5% per square meter as

proximity increases. Cemeteries had the opposite effect with prices per square meter decreasing as dwellings were located closer in the range of 2%.

To capture spatially varying amenity values of open spaces, certain studies have focused on the geographically weighted regression technique to allow the coefficient to vary across the study area with results indicating that there is significant variation in how different areas of a region value open spaces (Cho et al. 2006, Nilsson 2014). Cho et al. (2008) use a novel way to quantify open space and open space quality by exploring the effect of proximity to forest stands with different compositions of tree types (deciduous versus evergreens) and further by the shape of these stands.

In addition to proximity to open space amenities, some works have also studied the impact of higher concentrations of open spaces at the parcel level as a measure of neighbourhood open space and, indirectly, neighbourhood greenness. Irwin (2002) measures the percentage of different types of open spaces (cropland, pastures, forestry, conservation areas, public parks or military lands) within a 400-meter buffer to estimate the marginal impact of different open spaces and in general find that non-developable open spaces tend to have a positive impact on housing. In Castellón, Spain, Morancho (2003) find that distance to open space is a more significant determinant than the size of the open spaces with a 100-meter increase in proximity yielding a premium of €1,800.

Anderson and West (2006) consider both the proximity to different categories of open space as well as the size of the nearest open space to capture these influences for parks, cemeteries, golf courses, rivers and lakes and explore how the amenity values change under neighbourhoods with different characteristics. The authors include a range of interaction effects to estimate open space and its heterogeneous influence with neighbourhood demographics, crime rates, distance to central business district (CBD), income levels, density and lot size. Parks have positive impacts on housing prices, and the effect of being closer to a park is nearly

three or four times as high in neighbourhoods which are twice as dense as average and with twice as high income levels respectively.

Donovan and Butry (2010) estimate a spatial hedonic regression of street trees on property values in Portland, Oregon. The authors collect data on the number of street trees fronting a house, tree crown area in the front and within 100 feet of the house, height, volume and whether the trees are flowering, fruiting, coniferous, deciduous, appears to be sick or poorly pruned. Of these variables only meaningful results are obtained with the number of trees and the area of tree canopies within 100 feet, suggesting that in-situ sampling of trees to collect detailed data may not be necessary. Remote sensing techniques can be used to obtain the most relevant characteristics such as tree location and canopy with lower costs.

Machine learning methods are applied to satellite images or aerial photography, reading pixel level data and classifying ground objects based on patterns and relationships. While satellite images are freely available from various sources, they tend have a low spatial resolution which is unable to capture the heterogeneity of land use at the city level.[4] In order to distinguish urban tree canopies, high spatial resolutions are needed or alternatively images can be obtained via aerial photography on a low flying plane or drone. Remotely sensing environmental data from external sources is well developed, and images have been used to classify changes in land cover and detect objects on the ground.[5] Fewer studies, however, have incorporated the results of this classification into broader economic valuations, often favouring to use data from external sources which may or may not align with the scope of the research.

Remote sensing of tree canopies has most commonly been applied to larger regions or country level images as for example in Iizuka and Tateishi (2015) or Karlson et al. (2015).

[4] Both the United States Geological Survey and European Space Agency host free satellite images for the globe from the Landsat and Sentinel programs respectively.

[5] Applied remote sensing work has been used for, among others, the collection of urban pollution data, the classification of automobiles for the study of traffic density, the study of urban sprawl via changes in lights, capturing changes in rooftops and capturing urban vegetation.

Some studies have focused on tree canopies in an urban setting which is complicated by the high quality of images required to accurately indentify canopies located among the mixture of residences, commercial buildings, public spaces and dense network of roads. In the urban context, Li et al. (2015) test different classification algorithms to determine urban tree canopy coverage in two areas of Beijing and conclude that a SVM algorithm is prefered.

Parmehr et al. (2016) further highlight the power of remote sensing techniques by comparing remotely classified tree canopies with canopy coverage obtained from the *i*-Tree software, a software program developed to assist in the classification of trees, in a suburb of Melbourne, Australia. The authors determine that remote sensing techniques provide the flexibility of capturing continuous canopy coverage across a study area which is invaluable for research on the impacts of urban tree canopy.

A smaller subset of articles tie-in remote sensing to a broader applied urban economic analysis. In Baltimore, Troy and Grove (2008) include proximity to parks and crime interaction effects to examine the heterogeneity of impacts from different parks using remote sensing to classify parks as minimum two hectares with a high NDVI index, a measure of vegetation obtained using the near-infrared and red pixel information of the satellite image. Parks are valued positively up until a certain threshold level of crime of over 400% the national average, after which there is a negative effect on housing prices.

Troy, Grove and O'Neil-Dunne (2012) use 1 meter remote sensing data to estimate that a 10% increase in canopy coverage yields an approximate 12% decrease in crime rates. Wolfe and Mennis (2012) obtain a negative relationship between urban vegetation and crime with a decrease in levels of assault, robbery and burglary for census tracts with higher vegetation levels, as measured by the mean NDVI of the census tract based on 30 meter image resolution.

Conway et al. (2010) explore urban greenery by estimating the impact of different amounts of greenspace within buffer rings of a dwelling, controlling for spatial autocorrelation

in their housing prices. Their results indicate that a 1% increase in the amount of greenspace within 200 to 300 feet of a dwelling would result in an increase of 0.07% in the sales price.

Sander et al. (2010) obtained tree canopy coverage at a 30 meter resolution from the National Land Cover Database, a relatively low resolution which may limit the detail when it comes to identifying sparse or dense canopies of individual trees.[6] The authors use the percentage of tree coverage within buffers up to 1000 meters, and further the distance to lakes, the nearest park and the nearest trail. Under a spatial error specification, the authors argue that tree canopies are valued in direct proximity to dwellings with little influence further away in the magnitude of a 0.48% and a 0.29% increase in dwelling prices for a 10% in tree coverage within 100 meters and 250 meters respectively.

1 – 3. Study Region

The study area is the capital city of Lisbon, Portugal, covering approximately 100 km^2 with a population of 552,118 and 2 million residing in the greater metropolitan area. Lisbon was founded on the banks of the Tagus river and is one of the oldest capital cities in the world. The city has a rugged topography built on seven historic hills extending from the riverfront and a dense central area and peripheral zones emerging from the original medieval foundations. The river, running along the western and southern portions of the city, is an important amenity to both residents and the economy. The port is a major trading hub and further hosts many cruise ships every year. For the residents the riverfront offers tracks for running and cycling, viewpoints, modern amenities like restaurants and coffee kiosks and access to the water.

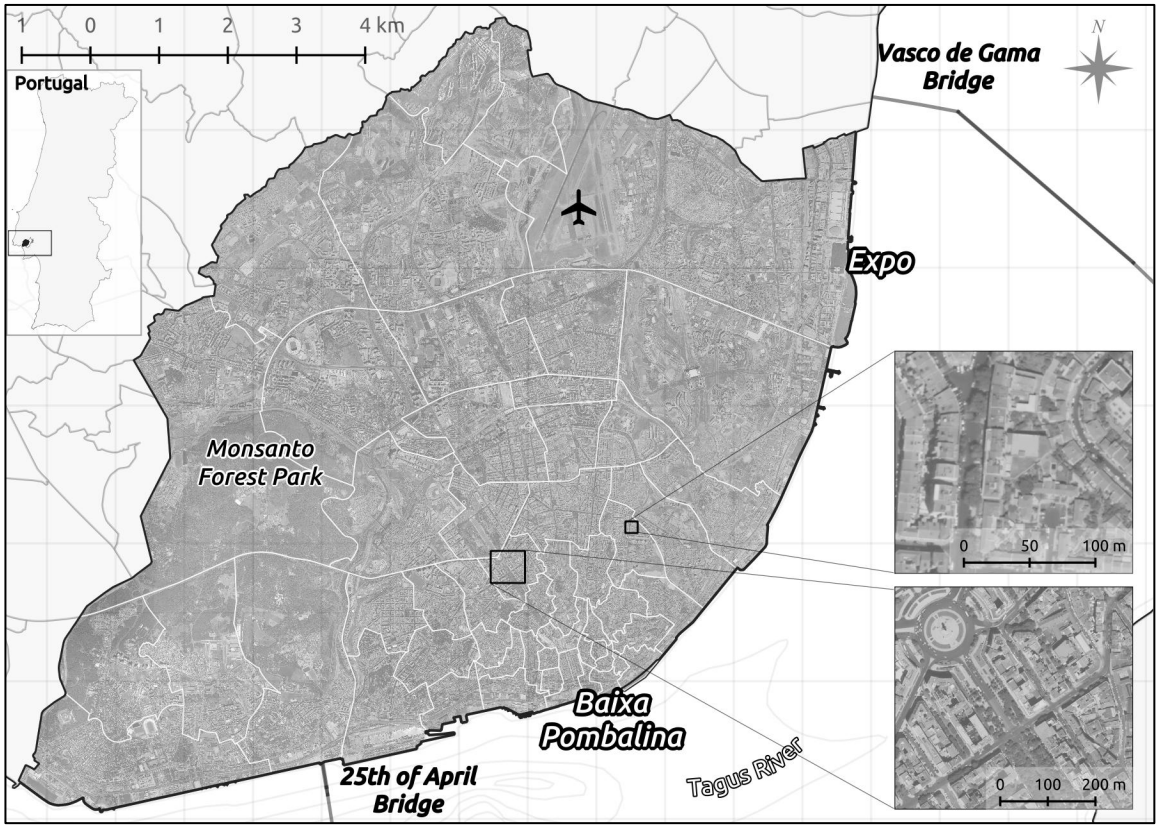
Today, Lisbon maintains its status as an economic and cultural hub in Europe and is the largest city in the country. The city has many of businesses with 311,000 firms primarily clustered across two predominant areas. The primary and historic CBD is known as *Baixa*

[6] A 30-meter resolution indicates that each pixel of the image represents 30 meters of ground coverage.

Pombalina and bordered by the river in the south. In 1998, Lisbon leveraged it's hosting of the World Expo to redevelop a previously idle area into a secondary CBD, known as *Parque das Nações* (or "Expo"), located further inland. This northern area of the city is also where the international airport is located. Two bridges connect the city to the alternate side of the river with many commuters entering the city each day.

Administratively, the city is divided into census enumeration tracts which generally align with street divisions, while *freguesias* (or civil parishes) are representative of a broader area within the city. The city is divided into 53 *freguesias* which align with the historic and cultural evolution of distinct neighbourhoods in the city.[7] *Freguesias* in *Baixa Pombalina* are among the smallest, on average 0.12 square kilometers, compared with the largest located out at the city periphery which is 11 square kilometers. Figure 1 shows aerial photography of the study region, key locations and *freguesia* level demarcations.

Figure 1. Study Region Aerial Photography: Lisbon, Portugal



[7] In 2012 the municipal council approved the reorganization from 53 *freguesias* to 24, however the most recent Census 2011 and many data sources continue to make use of the former classification of 53.

Lisbon has a Mediterranean climate with two distinctive seasons: a hot dry summer where temperatures commonly exceed 30°C (86°F) and a cool wet winter with sometimes intense periods of rain causing significant flooding. Yearly, during the rainy winter months, many parts of Lisbon experience sometimes severe flooding. The topography of the city means that these occurrences are not limited to the riverfront and we observe heterogeneity in this urban risk. Trees are a significant tool used in mitigating storm water runoff, and the municipality has recently approved a €170 million drainage plan to combat future flooding.

The diverse history has led to a diverse urban fabric across the city with a mixture of dense historic buildings and cobblestone streets juxtaposed against the newer buildings of modern Lisbon. Urban greening is a priority of the city, and there are many different types of green infrastructure including planted trees, open spaces, green corridors, and many different configurations of vegetation and flora. There are over 120 local parks and gardens ranging from small neighbourhood parks to those large and ornate. These amenities tend to be more manicured with walkways and flowerbeds, and may further be adorned with monuments or water features, often dating back to Portugal's Age of Discoveries in the 15th and 16th centuries.

There are 13 urban forests which are larger and offer denser and more natural tree stands and recreation facilities. The city skyline is dominated by the Monsanto Forest Park covering approximately 10% of the city and reaching 227 meters. This the largest urban forest in the region and offers many recreational amenities including trails, cycling, sport facilities and picnic areas, among others.

1 – 4. Measurement of Urban Greenness

The analysis is extended beyond traditional measures of proximity to open spaces by estimating variables via machine learning of remotely sensed, multi-spectral aerial photographs. This is done with georeferenced and orthorectified photographs covering Portugal taken by the *Direção Geral do Território* under the national Ministry of the

Environment. The images contain data in two important dimensions. Firstly, finer detail corresponds to a higher number of pixels in the image, known as a higher spatial resolution. The intensity of each pixel represents one unit of data in the form of a digital number, and thus more pixels translate to more data. It is from these digital numbers that a classification is built to identify high-detail greenness on the ground.

Secondly, aerial cameras are able to capture light from different parts of the electromagnetic spectrum, including what may not be visible to the human eye such as near-infrared (NIR). The different spectral bands capture the amount of light in each pixel representative of that specific part of the spectrum. When the visible bands (i.e. the *blue*, *green*, and *red* bands) are combined we have the image as would appear to the naked eye.

Aerial images of Lisbon were captured between July 27 and August 23, 2007 from an Intergraph Digital Mapping Camera DMC01-0037 and UltraCam-D. The images are available with a very high spatial resolution in four spectral bands: blue, green, red, and NIR. Every pixel represents 50 centimetres on the ground and for each we have a value representing the intensity of the four bands. The high spatial resolution of these images is what makes the detection of detailed objects such as tree canopy feasible, while the spectral resolution is used to classify objects as vegetation based on their colour and energy radiation.

1 – 4.1. Remote Sensing and Urban Greenness

Using geospatial and remotely sensed data helps in capturing detailed environmental characteristics and turning them into measures of urban greenness that can be used in a hedonic framework. A generated continuum of urban greenness across the city can thus be aggregated to capture different forms of local vegetation concentration including aboveground tree canopy and sparser vegetation. Administrative boundaries in Lisbon are used to define the spatial scale at which different measures are constructed, the largest representing *freguesias* and the smallest representing the over 3,623 census enumeration tracts (indicative of city blocks).

The focus of the machine learning process is to identify aboveground tree canopy coverage as identified by the pixel level data and in part through the normalized difference vegetation index (NDVI), an index used to develop additional measures of urban greenness. Classifications on pixel level data and broader patterns of surrounding the pixels are leveraged to identify the unique aspect of the green infrastructure of interest.

Vegetation Index

The NDVI is a key feature in the classification of urban canopy, and is in itself an important measure. Using the red (R) and NIR spectral bands, the NDVI is a synthetic band and calculated as NDVI = [NIR-R]/[NIR+R].[8] This index indicates the relative greenness of a pixel based on the reflectance of the red and the NIR spectral bands. Plants absorb visible (red) light during photosynthesis, with healthy plants absorbing more visible light. Further, the cell structure of plants reflect NIR thus the NDVI is a relative measure of healthy vegetation.

Different land covers yield different values of NDVI ranging from -1, representative of water to, +1 representative of the healthiest and lushest vegetation. Moderate and sparser vegetation tend to have values greater than 0.2 with more lush vegetation having higher values closer to 0.6 or 0.8. At the neighbourhood level, a high mean NDVI is representative of greener areas with more trees and vegetation. There is clear heterogeneity in the level of vegetation across Lisbon with low mean NDVI values in highly developed areas such as the primary CBD, *Baixa,* (-0.02) compared to the rest of Lisbon with higher levels of average vegetation (0.05).

Measures of Urban Greenness

The NDVI highlights areas rich in vegetation, however it may be vague in its interpretation and represent amenities with very different ecological functions. For example,

[8] The digital numbers associated to each pixel within each band is represented on a scale of 0-255. The NDVI, although measured on a scale of 0-1, is converted to 0-255. This keeps all variables in 8-bit format so that they remain integers as opposed to floating points, which increases file size and computation time significantly.

while both a closed canopy urban park and a sport field may both yield relatively high NDVI values, trees in an urban parks have an important role to play in carbon sequestration and balancing the urban micro climate while sport fields are primarily recreational for residents. The NDVI measure is used in a remote sensing algorithm to obtain the tree canopy coverage which, unlike the NDVI itself, identifies a particular land cover class.

This provides a measure of the pure canopy effect coming from the collection of above ground foliage from tree lined streets, public parks and private gardens. This is a valuable urban environmental variable given the important ecological function of trees including the mitigation of the urban heat island effect and storm water runoff and further for their aesthetic value and complementarity to the urban structure and in particular historic landscapes.

For each census tract the mean value of the continuous NDVI measure is calculated and differencing is used to estimate the value of all non-canopy vegetation remaining to capture the impact of other types of flora such as lower lying shrubs, lawns and flowers. The pixel level percentage of a tract covered by different types of vegetation are classified by the NDVI in two categories: firstly the percentage of pixels with NDVI values between 0.2 and 0.5, and secondly the percentage of pixels above values of 0.5. These measures capture the composition of vegetation where 0.2 and 0.5 represent low lying sparse vegetation and values above 0.5 represent healthy trees and greenness. Whether the local area of interest is the *freguesia* or the census tract, when the greenery is measured in percentages it can be interacted with the respective zonal area to determine the relative extent of average vegetation size in the area.

These measures are used in valuing the marginal implicit price of a range of different types of neighbourhood greenness. When combined with interaction effects, results indicate the substitutability and complementarity of this neighbourhood greenness with proximity to other urban open spaces such as parks and forests, ecological dis-amenities such as flood risk, and further other neighbourhood attributes such as being in a designated historic area.

1 – 5. Machine Learning Remote Sensing of Tree Canopies

With the increasing availability of high resolution satellite and aerial images, it is possible to distinguish detailed urban land use at a very fine scale. Examples of the types of urban features which can be extracted via remote sensing include, among others, pollution, vehicles and traffic, rooftops, land use and fragmentation, urban sprawl and green amenities. Here, machine learning remote sensing methods are used to classify urban tree canopies.

Results yield highly detailed data which can be used to provide context to various urban greening policy discussions. Accurately quantifying the size and location of tree canopy in a city provides important foundations for improved research and decision making regarding their economic and social benefits and costs. Further, urban tree canopies are distinctive in comparison to other green amenities in that they are the primary sources of carbon sequestration in municipalities and significant in offsetting total carbon emissions.[9]

1 – 5.1. One-Class SVM Learning Methodology and Training Sample Collection

Land use classification categorizes the elements of an aerial image based on the underlying pixel information. Images are composed of the different spectral bands which, when layered together, produce the image visible to the naked eye. Each layer represents the intensity, at each pixel, of the respective band. This intensity is stored as a digital number between 0-255, providing a large database from which to classify objects. A one-class SVM supervised machine learning algorithm as outlined by Schölkopf et al. (2001) analyzes underlying patterns in spectral information (and the synthetic NDVI band) to classify pixels. In this context a specific (one-class) algorithm identifies the unique class of interest, canopy.

Prior to the classification, a sample is used to train the algorithm in identifying canopies based on which pixels are true tree canopies as specified by the user. With the one-class SVM

[9] Nowak and Crane (2002) estimate that urban street trees in the USA store 700 million tons of carbon and sequester an additional 22.8 million tons annually, equivalent to $460 million per year.

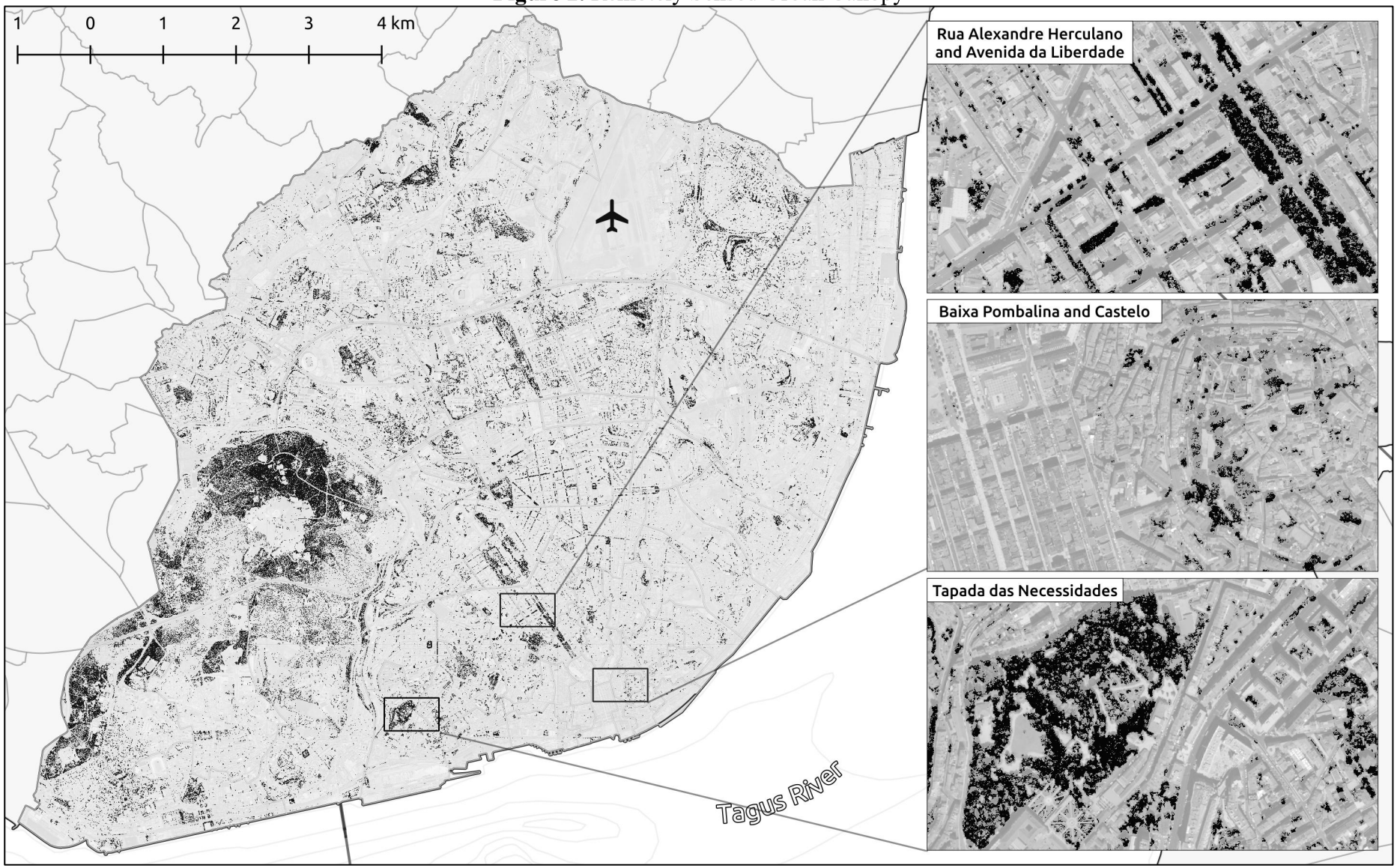
methodology, it is necessary only to provide positive training samples (i.e. a training sample composed of known tree canopies in the image) rather than negative classes, representative of everything else besides canopy. A function using these positive training samples estimates a high dimensional hypersphere which separates the cluster of training samples from all others based on the spectral information of each pixel with the largest margin possible.

A total of 506 positive training samples are collected across the region capturing a wide variety of tree canopy including individual trees planted in urban planters, clusters of trees in parks, forests and from a wide range of species with different vegetation levels (both healthy with strong green pigment or not). Visual detection at such high spatial resolution was used to classify the training samples, making use of underlying shadows to distinguish between trees and shrubs which have similar texture and comparison with satellite images from Google Earth.

For every pixel, a number of variables estimate the hypersphere including the digital number of the blue, green, red, NIR and synthetic NDVI bands. With these five bands, additional information from each is used to classify and detect tree canopies versus all other land classes. The idea is to determine what about the object in question is identifiable based on the underlying pixel information. For each band, we estimate the mean value of surrounding pixels under a three-pixel window. This determines, for example, whether a pixel classified as vegetation is surrounded by additional pixels characteristic of vegetation. Such a measure is used to help in distinguishing the boundaries and interiors of ground objects.

One important feature of tree canopies is the texture of the vegetation in comparison to other types of sparser vegetation such as grass and lawns. Again with a three-pixel window we estimate the standard deviation of the surrounding pixels. Green vegetation is detected based on the original bands and NDVI, and combined with the mean and standard deviation of surrounding pixels we narrow in and classify highly textured above ground biomass.

Figure 2. Remotely Sensed Urban Canopy

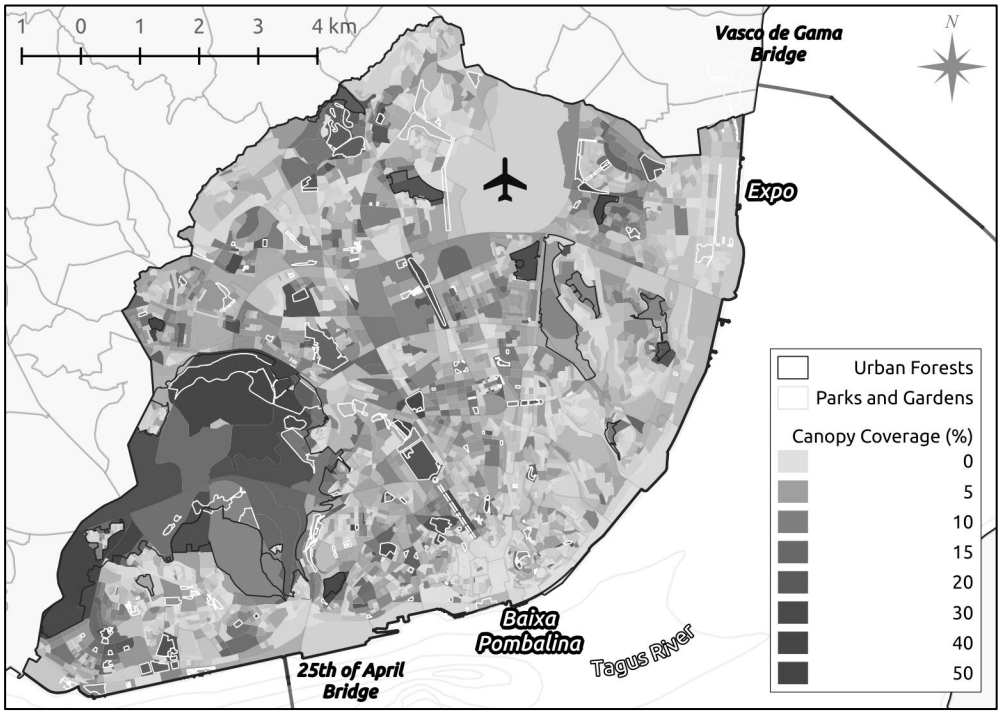


Thus there are 15 dimensions of digital numbers for each pixel in an image to use in the training and classification of the data: blue, green, red and NIR spectral bands, NDVI, and further the mean and standard deviation of surrounding pixels for each band. This methodology then takes an image pixel by pixel and, based on the pattern of digital numbers of the training sample, classifies each pixel as being either positive (i.e. the tree canopy class of interest) or negative (i.e. all other classes).

1 – 5.2. SVM Canopy Classification Results

To manage processing, images were clipped into 200 pixel by 200 pixel squares to which the algorithm was applied and the final products stitched together to create a seamless map of continuous canopy. Remotely sensed urban tree canopies are presented in figure 2. Estimates indicate that approximately 8% of Lisbon is covered directly by urban tree canopy.

Figure 3. Neighbourhood Aggregate Urban Tree Canopy



The algorithm successfully classifies both tree stands in urban forests or parks, and individual trees planted along the streets throughout the city. Thus, remote sensing of tree canopies can feasibly be implemented as means to obtain data on the urban ecology, including

all publicly and privately managed trees which may not have been accessible before under traditional in-situ methods. Figure 3 amalgamates this continuous canopy coverage to the city block (tract) level and highlights the heterogeneity in coverage. Areas in the historic CBD have significantly less tree canopy coverage, which has been shown to have important influences on the urban heat island effect and localized pollution.

1 – 5.3. Accuracy Assessment

Assessing the accuracy of classification results is important to understand whether are potential measurement errors are introduced in the econometric estimation. The accuracy of the tree canopy variable is assessed in two different areas of the city, based on a sample of 500 randomly drawn pixels from both. For each point the true land cover is determined by auxiliary images from Google Maps, Google Earth or Google Street View and compared to the output classification obtained from the SVM. It is possible to determine whether each point is classified correctly based on images from the same time frame. Figure 4 breaks down the results of the SVM algorithm cross-referenced with how that pixel appears in the auxiliary images. Through this error matrix, we can compare the relative proportion of tree canopies classified correctly as tree canopy or incorrectly as anything other (a negative class). Location and classification of tree canopy in the study area are presented in figure A1 – 1 of the appendix.

Figure 4. Remote Sensing Accuracy Assessment

Study Area A

SVM Classification	Land Truth (Aux. maps): Tree Canopy	Land Truth (Aux. maps): Non-Tree Canopy	Row Total
Tree Canopy	38	5	**43** *(88.4%)*
Non-Tree Canopy	37	420	**457** *(91.9%)*
Column Total	**75** *(50.6%)*	**425** *(98.8%)*	**500** *(91.6%)*

Study Area B

SVM Classification	Land Truth (Aux. maps): Tree Canopy	Land Truth (Aux. maps): Non-Tree Canopy	Row Total
Tree Canopy	80	20	**100** *(80.0%)*
Non-Tree Canopy	42	358	**400** *(89.5%)*
Column Total	**122** *(65.6%)*	**378** *(94.7%)*	**500** *(87.6%)*

Notes: Accuracies listed in parentheses.

The overall accuracy is the proportion of both negative and positive output classes that are correctly identified as so, represented by the highlighted diagonal entries of figure 4. The first study area has an overall accuracy of 91.6% while the second has an accuracy of 87.6%.

We can distinguish between two potential classification errors: the producer accuracy is a measure of the omission error, how many pixels on the land are not captured by the algorithm, while the user accuracy is a measure of the commission error, how many pixels from the classification are not truly what they should be classified as based on the auxiliary images. Focusing on the (positive class) tree canopy measure of interest, study area A indicates that the SVM algorithm classifies 38 positive tree canopy pixels, yet 37 actual tree canopies which should have been classified as so were not – a producer accuracy of 50.6%. However, of those pixels that the algorithm determined to be tree canopy, 88.4% are classified correctly. This pattern is seen also in study area B indicating that while we have a high accuracy related to those tree canopies correctly classified as so, there are situations in which not all tree canopy is captured, thus the variable represents an accurate lower bound on canopy coverage.

It is important to have consistently high user accuracy for the classified coverage to ensure that the tree canopy measure created is accurate based on the classification results. While this is the case for the algorithm, we see some limited producer accuracy indicated some omission errors. This suggests that not all (true) tree canopies on the ground are classified as such, and both study areas highlight why this may be the case. Lower producer accuracy is related to the heterogeneity of the sparsity of tree canopy coverage, especially for younger trees with less canopy. Pixels of these canopy types within continuous tracts of trees may be classified as negative if the canopy is not opaque or green enough to be identified by the algorithm. This problem may be rectified by considering large tracts of continuous canopy as determined by algorithm and filling in negative pixels surrounded by positive pixels.

Caution should be taken however in filling in all tracts of continuous tree canopies. Given that the measure of tree canopy is the mean taken at the neighbourhood level, we do not fill in contiguous tracts since we wish to also capture the heterogeneity of sparseness in canopy coverage across neighbourhoods.

1 – 6. Valuation of Open Spaces and Urban Greenness

The output of the machine learning canopy classification are used in a hedonic valuation of open space and urban greenness. The hedonic framework decomposes housing prices across Lisbon to determine the marginal implicit price of accessibility to open spaces, neighbourhood vegetation and tree canopy coverage.

1 – 6.1. Hedonic Pricing Data

The dataset includes a sampled cross section of 11,617 georeferenced two bedroom apartments from 2007 with listing price and relevant structural characteristics obtained from *Confidencial Imobiliário*.[10] These characteristics include area, the existence of parking, elevator, air conditioning, fireplace and whether the dwelling is new or not.

Census 2011 data is obtained from the *Instituto Nacional de Estatística* at the tract level representing the relevant sociodemographic and building characteristic variables. We control for the percentage of post-secondary educated, unemployed, residents above 65 and below 19 and population density. In terms of neighbourhood characteristics, the model includes the percentage of buildings built between different decades ranging back to 1919, the proportion of non-residential buildings and neglected or derelict buildings. Given the rich history of the city, we further control for the number of historic monuments. Accessibility to employment centers is captured by the weighted average commuting time to job centers.

[10] Although transaction prices are favored we are limited to listing prices, which may introduce a positive bias in the results. This bias is not expected to vary structurally by covariates, and estimation results remain meaningful.

Local urban amenities are obtained through the *Lisbon City Service Development Kit API* and *Câmara Municipal de Lisboa*, the municipal authority. Euclidean distance measures capture accessibility to a range of local amenities including the airport, hospitals, pharmacies, high schools, universities, train stations, shopping centers, galleries and museums, culture amenities, road infrastructure and metro stations. Ecological variables include proximity to the Tagus riverfront, a dummy variable for being located in an area of flooding risk and further we explore proximity to freeways and the impact of designated historic conservation areas. Variable descriptive statistics are located in table A1 – 1 of the appendix.

This work focuses on the impact of proximity to different heterogeneous categories of urban open spaces and overall urban greenness as measured by the remotely sensed variables. The location of open spaces in Lisbon are obtained from the municipality and include parks and gardens, urban forests (including Monsanto Forest Park), playgrounds, cemeteries or recreational football fields. Parks, urban forests and cemeteries are then classified by their size and average level of NDVI in order to capture the heterogeneity within these amenities.

1 – 6.2. Empirical Specification

Spatial dependence in housing prices may occur with pricing techniques comparing similar dwellings in the neighbourhood such that a dwellings price is determined in part by the value of neighbouring dwellings through a signaling mechanism. This implies a direct spatial relationship between property values potentially yielding biased and inefficient OLS estimates. Alternatively, omitted or unobserved variables such as outdoor maintenance expenditures or public perception of certain areas may be correlated in space through an externality mechanism, which in turn can influence local property prices. With unobservable spatial dependence OLS tends to underestimate standard errors in hedonic models, and if these unobserved amenities are correlated with neighbourhood housing prices, OLS also yields biased coefficient estimates. We therefore model and test for spatial dependence in the data using the following framework:

$$P = \alpha + S\beta_1 + B\beta_2 + D\beta_3 + O\beta_4 + G\beta_5 + I\beta_6 + \rho_{lag}\mathbb{W}P + \varepsilon$$
$$\varepsilon = \rho_{err}\mathbb{W}\varepsilon + u \;; \qquad u \sim iid(\mathbf{0}, \sigma^2 \mathrm{I_n})$$ Eq. 1 – 1

where log housing price, P, for an observation is influenced by a vector of structural characteristics of the dwelling, S, neighbourhood attributes, B, measures of accessibility to local urban amenities, D, and the variables of interest, O and G, which represent respectively the proximity to green amenities and local green coverage from one of the various measures. A number of interaction effects are included, I, between the measures of greenness and open space, environmental and local characteristics.

In particular, measures of urban greenness are interacted with neighbourhood environmental variables in the form of proximity to the riverfront and location in a high flooding zone. Vegetation, and trees specifically, may have important pollution mitigating influences and thus we control for proximity to freeways. Finally, complementary effects may be found in conservation areas where these types of amenities are jointly valued.

Further interaction effects capture influences between measures of greenness and accessibility to open spaces. In this context, we explore how prices may showcase a tradeoff between the greenness of the residents' neighbourhood and proximity to different types of amenities including parks, forests and playgrounds. Additionally, greenness and area of these open spaces are categorized to capture how different green quality and size are valued differently by the residential real estate market. Although we control for relevant observable characteristics influencing housing prices, locational fixed effects are introduced at the *freguesia* level, F, to control for any potentially remaining omitted variables.

Spatial Dependence

Spatial dependence is evaluated from test statistics related to the parameters in the above specification where neighbouring prices and, or, the error term are weighted by a spatial

weight matrix, $\mathbb{W}$, defining a neighbouring relationship.[11] From the above general specification, when $\rho_{lag} = 0$ we have a spatial error model (SEM) and with $\rho_{err} = 0$, the spatial autoregressive model (SAR). With economic intuition for the inclusion of both such spatial dependencies, we further estimate the mixed spatial autoregressive model with autoregressive disturbances (SARAR).

The spatial Lagrange multiplier (LM) test diagnostics compares the baseline OLS as the restricted model (null hypothesis) and the spatial model as the unrestricted model (alternative hypothesis) and can thus effectively consider differences between spatial and non-spatial models, and whether spatial dependence is a result of an omitted spatial lag of the dependent variable or through the error component, ultimately indicating whether the SEM, SAR or SARAR models are appropriate.

Without *freguesia* spatial fixed effects, results indicate positive spatial autocorrelation influencing both the dependant variable and the residuals. After introducing the spatial fixed effects we find that for all models the LM statistics are not significant. Given these results a standard OLS specification is estimated with spatial fixed effects at the *freguesia* level. Full spatial test results are in table A1 – 3 of the appendix.

1 – 6.3. OLS Hedonic Estimates

The characteristics (size and quality) and design of open spaces influence the way in which people use and visit these amenities, and further determine their impacts on the urban environment. Through the hedonic model we examine the value to residents of neighbourhood green cover and greenness of urban open spaces, the complementarity and substitutability between proximity to urban open spaces and a resident's neighbourhood greenness, and the interaction between urban greenness and broader ecological and neighbourhood characteristics.

[11] Two spatial weight matrices are used to represent neighbour relationships between properties: an inverse distance and inverse squared distance weight for neighbours within 500 meters. Spatial weight matrices are summarized in appendix table A1 – 2.

It is important in the development of these urban environmental measures that the scope of measurement aligns with how these characteristics are perceived by local residents. We control for neighbourhood covariates at the tract (city block) level to capture differences in the demographic and building stock of a smaller collection of buildings, on average corresponding to approximately a 100 to 200-meter radius. However, the influence of urban canopy and vegetation is likely more far reaching than a few city blocks with contiguous canopies running throughout neighbourhoods valued by residents more than lesser patches. The measurement of greenness is introduced thus at the *freguesia* level.

1 – 6.3.1. Baseline Models and Open Space Heterogeneity

Table 1 presents the baseline models. Model 1 looks at the effect of standard structural and neighbourhood variables with the traditional measures of accessibility to alternative types of open space including parks, urban forests, cemeteries, football fields and playgrounds. Models 2, 3 and 4 explore how the size and greenness of different types of urban open spaces (parks, urban forests, and cemeteries) impact housing prices. Table 1 highlights the green variables of interest with full results for all structural, neighbourhood and accessibility characteristics in table A1 – 4 of the appendix.

Based on model 1, the structural characteristics behave as expected with positive impacts on the price of a dwelling. The strongest drivers is whether the dwelling is newer, yielding a premium of 16.94%, whether the dwelling has parking, yielding a premium of 7.37%, whether the dwelling has air conditioning, yielding a premium of 15.05% and the area of the dwelling, with a price premium of 0.79% per square meter increase. On the other hand, the neighbourhood characteristic that influences housing prices the most is income levels, increasing prices approximately 0.39% per percent increase in the neighbourhood average income. Higher commuting time has a consistently negative impact on price capturing the decreasing gradient from major employment centers. This is highlighted further by the increase

in prices for dwellings closer to train stations and metro stops, approximately 0.015% increase per kilometer increase in proximity.

Table 1. Hedonic Valuation of Open Space Heterogeneity

Dep. Variable: ln(Price)	**Model 1**	***Open Space Heterogeneity***		
		Model 2	**Model 3**	**Model 4**
Intercept	4.95977**	4.50317**	4.93687**	4.49936**
	(2.21782)	*(2.22333)*	*(2.21808)*	*(2.22362)*
Open Space Accessibility				
ln(Distance to football field)	-0.01522**	-0.01464**	-0.01601**	-0.01500**
	(0.00623)	*(0.00632)*	*(0.00626)*	*(0.00634)*
ln(Distance to playground)	0.00395	0.00273	0.00631	0.00348
	(0.00620)	*(0.00653)*	*(0.00634)*	*(0.00668)*
ln(Distance to forest)	-0.01892***	-0.02656***	-0.02675***	-0.02727***
	(0.00270)	*(0.00318)*	*(0.00668)*	*(0.00702)*
ln(Distance to park)	-0.00139	0.04511***	-0.00027	0.04680***
	(0.00450)	*(0.01328)*	*(0.00481)*	*(0.01336)*
ln(Distance to cemetery)	-0.00614*	-0.02002	0.00305	-0.00384
	(0.00353)	*(0.01282)*	*(0.01089)*	*(0.01789)*
Open Space Heterogeneity				
ln(Forest size)		-0.00259		-0.0036
		(0.00269)		*(0.00279)*
ln(Park size)		0.01498***		0.01572***
		(0.00439)		*(0.00444)*
ln(Cemetery size)		0.00382		0.00337
		(0.00436)		*(0.00460)*
ln(Distance to forest) × ln(Forest size)		-0.00872***		-0.00920***
		(0.00182)		*(0.00194)*
ln(Distance to park) × ln(Park size)		0.00960***		0.00956***
		(0.00259)		*(0.00262)*
ln(Distance to cemetery) × ln(Cemetery size)		-0.00741		-0.00699
		(0.00542)		*(0.00545)*
ln(Distance to forest) × Forest NDVI			0.04997	-0.00009
			(0.03942)	*(0.04399)*
ln(Distance to park) × Park NDVI			-0.04924	-0.06486
			(0.04420)	*(0.04471)*
ln(Distance to cemetery) × Cemetery NDVI			-0.06613	-0.10653
			(0.07088)	*(0.07535)*
Freguesia F.E.	Yes	Yes	Yes	Yes
Structural Characteristics	Yes	Yes	Yes	Yes
Neighbourhood Characteristics	Yes	Yes	Yes	Yes
Accessibility Characteristics	Yes	Yes	Yes	Yes
Local Environmental Characteristics	Yes	Yes	Yes	Yes
Observations	11,617	11,617	11,617	11,617
Adjusted R^2	0.6652	0.6661	0.6653	0.6662

*Notes: ***Significance at 1 % level; **Significance at 5 % level; *Significance at 10 % level.*
Heteroskedastic consistent errors.

Model 1 yields significant and negative coefficients on three measures of open space, distance to football field, forest and cemeteries. This indicates that residential dwellings located closer to these open spaces have higher values and, for example, cemeteries increase prices in a range of 0.006% per kilometer increase in proximity. Moreover, residents value proximity to the Tagus river in the range of 0.015% per kilometer decrease in distance.

It is important to capture the heterogeneity of these amenities as determined by their size and vegetation level. In models 2, 3 and 4 we explore the impact of different types of parks, urban forests and cemeteries through interactions of distance and the size of the open space and also with the average value of NDVI for the open space. Comparing model 1 with model 4, adding both the open space size and level of greenness reduces slightly the size of the coefficients on the distance to the nearest open space of a particular type. In particular, proximity to football fields and urban forests tend to have a positive impact on housing prices with prices increasing by approximately 0.015% and 0.019% respectively for every kilometer decrease in distance. No significant marginal effect is found for proximity to parks, however these results may be driven by their heterogeneity as we obtain significant interaction effects when controlling for parks of different sizes. Dwelling prices capitalize on proximity to smaller parks while there tends to be a negative impact with proximity to the largest parks.

Increasing the size of the nearest park marginally however is valued at approximately 0.015% per square kilometer increase in park space. These results should be compared in combination with the interaction effects as an increased number of larger parks may tend to drive negative price effects. Given the variability in park sizes, from small neighbourhood parks to large landscaped gardens, this measure may capture congestion and noise effects associated with living next to the largest most visited spaces.

When controlling for the size of the nearest forest proximity is valued at approximately 0.03% per kilometer decrease in distance, larger when compared to the baseline model without accounting for size heterogeneity. This effect of proximity is compounded with additional benefits of living nearer to larger urban forests which tend to offer the most recreational facilities which are not readily available in the neighbourhood or park environment.

Another interesting result from models 2, 3 and 4 is that residents clearly value size over vegetation for given urban open spaces, as the average NDVI of those open spaces is never

significant while area is. This seems to suggest that greenness of different types of urban open spaces may then be valued for their broader effects with the urban environment rather than for their complementarity with the spaces specific use by an urban resident. In fact, the view of urban planning for open spaces has been extended from an aesthetic view to consider social impacts related to recreation, health and psychology to environmental and ecological functions.

1 – 6.3.2. Impact of Overall Urban Greenness

Using model 4 as the base, different measures of urban greenness are introduced as a proxy for the overall natural environment of a neighbourhood. The marginal impact of urban greenness on the residential real estate market is estimated and interacts the three measures of *freguesia* level greenness with the size of the *freguesia* to capture the extent of green coverage. The measures include the percentage of canopy coverage, which when interacted with the natural log of the size of the *freguesia* yields the impact for increasing relative canopy coverage per square kilometer. The mean NDVI measure interacted with size captures the extent of the average quality of greenness of a neighbourhood and finally the percentage of NDVI pixels in between different values represent the percentage of sparse or dense and lush vegetation. When interacted with size, this percentage represents the impact of increasing the relative size of these types of vegetation coverages in the neighbourhood.

Urban Greenness

Table 2 introduces measures of urban greenness in three manners: firstly by percentage of canopy coverage, secondly by mean NDVI, and thirdly by percentage of NDVI representative of sparse or lush vegetation. As we move to including these measures, the explanatory power of the OLS model specifications 5 through 13 improve. We see a reduction in the sum of squared error and the AIC value which indicate better fit by controlling for additional elements of urban open spaces which are typically excluded from hedonic analyses estimating impacts on property values.

Table 2. Impacts of Urban Greenness and Open Space

Dep. Variable: ln(Price)	***Urban Tree Canopy Coverage***			***Mean NDVI***			***NDVI by Vegetation Type***		
	Model 5	**Model 6**	**Model 7**	**Model 8**	**Model 9**	**Model 10**	**Model 11**	**Model 12**	**Model 13**
Intercept	4.55710**	4.41484**	4.92013**	4.62164**	4.36259*	4.89187**	4.65924**	5.78615***	5.30349**
	(2.22373)	*(2.22520)*	*(2.22843)*	*(2.22522)*	*(2.23045)*	*(2.22751)*	*(2.22554)*	*(2.22611)*	*(2.23374)*
Neighbourhood and Environmental Characteristics									
ln(Distance to Tagus)	-0.01114*	-0.01244**	-0.00763	-0.01068*	-0.01580**	-0.00847	-0.01141*	0.00193	0.00895
	(0.00582)	*(0.00606)*	*(0.00603)*	*(0.00582)*	*(0.00615)*	*(0.00609)*	*(0.00583)*	*(0.00716)*	*(0.00753)*
ln(Distance to freeway)	0.01438*	0.01545**	0.02774***	0.01508**	0.01390*	0.02164*	0.01383*	0.01309*	0.04429***
	(0.00746)	*(0.00759)*	*(0.00971)*	*(0.00746)*	*(0.00773)*	*(0.01149)*	*(0.00753)*	*(0.00763)*	*(0.01300)*
Flood risk dummy	-0.02579**	-0.02017	-0.06991***	-0.02551**	-0.03261**	-0.05494***	-0.02570**	-0.02999**	-0.08334***
	(0.01298)	*(0.01322)*	*(0.02115)*	*(0.01299)*	*(0.01324)*	*(0.01808)*	*(0.01300)*	*(0.01316)*	*(0.02479)*
Located in Conservation Area	-0.01511	-0.01702	-0.07279***	-0.01515	-0.01782	-0.09139***	-0.01705	0.01464	-0.06881**
	(0.01362)	*(0.01421)*	*(0.02088)*	*(0.01363)*	*(0.01383)*	*(0.02638)*	*(0.01371)*	*(0.01470)*	*(0.02726)*
View of the Tagus river	0.06451***	0.06391***	0.06378***	0.06454***	0.06424***	0.06384***	0.06450***	0.06384***	0.06193***
	(0.00864)	*(0.00864)*	*(0.00863)*	*(0.00864)*	*(0.00864)*	*(0.00863)*	*(0.00864)*	*(0.00861)*	*(0.00863)*
Open Space Accessibility									
ln(Distance to football field)	-0.01581**	-0.01639**	-0.01611**	-0.01547**	-0.01893***	-0.01814***	-0.01596**	-0.02034***	-0.01704***
	(0.00636)	*(0.00643)*	*(0.00642)*	*(0.00635)*	*(0.00651)*	*(0.00643)*	*(0.00636)*	*(0.00646)*	*(0.00641)*
ln(Distance to playground)	0.00409	-0.01099	0.00306	0.00303	0.00273	0.00169	0.00423	-0.00461	0.0037
	(0.00669)	*(0.00793)*	*(0.00670)*	*(0.00669)*	*(0.00716)*	*(0.00671)*	*(0.00669)*	*(0.01136)*	*(0.00677)*
ln(Distance to forest)	-0.02777***	-0.03119***	-0.01997***	-0.02723***	-0.02988***	-0.02227***	-0.02755***	0.02276	-0.02455***
	(0.00703)	*(0.01142)*	*(0.00729)*	*(0.00702)*	*(0.01123)*	*(0.00724)*	*(0.00703)*	*(0.01471)*	*(0.00729)*
ln(Distance to park)	0.04753***	0.05519***	0.03834***	0.04687***	0.06436***	0.03227**	0.04664***	0.02165	0.02767**
	(0.01337)	*(0.01467)*	*(0.01356)*	*(0.01336)*	*(0.01550)*	*(0.01399)*	*(0.01338)*	*(0.01587)*	*(0.01386)*
ln(Distance to cemetery)	-0.00073	-0.00512	0.01273	-0.00019	-0.01061	0.00427	0.00156	-0.00894	0.00975
	(0.01799)	*(0.01808)*	*(0.01954)*	*(0.01808)*	*(0.01880)*	*(0.01981)*	*(0.01811)*	*(0.01925)*	*(0.02028)*
Open Space Heterogeneity									
ln(Forest size)	-0.00404	-0.00597**	-0.00356	-0.00383	-0.00313	-0.00305	-0.00424	-0.00740**	-0.00486*
	(0.00280)	*(0.00287)*	*(0.00280)*	*(0.00279)*	*(0.00284)*	*(0.00283)*	*(0.00281)*	*(0.00288)*	*(0.00286)*
ln(Park size)	0.01504***	0.01530***	0.01339***	0.01521***	0.01628***	0.01316***	0.01486***	0.02181***	0.01543***
	(0.00446)	*(0.00452)*	*(0.00449)*	*(0.00445)*	*(0.00459)*	*(0.00458)*	*(0.00446)*	*(0.00484)*	*(0.00452)*
ln(Cemetery size)	0.00381	0.0033	0.00229	0.00371	0.00326	0.00183	0.00416	0.00185	0.00167
	(0.00461)	*(0.00464)*	*(0.00465)*	*(0.00460)*	*(0.00468)*	*(0.00473)*	*(0.00462)*	*(0.00468)*	*(0.00469)*
ln(Distance to forest) × ln(Forest size)	-0.00927***	-0.00987***	-0.00989***	-0.00914***	-0.00882***	-0.00915***	-0.00924***	-0.0017	-0.01195***
	(0.00194)	*(0.00212)*	*(0.00196)*	*(0.00194)*	*(0.00201)*	*(0.00199)*	*(0.00194)*	*(0.00282)*	*(0.00207)*
ln(Distance to park) × ln(Park size)	0.00958***	0.00996***	0.00762***	0.00956***	0.01058***	0.00708***	0.00936***	0.01188***	0.00774***
	(0.00262)	*(0.00265)*	*(0.00265)*	*(0.00262)*	*(0.00272)*	*(0.00272)*	*(0.00263)*	*(0.00280)*	*(0.00271)*
ln(Distance to cemetery) × ln(Cemetery size)	-0.0058	-0.00747	-0.00396	-0.00579	-0.00666	-0.00529	-0.00568	-0.00489	-0.00395
	(0.00550)	*(0.00552)*	*(0.00567)*	*(0.00552)*	*(0.00558)*	*(0.00566)*	*(0.00553)*	*(0.00591)*	*(0.00578)*
ln(Distance to forest) × Forest NDVI	0.00578	0.01857	-0.0557	0.00276	0.00962	-0.03595	0.0044	-0.005	-0.03998
	(0.04413)	*(0.05083)*	*(0.04647)*	*(0.04404)*	*(0.05085)*	*(0.04617)*	*(0.04409)*	*(0.04555)*	*(0.04640)*

ln(Distance to park) × Park NDVI	-0.07294 *(0.04498)*	-0.0593 *(0.04526)*	-0.09950** *(0.04628)*	-0.07358 *(0.04514)*	-0.07292 *(0.04581)*	-0.06579 *(0.04671)*	-0.07725* *(0.04517)*	-0.08751* *(0.04772)*	-0.06237 *(0.04814)*
ln(Distance to cemetery) × Cemetery NDVI	-0.10822 *(0.07535)*	-0.10045 *(0.07595)*	-0.17059** *(0.08095)*	-0.11079 *(0.07540)*	-0.05305 *(0.07935)*	-0.13696* *(0.08324)*	-0.11922 *(0.07577)*	-0.02961 *(0.07816)*	-0.14569* *(0.08540)*
Measures of Urban Greenness (Measured at the Freguesia Scale)									
% Canopy Coverage × ln(*Freguesia* size)	0.14721* *(0.08935)*	0.19216** *(0.09108)*	0.20385* *(0.10630)*						
Mean NDVI × ln(*Freguesia* size)				0.10897 *(0.07728)*	0.04931 *(0.08449)*	0.16251** *(0.08117)*			
% NDVI [0.2, 0.5] Coverage × ln(*Freguesia* size)							0.0338 *(0.03826)*	0.05355 *(0.04116)*	0.09129* *(0.04775)*
% NDVI [0.5, 1] Coverage × ln(*Freguesia* size)							0.53408 *(0.47173)*	0.06843 *(0.48321)*	-0.19199 *(0.52411)*
Freguesia F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Structural Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Accessibility Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Local Environmental Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Open Space Interactions	No	Yes	No	No	Yes	No	No	Yes	No
Environmental Interactions	No	No	Yes	No	No	Yes	No	No	Yes
AIC	-3056.10	-3060.90	-3070.70	-3055.40	-3058.20	-3064.80	-3055.40	-3120.30	-3097.60
SSE	512.59	512.03	511.60	512.62	512.15	511.86	512.53	508.98	509.97
Observations	11,617	11,617	11,617	11,617	11,617	11,617	11,617	11,617	11,617
Adjusted R^2	0.6663	0.6665	0.6668	0.6662	0.6664	0.6666	0.6663	0.6683	0.6677

*Notes: ***Significance at 1 % level; **Significance at 5 % level; *Significance at 10 % level.*
Heteroskedastic consistent errors

Overall, urban greenness has a positive effect on housing prices with results from table 2 revealing that residents value to be in neighbourhoods with higher canopy coverage. While NDVI is unitless, it nonetheless indicates that dwellings have higher prices in neighbourhoods with higher levels of vegetation and is significant when interaction effects are included.

Additionally, when decomposing the NDVI by levels, there is a positive amenity effect, yet much weaker than the effect of canopy coverage, for sparse vegetation coverage, which includes lawns and bushes. In terms of coverage of tree canopy and of sparse vegetation, this measure indicates that as we increase the relative size of either measure by a square kilometer, housing prices would increase, on average, by approximately 0.203% for increased tree canopy or 0.091% for increased sparse vegetation, after controlling for environmental interactions. For an average dwelling of €203,483 this corresponds to an increased price of slightly over €400 per dwelling for more tree canopy or €185 for sparser vegetation.

It should also be highlighted that the tree canopy coverage measure of greenness seems to perform better in general compare to the other measures of overall greenness. For all three specifications (models 5 to 8) the coefficients associated with the pure effects are always positive and significant. In contrast, the coefficients associated with mean NDVI and with NDVI by type are only positively significant when we introduce interaction variables between the overall measures of greenness and other elements of the urban environment such as flood risk or historic conservation zones. Moreover, the coefficients associated with distance to alternative urban spaces are also lower compared to model 4. This suggests that model 4 still tends to overestimate the value residents place on proximity to an urban open space, even when we control for urban open space heterogeneity. Thus, overall greenness of a neighbourhood seems to be an important green element that should also be taken into account when examining the impact of alternative urban open spaces and sizes of open space in housing prices.

1 – 6.3.3. Interactions of Urban Greenness with other Urban Variables

Table 3 presents the interaction effects between open space accessibility and measures of greenness to explore the complementarity and substitutability between these two types of green amenities (full results available in table A1 – 5 of the appendix). This is accompanied by models on the interaction effects of urban greenness and broader neighbourhood environmental characteristics or historic zones. Though open spaces may provide recreational and relaxation opportunities to residents, the results indicate that such spaces may further have a wide range of influences on various facets of the urban environment such as controlling the risk of flooding and accentuating neighbourhood aesthetics through interactions with historic qualities.

Complementary versus Substitutability with Other Types of Urban Open Space

For each measure of urban greenness, we explore how proximity to different urban open spaces are complemented or substituted by the overall greenness of a resident's own immediate area (represented by a collection of census tracts as opposed to the *freguesia* level). From model 6, higher residential canopy coverage appears to compensate for living at greater distance from a playground.

On the other hand, higher canopy coverage may complement proximity to an urban park in some manner, whereby households who like trees choose zones of the *freguesia* that have lots of trees and are located closer to an urban park. This result may be justified because urban parks may offer more diverse vegetation in the form of grassland and large trees. Yet, when looking at the coefficient on the distance to a particular type of open space, we see that households prefer nevertheless to be closer to smaller rather than larger urban parks. The same qualitative conclusions can also be drawn if we consider the interactions between distance to an urban open space and the mean NDVI of the collection of census tracts (model 9).

Table 3. Complementarity and Substitutability of Local Greenness

Dep. Variable: ln(Price)	**Urban Tree Canopy Coverage**			**Mean NDVI**			**NDVI by Vegetation Type**		
	Model 5	**Model 6**	**Model 7**	**Model 8**	**Model 9**	**Model 10**	**Model 11**	**Model 12**	**Model 13**
Intercept	4.55710**	4.41484**	4.92013**	4.62164**	4.36259*	4.89187**	4.65924**	5.78615***	5.30349**
	(2.22373)	*(2.22520)*	*(2.22843)*	*(2.22522)*	*(2.23045)*	*(2.22751)*	*(2.22554)*	*(2.22611)*	*(2.23374)*
	Greenness†: % Canopy Coverage			**Greenness†: Mean NDVI Value**			**Greenness†: % NDVI in [0.2, 0.5]**		
Greenness × ln(Distance to forest)		0.02695			0.02592			-0.30215***	
		(0.11200)			*(0.08138)*			*(0.06937)*	
Greenness × ln(Distance to park)		-0.24100*			-0.25484***			0.22178***	
		(0.13648)			*(0.08606)*			*(0.08477)*	
Greenness × ln(Distance to playground)		0.55430***			0.15198*			0.07369	
		(0.15777)			*(0.08605)*			*(0.11518)*	
Greenness × ln(Distance to freeway)			-0.24641			-0.08031			-0.21557**
			(0.15948)			*(0.14477)*			*(0.09100)*
Greenness × Flood risk dummy			0.97734**			0.46593*			0.21301
			(0.40832)			*(0.26270)*			*(0.23989)*
Greenness × Conservation zone dummy			0.80115***			0.88043***			-0.10131
			(0.28112)			*(0.29276)*			*(0.15086)*
Greenness × ln(Distance to Tagus)		0.08932	-0.0598		-0.10774	-0.08709		-0.17067**	-0.35203***
		(0.11203)	*(0.09418)*		*(0.06609)*	*(0.06262)*		*(0.07209)*	*(0.06188)*
							Greenness†: % NDVI > 0.5		
Greenness × ln(Distance to forest)								0.26095	
								(0.48057)	
Greenness × ln(Distance to park)								-0.49237	
								(0.44844)	
Greenness × ln(Distance to playground)								0.78442	
								(0.67818)	
Greenness × ln(Distance to freeway)									-0.22102
									(0.63224)
Greenness × Flood risk dummy									1.47264
									(1.49857)
Greenness × Conservation zone dummy									3.29077***
									(1.09931)
Greenness × ln(Distance to Tagus)								0.3716	0.98485***
								(0.37735)	*(0.29024)*
Table 2 F.E. and Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AIC	-3056.10	-3060.90	-3070.70	-3055.40	-3058.20	-3064.80	-3055.40	-3120.30	-3097.60
SSE	512.59	512.03	511.60	512.62	512.15	511.86	512.53	508.98	509.97
Observations	11,617	11,617	11,617	11,617	11,617	11,617	11,617	11,617	11,617
Adjusted R^2	0.6663	0.6665	0.6668	0.6662	0.6664	0.6666	0.6663	0.6683	0.6677

*Notes: ***Significance at 1 % level; **Significance at 5 % level; *Significance at 10 % level.*

† Local Greenness measured at a scale including the collection of census tracts in which the dwelling is located.

Heteroskedastic consistent errors

Further, results from model 12 suggest that if a resident's immediate area of residence is relatively more covered by sparse vegetation such as lawns and shrubs, then the trade-offs households are willing to make are different from above, stemming from the different measures of greenness. In particular, greater sparse vegetation coverage can substitute for living further away from urban parks (since it is likely that the same type of vegetation is found in both places) but it is complementary with proximity to an urban forest. This latter effect may once more possibly be explained because the type of vegetation found in both locations differ and residents may enjoy living in an area of sparse vegetation with proximity to urban forests where there are different types of vegetation coverages and recreational activities compared to the neighbourhood where this is lacking.

Finally, when looking at model 13, proximity to the Tagus river is more valuable in areas with a high relative proportion of sparse vegetation, while the opposite is true for areas with high proportions of lush vegetation.

Influences of Urban Greenness on the Urban Environment

When examining the interaction variables of urban greenness with the urban environment (models 7, 10 and 13) open spaces have further significant positive effects in *freguesias* with high flood risk. Therefore, both levels of canopy coverage or urban greenness provide valuable amenity services in mitigating of flood, a common occurrence during the winter months. The residential real estate market successfully captures these environmental spillover effect that increased canopy coverage has in flooding areas.

Moreover, increasing the *freguesia* relative proportion of tree canopy coverage or urban greenness in general in historic (conservation) areas of the city has a significant positive compounding effect and residents value the aesthetics and characteristics of their neighbourhood for the conjoined effect of historic and green amenities. This then seems to suggest that historic amenities and urban greenness are complementary in housing prices.

1 – 6.4. Policy Implications

The results provide support for the claim that urban green amenities lead to higher housing prices suggesting that urban residents are willing to trade off higher prices for nonmarket amenities such as green spaces. Moreover, the interaction influences of urban greenness on the local environment highlight the importance of understanding the measure of urban greenness being used and the respective scale.

From a policy perspective, these results indicate the importance of considering not only the heterogeneity of different types of open spaces but also the heterogeneous interaction effects that these urban green spaces have on the local neighbourhood. Urban greening policies may benefit from focusing on the overall greening of a neighbourhood through increased tree canopy or vegetation as opposed to the development of new open spaces which require land, sometimes quite scarce and extremely expensive in central areas of historic cities. However, these green amenities, when bundled with natural or historic amenities, can further enhance the pleasantness and competitiveness of a city and boost tourism. Urban green amenities contribute to neighbourhood aesthetics and appeal and provide essential services that are critical to both urban ecological functioning and integrity, all of which influence human quality of life and health (Kahn and Walsh 2015, Zupancic et al. 2015).

As urban quality of life and cross city competitiveness improves, property values rise. By raising city property values, private sector investments in upgrading buildings and better restaurants and retail shopping can also be triggered. In addition, property tax revenues rise and local governments can finance more urban projects that further enhance local quality of life and competitiveness for residents, workers or firms. This in turn creates incentives for cities, especially modern consumer cities, to invest in green infrastructure, expand their green

space resources near waterfronts, and adapt obsolete or underused infrastructure such as rail corridors into green infrastructures for walking and biking, exercise, and social interaction.[12]

A good example is the Lisbon riverfront renewal focusing on beautifying and promoting tourism around the waterfront. In 2008, the municipality developed the Riverfront General Plan that has guided public projects along the 19 km Tagus waterfront over the last decade. Alongside this plan, some port and industrial coastal areas that were abandoned or obsolete were also released to public use and converted into urban parks and recreational areas associated to boating, cycling or pedestrian use as well as landscaping and paths connecting the river and city. The greening of the riverfront has been particular important in the central section where, due to the dense medieval urban structure, this area of the city had almost no green space. One interesting feature of Lisbon's riverfront is nevertheless the lack of dense construction. The exception has been in the renewed Expo area, far from the historic center, where some residential construction has been allowed near the water line. Lisbon's waterfront has become one of the most popular destinations in the city, attracting millions of people each year. This type of urban renewal is now also taking place in nearby neighbouring coastal municipalities such as Oeiras and Cascais.

Yet, such urban green space strategies may have contradictory results (Wolch et al. 2014, Kahn and Walsh 2015). If they are successful from the perspective of urban residents and businesses, they may ultimately exclude those whose need for access is most critical. Many studies further reveal that the distribution of urban green space and tree canopy often disproportionately benefits more affluent communities (Danford et al. 2014, Wolch et al. 2014, Watkins et al. 2016).

[12] Many coastal cities around the world have evolved from being producer cities to becoming consumer cities (Glaeser et al. 2001). Lisbon is an example of a coastal city that used to be a major hub of production and transportation that has evolved into a green area hub. Its limited land supply, due to its topography and building height restrictions, has also contributed to the high residential prices and to the development of endogenous local attributes that have further enhanced local amenities and the attractiveness of the city.

In many cities, including Lisbon, low income neighbourhoods often also have relatively poor access to safe and well-maintained parks and other open spaces. Thus, by simultaneously making older and typically low-income or industrial areas of existing cities more livable and attractive and therefore targets for new and more upscale development, as was the case in Expo, urban greening projects can set off rounds development and potential gentrifying pressures.

Since higher-quality neighbourhoods and cities require rent premiums, this implies that such areas self-select a subset of firms and households, usually highly educated and wealthy individuals, to locate there. This market pressure in turn can substantially alter housing opportunities and the retail infrastructure that supports lower income communities, forcing poor residents out of improved neighbourhoods or city areas, only to resettle in neighbourhoods or communities with worse environmental quality but more affordable (Zukin et al. 2009).

It is then important that urban planners also develop urban green space strategies that protect social as well as environmental sustainability. One possible solution is green space interventions that are small-scale and scattered rather than large space projects that geographically concentrate resources and with limited localized benefits. These types of greening interventions such as the planting of street trees, flower beds, pocket parks and small neighbourhood playgrounds can be implemented in such a way as to distribute any pressures or influences of green amenities more consistently across a municipality, rather than having the impacts concentrated around larger dedicated green interventions.

Moreover, these types of green initiatives can be complemented with anti-gentrification measures that include the provision of affordable housing so that existing residents may have a stake in an improving neighbourhood. While the pressures potentially exerted from the effects of urban green strategies is certainly an important topic that deserves further investigation, it is beyond the scope of this thesis.

1 – 7. Conclusions

This chapter highlights how machine learning remote sensing of aerial images can generate measures to incorporate into broader economic valuations of urban environmental amenities. Specifically the remote sensing of tree canopy coverage and urban greenness is used in a hedonic framework to examine how open spaces and canopy coverage influence the local housing market via direct and interactive influences. Results indicate it is important to consider how enviornmental variables are measured and introduced into the empirical specification. These variables should be constructed so as to best reflect how they are perceived by residents and further, especially given the consideration of environmental characteristics, how they interact with other aspects of the built and natural environment.

Through the use of interaction effects, the model estimates can be used to explore how green amenities relate to the broader environment and accesibility to open space. While housing prices tend to capitalize proximity to smaller neighbourhood parks, larger forests are prefered. Results indicate that different measures of green amenities provide consistent positive effects on dwelling prices. Estimates range that for a square kilometer increase in relative tree canopy coverage of a neighbourhood, the impact would average around €400 per dwelling. This effect is strong when compared to the effect due to proximity to various open spaces.

By valuing urban greenness and canopy coverage, there is increased discussion to be had regarding the use of neighbourhood greening through increased vegetation and canopy as an alternative to the provision of large areas of green space. Further, the heterogeneity of the use value of these amenities is important in considering the types of recreational facilities provided by parks and forests.

With increasingly detailed data and computational power, remote sensing and geospatial data continue to provide viable ways to explore urban policy analyses and assess the distribution and access of urban amenities such as tree canopy and urban green spaces. High

resolution aerial photography has shown itself to provide accurate data regarding the location of tree canopies, distinguishing them from all other land uses such as roads and buildings. Making use of this data we are able to extend the standard hedonic analysis to include tree canopy vegetation, an important urban environmental variable traditionally overlooked.

1 – References

Anderson, L. M., and H. K. Cordell. 1988. “Influence of Trees on Residential Property Values in Athens, Georgia (U.S.A.): A Survey Based on Actual Sales Prices.” *Landscape and Urban Planning* 15: 153-164.

Anderson, Soren T., and Sarah E. West. 2006. “Open space, residential property values, and spatial context.” *Regional Science and Urban Economics* 36: 773-789.

Câmara Municipal de Lisboa. *Biodiversidade na cidade de Lisboa: Uma Estratégia para 2020*. 2015a.

Câmara Municipal de Lisboa. *Plano Diretor Municipal*. 2015b. http://www.cm-lisboa.pt/viver/urbanismo/planeamento-urbano/plano-diretor-municipal.

Cho, Seong-Hoon, J. M. Bowker, and William M. Park. 2006. "Measuring the Contribution of Water and Green Space Amenities to Housing Values: An Application and Comparison of Spatially Weighted Hedonic Models." *Journal of Agricultural and Resource Economics* 31 (3): 485-507.

Cho, Seong-Hoon, Neelam C. Poudyal, and Roland K. Roberts. 2008. "Spatial analysis of the amenity value of green open space." *Ecological Economics* 66: 403-416.

Conway, Delores, Christina Q. Li, Jennifer Wolch, Christopher Kahle, and Michael Jerrett. 2010. “A Spatial Autocorrelation Approach for Examining the Effects of Urban Greenspace on Residential Property Values.” *Journal of Real Estate Finance and Economics* 41 (2): 150-169.

Czembrowski, Piotr, and Jakub Kronenberg. 2016. “Hedonic pricing and different urban green space types and sizes: Insights into the discussion on valuing ecosystem services.” *Landscape and Urban Planning* 146: 11-19.

Danford, Rachel S., Chingwen Cheng, Michael W. Strohbach, Robert Ryan, Craig Nicolson, Paige S. Warren. 2014. “What does it take to achieve equitable urban tree canopy distribution? A Boston case study.” *Cities and the Environment* 7 (1): Article 2.

Des Rosiers, Francois, Marius Thériault, Yan Kestens and Paul Villeneuve. 2002. "Landscaping and House Values: An Empirical Investigation" *Journal of Real Estate Research* 23 (1-2): 139-161.

Dombrow, Jonathan, Mauricio Rodriguez, and C. F. Sirmans. 2000. “The Market Value of Mature Trees in a Single-Family Housing Market.” *Appraisal Journal* 68: 39-43.

Donovan, Geoffrey H., and David T. Butry. 2010. “Trees in the city: Valuing street trees in Portland, Oregon.” *Landscape and Urban Planning* 94: 77-83.

Geoghegan, Jacqueline. 2002. “The value of open spaces in residential land use.” *Land Use Policy* 19: 91-98.

Glaeser, Edward L., Jed Kolko, and Albert Saiz. 2001. "Consumer city." *Journal of Economic Geography* 1: 27–50.

Iizuka, Kotaro, and Ryutaro Tateishi. 2015. "Estimation of CO2 Sequestration by the Forests in Japan by Discriminating Precise Tree Age Category using Remote Sensing Techniques." *Remote Sensing* 7: 15082-15113.

Irwin, Elena G. 2002. "The Effects of Open Space on Residential Property Values." *Land Economics* 78 (4): 465-480.

Kahn, Mathew E., and Randall Walsh. ‘Chapter 7 - Cities and the Environment,’ in *Handbook of Regional and Urban Economics Vol. 5*, ed. Gilles Duranton, J. Vernon Henderson, and William C. Strange (Elsevier, 2015), 405-465.

Karlson, Martin, Madelene Ostwald, Reese. Heather, Josias Sanou, Boalidioa Tankoano, and Eskil Mattsson. 2015. "Mapping Tree Canopy Cover and Aboveground Biomass in Sudano-Sahelian Woodlands Using Landsat 8 and Random Forest." *Remote Sensing* 7: 10017-10041.

Kaufman, Dennis A., and Norman R. Cloutier. 2006. "The Impact of Small Brownfields and Greenspaces on Residential Property Values." *The Journal of Real Estate Finance and Economics* 33 (1): 19-30.

Kestens, Yan, Marius Thériault, and Francois Des Rosiers. 2004. "The Impact of Surrounding Land Use and Vegetation on Single-Family House Prices." *Environment and Planning* 31 (4): 539-567.

Li, Dan, Yinghai Ke, Huili Gong, and Xiaojuan Li. 2015. "Object-Based Urban Tree Species Classification Using Bi-Temporal WorldView-2 and WorldView-3 Images." *Remote Sensing* 12 (7): 16917-16937.

McConnell, Virginia, and Margaret Walls. 2005. The Value of Open Space: Evidence from Studies of Nonmarket Benefits. Report, Washington: Resources for the Future.

Morancho, Aurelia Bengochea. 2003. “A hedonic valuation of urban green areas.” *Landscape and Urban Planning* 66 (1): 35-41.

Nilsson, Pia. 2014. "Natural amenities in urban space - A geographically weighted regression approach." *Landscape and Urban Planning* 121: 45-54.

Nowak, David J., and Daniel E. Crane. 2002. “Carbon storage and sequestration by urban trees in the USA.” *Environmental Pollution* 116: 381-389.

Parmehr, Ebadat G., Marco Amati, Elizabeth J. Taylor and Stephen J. Livesley. 2016. “Estimation of urban tree canopy cover using random point sampling and remote sensing methods.” *Urban Forestry & Urban Greening* 20: 160-171.

Sander, Heather A., and Stephen Polasky. 2009. "The value of views and open space: Estimates from a hedonic pricing model for Ramsey County, Minnesota, USA." *Land Use Policy* 26 (3): 837-845.

Sander, Heather A., Stephen Polasky, and Robert G. Haight. 2010. “The value of urban tree cover: A hedonic property price model in Ramsey and Dakota Counties, Minnesota, USA.” *Ecological Economics* 69 (8): 1646-1656.

Schölkopf, Bernhard, John C. Platt, John Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. 2001. "Estimating the support of a high-dimensional distribution." Neural Computation 13 (7): 1443-1471.

Troy, Austin, and J. Morgan Grove. 2008. "Property values, parks, and crime: A hedonic analysis in Baltimore, MD." *Landscape and Urban Planning* 87: 233-245.

Troy, Austin, J. Morgan Grove, and Jarlath O'Neil-Dunne. 2012. "The relationship between tree canopy and crime rates across an urban-rural gradient in the greater Baltimore region." *Landscape and Urban Planning* 106: 262-270.

Tyrväinen, Liisa, and Antti Miettinen. 2000. "Property Prices and Urban Forest Amenities." *Journal of Environmental Economics and Management* 39 (2): 205-223.

Voicu, Ioan, and Vicki Been. 2008. "The Effect of Community Gardens on Neighbouring Property Values." *Real Estate Economics* 36 (2): 241-283.

Waltert, Fabian, and Felix Schläpfer. 2010. "Landscape amenities and local development: A review of migration, regional economic and hedonic pricing studies." *Ecological Economics* 70 (2): 141-152.

Watkins, Shannon Lea, Sarah K. Mincey, Jess Vogt, and Sean P. Sweeney. 2016. "Is Planting Equitable? An Examination of the Spatial Distribution of Nonprofit Urban Tree-Planting Programs by Canopy Cover, Income, Race, and Ethnicity." Environment and Behavior 49 (4): 452-482.

Watt, Fiona, and Bram Gunther. "Tree Cover % - How Does Your City Measure Up?" deeproot. April 25, 2010. http://www.deeproot.com/blog/blog-entries/tree-cover-how-does-your-city-measure-up (accessed September, 2016).

Wolch, Jennifer R., Jason Byrne, Joshua P. Newell. 2014. "Urban green space, public health, and environmental justice: the challenge of making cities 'just green enough'." *Landscape and Urban Planning* 125: 234–244.

Wolfe, Mary K., and Jeremy Mennis. 2012. "Does vegetation encourage or suppress urban crime? Evidence from Philadelphia, PA." *Landscape and Urban Planning* 108: 112-122.

Zukin, Sharon, Valerie Trujillo, Peter Frase, Danielle Jackson, Time Recuber, and Abraham Walker. 2009. "New retail capital and neighbourhood change: Boutiques and gentrification in New York City?" *City and Community* 8 (1): 47–64.

Zupancic, Tara, Claire Westmacott, and Mike Bulthuis. 2015. The impact of green space on heat and air pollution in urban communities: A meta-narrative systematic review. Vancouver: David Suzuki Foundation.

1 – Appendix

Figure A1 – 1. Accuracy Assessment Study Area
Area A: *Eduardo VII Park* and **Area B:** *Praça de Espanha*

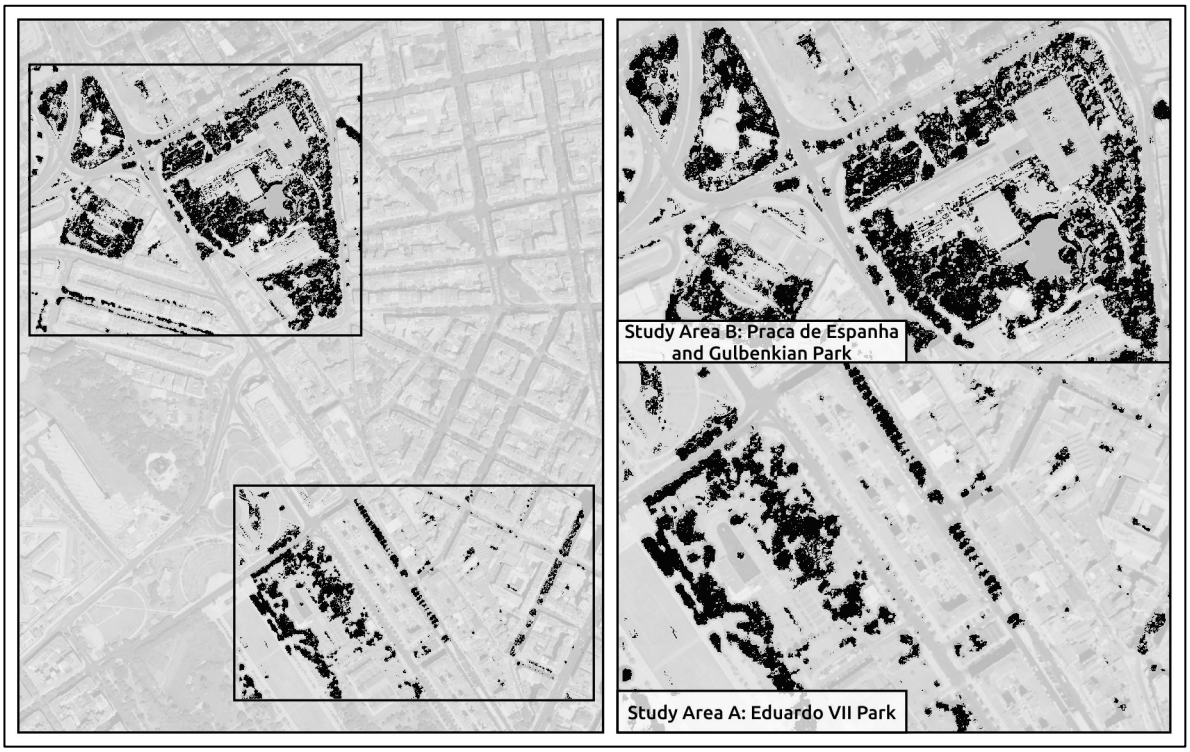


Table A1 – 1. Descriptive Statistics

	N	Mean	St. Dev	Min	Max	Unit
Price (€)	11,617	203,483	84,391	35,000	1,100,000	Euro
Square meters	11,617	85.04	23.29	25	240	Sq. m.
New dummy	11,617	0.180	0.384	0	1	Dummy
Pool dummy	11,617	0.008	0.086	0	1	Dummy
Parking dummy	11,617	0.114	0.318	0	1	Dummy
Fireplace dummy	11,617	0.025	0.157	0	1	Dummy
Double windows dummy	11,617	0.207	0.405	0	1	Dummy
Air conditioning dummy	11,617	0.119	0.324	0	1	Dummy
Elevator dummy	11,617	0.228	0.420	0	1	Dummy
% Non-residential buildings	11,617	0.025	0.116	0.00	1.00	Percent
Neglected Buildings in 200 m	11,617	6.426	6.182	0.00	45.00	Count
% Buildings built pre 1919	11,617	0.104	0.246	0.00	1.00	Percent
% Buildings built 1919-1945	11,617	0.169	0.279	0.00	1.00	Percent
% Buildings built 1946-1960	11,617	0.221	0.336	0.00	1.00	Percent
% Buildings built 1961-1970	11,617	0.127	0.266	0.00	1.00	Percent
% Buildings built 1971-1980	11,617	0.125	0.300	0.00	1.00	Percent
% Buildings built 1981-1990	11,617	0.023	0.117	0.00	1.00	Percent
% Buildings built 1991-1995	11,617	0.044	0.160	0.00	1.00	Percent
% Buildings built 1996-2000	11,617	0.108	0.292	0.00	1.00	Percent
% Buildings built 2000-2005	11,617	0.036	0.152	0.00	1.00	Percent
Neighbourhood Mean Income	11,617	27,840	8,657	10,619	49,968	Euro
Population density	11,617	0.015	0.013	0.00	0.08	Persons/ Sq. m.
% Higher educated	11,617	0.268	0.184	0.00	1.00	Percent
% Population over 65	11,617	0.280	0.165	0.00	1.00	Percent
% Population under 19	11,617	0.159	0.093	0.00	0.80	Percent
Average commute time	11,617	21.87	10.20	0.00	70.67	Minutes

# Neighbourhood Monuments	11,617	4.593	4.355	0.00	15.00	Count
Distance to airport	11,617	4.489	1.795	0.59	10.65	Kilometer
Distance to cultural amenity	11,617	0.639	0.632	0.02	4.17	Kilometer
Distance to art gallery	11,617	0.950	0.662	0.01	3.45	Kilometer
Distance to museum	11,617	0.673	0.402	0.01	2.48	Kilometer
Distance to pharmacy	11,617	0.168	0.151	0.01	1.25	Kilometer
Distance to hospital	11,617	0.938	0.577	0.01	3.19	Kilometer
Distance to mall	11,617	0.640	0.695	0.01	5.97	Kilometer
Distance to supermarket	11,617	0.208	0.162	0.01	1.23	Kilometer
Distance to university	11,617	0.621	0.406	0.03	2.50	Kilometer
Distance to high school	11,617	0.622	0.301	0.04	2.38	Kilometer
Distance to stadium	11,617	1.841	1.112	0.00	4.56	Kilometer
Distance to fire station	11,617	1.080	0.734	0.03	3.76	Kilometer
Distance to police	11,617	0.622	0.324	0.02	1.90	Kilometer
Distance to metro	11,617	0.611	0.719	0.03	6.25	Kilometer
Distance to train	11,617	1.149	0.771	0.00	4.79	Kilometer
Distance to parking	11,617	0.492	0.417	0.01	3.04	Kilometer
Distance to freeway	11,617	1.520	1.006	0.00	4.06	Kilometer
Conservation zone dummy	11,617	0.142	0.349	0	1	Dummy
Flood risk dummy	11,617	0.097	0.296	0	1	Dummy
View of the Tagus river	11,617	0.062	0.241	0	1	Dummy
Distance to Tagus	11,617	2.433	1.802	0.00	7.40	Kilometer
Distance to playground	11,617	0.424	0.181	0.02	1.19	Kilometer
Distance to football field	11,617	1.141	0.686	0.06	3.68	Kilometer
Distance to forest	11,617	1.098	0.675	0.00	2.75	Kilometer
Distance to cemetery	11,617	1.275	0.704	0.00	2.99	Kilometer
Distance to park	11,617	0.342	0.223	0.01	1.27	Kilometer
Forest size	11,617	0.841	2.051	0.04	10.29	Sq. kilometer
Cemetery Size	11,617	0.116	0.090	0.00	0.27	Sq. kilometer
Park size	11,617	0.028	0.083	0.00	1.00	Sq. kilometer
Cemetery NDVI	11,617	0.085	0.040	0.00	0.17	NDVI
Forest NDVI	11,617	0.138	0.064	0.01	0.27	NDVI
Park NDVI	11,617	0.051	0.055	-0.06	0.22	NDVI
% Canopy Coverage	11,617	0.049	0.046	0.00	0.21	Percent
Mean NDVI	11,617	0.063	0.049	-0.04	0.17	NDVI
% NDVI [0.2, 0.5] Coverage	11,617	0.142	0.084	0.03	0.34	Percent
% NDVI [0.5, 1.0] Coverage	11,617	0.012	0.013	0.00	0.05	Percent

Table A1 – 2. Spatial Weights (SW)

	Description	Nº. Locations	Nº. Non-zero Links	% Non-zero Links	Avg. Nº. of Links
SW1	Inverse distance of all properties in 500m	11,616	4,617,282	3.42	397.49
SW2	Inverse sq. distance of all properties in 500m	11,616	4,617,282	3.42	397.49

Table A1 – 3. Tests of Spatial Dependence

	Global Moran's I (Dep.)	Z-Value (Dep.)	Global Moran's I (Res.)	Z-Value (Res.)	LM Error	LM Lag	Rob. LM Error	Rob. LM Lag
Model 1								
SW1	0.1039***	21.19	0.0051*	1.50	1.068	0.172	1.495	0.598
SW2	0.1118***	12.42	0.0082	1.18	0.832	0.054	1.352	0.573
Model 2								
SW1	0.1039***	21.19	0.0027	1.03	0.3082	0.5931	0.6976	0.9824
SW2	0.1118***	12.42	0.0055	0.88	0.3717	0.2949	0.9726	0.8958
Model 3								
SW1	0.1039***	21.19	0.0029	1.0770	0.3453	0.3553	0.6599	0.6699
SW2	0.1118***	12.42	0.0061	0.9590	0.4586	0.1553	0.9642	0.6609
Model 4								
SW1	0.1039***	21.19	0.0036	1.2200	0.5427	0.2111	0.8553	0.5237
SW2	0.1118***	12.42	0.0060	0.9392	0.4378	0.0861	0.8288	0.4771
Model 5								
SW1	0.1039***	21.19	0.0019	0.89	0.155	0.700	0.472	1.017
SW2	0.1118***	12.42	0.0051	0.86	0.326	0.367	0.953	0.994
Model 6								
SW1	0.1039***	21.19	0.0015	0.81	0.096	0.733	0.367	1.004
SW2	0.1118***	12.42	0.0046	0.80	0.259	0.394	0.840	0.975
Model 7								
SW1	0.1039***	21.19	0.0013	0.7661	0.0665	0.7979	0.3162	1.0480
SW2	0.1118***	12.42	0.0040	0.7382	0.1965	0.4518	0.7500	1.0050
Model 8								
SW1	0.1039***	21.19	0.0012	0.76	0.064	0.919	0.334	1.189
SW2	0.1118***	12.42	0.0043	0.77	0.229	0.520	0.868	1.159
Model 9								
SW1	0.1039***	21.19	0.0017	0.84	0.114	0.628	0.375	0.890
SW2	0.1118***	12.42	0.0048	0.82	0.282	0.320	0.825	0.863
Model 10								
SW1	0.1039***	21.19	0.0013	0.82	0.087	0.777	0.400	0.834
SW2	0.1118***	12.42	0.0060	0.70	0.290	0.420	0.600	1.053
Model 11								
SW1	0.1039***	21.19	0.0014	0.79	0.079	0.796	0.343	1.060
SW2	0.1118***	12.42	0.0046	0.81	0.264	0.438	0.883	1.058
Model 12								
SW1	0.1039***	21.19	0.0016	0.830	0.107	0.675	0.374	0.942
SW2	0.1118***	12.42	0.0047	0.816	0.275	0.352	0.838	0.916
Model 13								
SW1	0.1039***	21.19	0.0027	1.084	0.307	0.377	0.614	0.683
SW2	0.1118***	12.42	0.0059	0.966	0.430	0.169	0.936	0.675

Table A1 – 4. Hedonic Valuation of Open Space Heterogeneity (Full Results)

Dep. Variable: ln(Price)	**Model 1**	***Open Space Heterogeneity***		
		Model 2	**Model 3**	**Model 4**
Intercept	4.95977**	4.50317**	4.93687**	4.49936**
	(2.21782)	(2.22333)	(2.21808)	(2.22362)
Structural Characteristics				
ln(Sq. meter)	0.78928***	0.78878***	0.78924***	0.78859***
	(0.00828)	(0.00827)	(0.00828)	(0.00827)
New dummy	0.15645***	0.15644***	0.15690***	0.15687***
	(0.00571)	(0.00570)	(0.00571)	(0.00571)
Pool dummy	0.11721***	0.11572***	0.11760***	0.11678***
	(0.02357)	(0.02355)	(0.02358)	(0.02355)
Parking dummy	0.07115***	0.07156***	0.07154***	0.07190***
	(0.00738)	(0.00737)	(0.00738)	(0.00738)
Fireplace dummy	0.03117**	0.02920**	0.03084**	0.02891**
	(0.01303)	(0.01302)	(0.01303)	(0.01302)

Double windows dummy	0.01481***	0.01526***	0.01477***	0.01529***
	(0.00555)	(0.00555)	(0.00555)	(0.00555)
Air conditioning dummy	0.14027***	0.13860***	0.14010***	0.13850***
	(0.00669)	(0.00668)	(0.00669)	(0.00669)
Elevator dummy	0.01380**	0.01417**	0.01378**	0.01427**
	(0.00562)	(0.00561)	(0.00562)	(0.00561)
Neighbourhood Characteristics				
% Non-residential buildings	0.0529	0.04347	0.06227*	0.05302
	(0.03326)	(0.03462)	(0.03388)	(0.03517)
% Buildings built pre 1919	-0.02998	-0.02341	-0.02932	-0.02193
	(0.02265)	(0.02303)	(0.02266)	(0.02310)
% Buildings built 1919-1945	-0.07547***	-0.05887***	-0.07971***	-0.06146***
	(0.01853)	(0.01975)	(0.01869)	(0.01991)
% Buildings built 1946-1960	-0.01211	-0.00813	-0.01661	-0.00746
	(0.01926)	(0.01985)	(0.01965)	(0.02003)
% Buildings built 1961-1970	-0.05135**	-0.01663	-0.05004**	-0.01519
	(0.02237)	(0.02365)	(0.02259)	(0.02392)
% Buildings built 1971-1980	-0.02821	-0.0249	-0.03312*	-0.03035
	(0.01961)	(0.02011)	(0.01978)	(0.02027)
% Buildings built 1981-1990	-0.19881***	-0.18190***	-0.20910***	-0.18728***
	(0.03549)	(0.03567)	(0.03600)	(0.03621)
% Buildings built 1991-1995	0.04771	0.03876	0.04407	0.04162
	(0.03412)	(0.03495)	(0.03448)	(0.03528)
% Buildings built 1996-2000	0.01675	0.03812	0.00505	0.02586
	(0.02351)	(0.02428)	(0.02527)	(0.02616)
% Buildings built 2000-2005	-0.13415***	-0.11712***	-0.13749***	-0.11655***
	(0.03600)	(0.03628)	(0.03613)	(0.03640)
ln(Average income)	0.39090*	0.44002**	0.39286*	0.43938**
	(0.22304)	(0.22369)	(0.22307)	(0.22373)
Population density	0.27617	0.31428	0.45275	0.49065
	(0.33709)	(0.34855)	(0.35128)	(0.36147)
% Higher educated	0.01848	0.01295	0.01039	0.00108
	(0.02548)	(0.02695)	(0.02610)	(0.02797)
% Population over 65	-0.05749	-0.0458	-0.05595	-0.05084
	(0.03958)	(0.04000)	(0.03991)	(0.04027)
% Population under 19	0.01202	-0.04133	0.00584	-0.05178
	(0.06714)	(0.06822)	(0.06740)	(0.06883)
Average commute time	-0.00106**	-0.00054	-0.00095*	-0.00037
	(0.00048)	(0.00050)	(0.00049)	(0.00051)
Nº. Neglected Buildings in 200 m	-0.00358***	-0.00298***	-0.00344***	-0.00291***
	(0.00088)	(0.00092)	(0.00089)	(0.00092)
Nº. Neighbourhood Monuments	0.00185	0.00314***	0.00178	0.00312***
	(0.00115)	(0.00118)	(0.00116)	(0.00118)
Accessibility Characteristics				
ln(Distance to airport)	-0.0153	-0.00957	-0.01752	-0.0118
	(0.01213)	(0.01261)	(0.01223)	(0.01274)
ln(Distance to stadium)	-0.00236	-0.00074	-0.00213	0.0007
	(0.00643)	(0.00649)	(0.00652)	(0.00658)
ln(Distance to supermarket)	0.00278	0.0032	0.0031	0.00285
	(0.00548)	(0.00583)	(0.00566)	(0.00602)
ln(Distance to cultural amenity)	0.00162	0.00027	0.00052	-0.00127
	(0.00730)	(0.00755)	(0.00732)	(0.00760)
ln(Distance to art gallery)	-0.02767***	-0.03079***	-0.02858***	-0.03172***
	(0.00784)	(0.00800)	(0.00789)	(0.00804)
ln(Distance to museum)	-0.00258	0.00179	-0.00172	0.00322
	(0.00834)	(0.00861)	(0.00838)	(0.00864)
ln(Distance to pharmacy)	-0.00122	0.00265	0.00003	0.00248
	(0.00563)	(0.00607)	(0.00576)	(0.00610)
ln(Distance to parking)	0.00447	0.00277	0.00464	0.00198
	(0.00483)	(0.00511)	(0.00486)	(0.00516)
ln(Distance to train)	-0.01583***	-0.01508**	-0.01693***	-0.01587***
	(0.00588)	(0.00608)	(0.00591)	(0.00613)
ln(Distance to metro)	-0.01580***	-0.01912***	-0.01622***	-0.02016***
	(0.00524)	(0.00534)	(0.00530)	(0.00541)
ln(Distance to high school)	-0.00628	-0.00657	-0.01063	-0.00954
	(0.00734)	(0.00752)	(0.00768)	(0.00789)

ln(Distance to university)	0.01397**	0.01517**	0.01215*	0.01420**
	(0.00634)	(0.00640)	(0.00641)	(0.00650)
ln(Distance to hospital)	0.01109*	0.01714**	0.01261*	0.01892***
	(0.00641)	(0.00674)	(0.00650)	(0.00684)
ln(Distance to mall)	-0.0068	-0.00925	-0.00471	-0.00584
	(0.00663)	(0.00668)	(0.00678)	(0.00686)
ln(Distance to police)	0.01433**	0.01272**	0.01554***	0.01417**
	(0.00589)	(0.00597)	(0.00593)	(0.00602)
ln(Distance to fire station)	0.03311***	0.02251***	0.03158***	0.02174***
	(0.00710)	(0.00765)	(0.00720)	(0.00766)
Local Environmental Characteristics				
Flood risk dummy	-0.02952**	-0.03235***	-0.02919**	-0.02658**
	(0.01204)	(0.01224)	(0.01247)	(0.01297)
Conservation zone dummy	-0.01507	-0.01254	-0.01512	-0.0137
	(0.01342)	(0.01355)	(0.01349)	(0.01359)
View of the Tagus river	0.06401***	0.06511***	0.06342***	0.06455***
	(0.00864)	(0.00863)	(0.00864)	(0.00864)
ln(Distance to Tagus)	-0.01560***	-0.01374**	-0.01438**	-0.01101*
	(0.00552)	(0.00563)	(0.00569)	(0.00582)
ln(Distance to freeway)	0.02423***	0.01474**	0.02358***	0.01495**
	(0.00724)	(0.00744)	(0.00728)	(0.00746)
Open Space Accessibility				
ln(Distance to football field)	-0.01522**	-0.01464**	-0.01601**	-0.01500**
	(0.00623)	(0.00632)	(0.00626)	(0.00634)
ln(Distance to playground)	0.00395	0.00273	0.00631	0.00348
	(0.00620)	(0.00653)	(0.00634)	(0.00668)
ln(Distance to forest)	-0.01892***	-0.02656***	-0.02675***	-0.02727***
	(0.00270)	(0.00318)	(0.00668)	(0.00702)
ln(Distance to park)	-0.00139	0.04511***	-0.00027	0.04680***
	(0.00450)	(0.01328)	(0.00481)	(0.01336)
ln(Distance to cemetery)	-0.00614*	-0.02002	0.00305	-0.00384
	(0.00353)	(0.01282)	(0.01089)	(0.01789)
Open Space Heterogeneity				
ln(Forest size)		-0.00259		-0.0036
		(0.00269)		(0.00279)
ln(Park size)		0.01498***		0.01572***
		(0.00439)		(0.00444)
ln(Cemetery Size)		0.00382		0.00337
		(0.00436)		(0.00460)
ln(Distance to forest) × ln(Forest size)		-0.00872***		-0.00920***
		(0.00182)		(0.00194)
ln(Distance to park) × ln(Park size)		0.00960***		0.00956***
		(0.00259)		(0.00262)
ln(Distance to cemetery) × ln(Cemetery Size)		-0.00741		-0.00699
		(0.00542)		(0.00545)
ln(Distance to forest) × Forest NDVI			0.04997	-0.00009
			(0.03942)	(0.04399)
ln(Distance to park) × Park NDVI			-0.04924	-0.06486
			(0.04420)	(0.04471)
ln(Distance to cemetery) × Cemetery NDVI			-0.06613	-0.10653
			(0.07088)	(0.07535)
Freguesia F.E.	Yes	Yes	Yes	Yes
Breusch-Pagan	486.62***	472.95***	496.02***	478.51***
Breusch-Godfrey	26.53***	24.03***	26.36***	24.02***
Durbin-Watson	1.90***	1.91***	1.90***	1.91***
AIC	-3030.10	-3056.40	-3028.20	-3055.40
SSE	514.62	512.93	514.44	512.71
Observations	11,617	11,617	11,617	11,617
Adjusted R^2	0.66527	0.66619	0.6653	0.66625
Residual Std. Error	0.21142	0.21113	0.21141	0.21111
F Statistic	225.13***	213.68***	218.82***	208.03***

*Notes: ***Significance at 1 % level; **Significance at 5 % level; *Significance at 10 % level.*
Heteroskedastic Consistent Errors

Table A1 – 5. Impacts of Urban Greenness and Open Space (Full Results)

Dep. Variable: ln(Price)	***Urban Tree Canopy Coverage***			***Mean NDVI***			***NDVI by Vegetation Type***		
	Model 5	**Model 6**	**Model 7**	**Model 8**	**Model 9**	**Model 10**	**Model 11**	**Model 12**	**Model 13**
Intercept	4.55710**	4.41484**	4.92013**	4.62164**	4.36259*	4.89187**	4.65924**	5.78615***	5.30349**
	(2.22373)	*(2.22520)*	*(2.22843)*	*(2.22522)*	*(2.23045)*	*(2.22751)*	*(2.22554)*	*(2.22611)*	*(2.23374)*
Structural Characteristics									
ln(Sq. meter)	0.78871***	0.78789***	0.78722***	0.78873***	0.78813***	0.78736***	0.78875***	0.78429***	0.78614***
	(0.00827)	*(0.00827)*	*(0.00827)*	*(0.00827)*	*(0.00827)*	*(0.00827)*	*(0.00827)*	*(0.00826)*	*(0.00827)*
New dummy	0.15687***	0.15626***	0.15678***	0.15684***	0.15662***	0.15690***	0.15679***	0.15473***	0.15583***
	(0.00571)	*(0.00571)*	*(0.00570)*	*(0.00571)*	*(0.00571)*	*(0.00570)*	*(0.00571)*	*(0.00570)*	*(0.00570)*
Pool dummy	0.11637***	0.11553***	0.11639***	0.11631***	0.11566***	0.11648***	0.11628***	0.11475***	0.12115***
	(0.02355)	*(0.02355)*	*(0.02354)*	*(0.02355)*	*(0.02355)*	*(0.02354)*	*(0.02355)*	*(0.02351)*	*(0.02352)*
Parking dummy	0.07173***	0.07155***	0.07187***	0.07181***	0.07163***	0.07217***	0.07184***	0.07150***	0.07191***
	(0.00738)	*(0.00737)*	*(0.00737)*	*(0.00738)*	*(0.00737)*	*(0.00737)*	*(0.00738)*	*(0.00735)*	*(0.00736)*
Fireplace dummy	0.02853**	0.02835**	0.02811**	0.02866**	0.02835**	0.02788**	0.02844**	0.02630**	0.02698**
	(0.01302)	*(0.01302)*	*(0.01302)*	*(0.01302)*	*(0.01302)*	*(0.01303)*	*(0.01302)*	*(0.01299)*	*(0.01302)*
Double windows dummy	0.01540***	0.01592***	0.01605***	0.01551***	0.01572***	0.01652***	0.01562***	0.01633***	0.01683***
	(0.00555)	*(0.00555)*	*(0.00555)*	*(0.00555)*	*(0.00555)*	*(0.00556)*	*(0.00555)*	*(0.00554)*	*(0.00555)*
Air conditioning dummy	0.13832***	0.13857***	0.13871***	0.13839***	0.13836***	0.13871***	0.13821***	0.13680***	0.13815***
	(0.00669)	*(0.00669)*	*(0.00668)*	*(0.00669)*	*(0.00669)*	*(0.00669)*	*(0.00669)*	*(0.00667)*	*(0.00668)*
Elevator dummy	0.01426**	0.01438**	0.01422**	0.01424**	0.01438**	0.01385**	0.01418**	0.01574***	0.01502***
	(0.00561)	*(0.00561)*	*(0.00561)*	*(0.00561)*	*(0.00561)*	*(0.00561)*	*(0.00562)*	*(0.00560)*	*(0.00561)*
Neighbourhood Characteristics									
% Non-residential buildings	0.04832	0.06627*	0.06559*	0.05398	0.06464*	0.06575*	0.05243	0.12519***	0.08888**
	(0.03528)	*(0.03595)*	*(0.03562)*	*(0.03517)*	*(0.03577)*	*(0.03590)*	*(0.03517)*	*(0.03651)*	*(0.03657)*
% Buildings built pre 1919	-0.01241	-0.01934	-0.01	-0.01311	-0.02104	-0.01537	-0.01132	-0.01526	0.00227
	(0.02381)	*(0.02405)*	*(0.02407)*	*(0.02393)*	*(0.02497)*	*(0.02451)*	*(0.02457)*	*(0.02530)*	*(0.02527)*
% Buildings built 1919-1945	-0.05543***	-0.05164**	-0.04449**	-0.05675***	-0.04730**	-0.05101**	-0.05506***	-0.03023	-0.03731*
	(0.02024)	*(0.02057)*	*(0.02075)*	*(0.02019)*	*(0.02043)*	*(0.02068)*	*(0.02060)*	*(0.02094)*	*(0.02118)*
% Buildings built 1946-1960	-0.00742	-0.00314	-0.00182	-0.00798	0.00066	-0.00443	-0.00856	-0.01093	-0.00619
	(0.02003)	*(0.02054)*	*(0.02024)*	*(0.02004)*	*(0.02030)*	*(0.02034)*	*(0.02029)*	*(0.02033)*	*(0.02060)*
% Buildings built 1961-1970	-0.01597	-0.00015	-0.0111	-0.01972	-0.01436	-0.02694	-0.02142	-0.01399	-0.02169
	(0.02392)	*(0.02473)*	*(0.02412)*	*(0.02413)*	*(0.02469)*	*(0.02455)*	*(0.02459)*	*(0.02511)*	*(0.02478)*
% Buildings built 1971-1980	-0.03467*	-0.03313	-0.02491	-0.03283	-0.02221	-0.01808	-0.03634*	-0.01061	-0.0122
	(0.02044)	*(0.02096)*	*(0.02128)*	*(0.02034)*	*(0.02086)*	*(0.02143)*	*(0.02054)*	*(0.02104)*	*(0.02153)*
% Buildings built 1981-1990	-0.18584***	-0.14373***	-0.16035***	-0.18662***	-0.18057***	-0.17241***	-0.18690***	-0.12717***	-0.15853***
	(0.03621)	*(0.03874)*	*(0.03689)*	*(0.03621)*	*(0.03677)*	*(0.03689)*	*(0.03715)*	*(0.03841)*	*(0.03763)*
% Buildings built 1991-1995	0.0426	0.03602	0.04577	0.04248	0.02812	0.03983	0.04365	0.02077	0.06318*
	(0.03529)	*(0.03572)*	*(0.03572)*	*(0.03529)*	*(0.03587)*	*(0.03559)*	*(0.03533)*	*(0.03551)*	*(0.03609)*
% Buildings built 1996-2000	0.02688	0.03205	0.02886	0.02774	0.02233	0.02493	0.02799	-0.00398	0.03129
	(0.02617)	*(0.02653)*	*(0.02644)*	*(0.02619)*	*(0.02659)*	*(0.02672)*	*(0.02620)*	*(0.02650)*	*(0.02694)*

% Buildings built 2000-2005	-0.12570***	-0.13233***	-0.13063***	-0.12209***	-0.10846***	-0.13161***	-0.12380***	-0.12265***	-0.16197***
	(0.03682)	*(0.03704)*	*(0.03704)*	*(0.03661)*	*(0.03703)*	*(0.03677)*	*(0.03698)*	*(0.03964)*	*(0.03875)*
ln(Average income)	0.43276*	0.44894**	0.39248*	0.42594*	0.45796**	0.39528*	0.42262*	0.30586	0.35426
	(0.22375)	*(0.22392)*	*(0.22423)*	*(0.22392)*	*(0.22456)*	*(0.22418)*	*(0.22391)*	*(0.22401)*	*(0.22473)*
Population density	0.47014	0.25258	0.42851	0.50885	0.47531	0.598	0.50099	0.78272**	0.86416**
	(0.36165)	*(0.38309)*	*(0.36534)*	*(0.36168)*	*(0.36793)*	*(0.36385)*	*(0.36161)*	*(0.37469)*	*(0.37739)*
% Higher educated	0.01125	0.00272	0.01471	0.01272	-0.00833	0.01837	0.01295	0.02626	0.02442
	(0.02864)	*(0.02906)*	*(0.02908)*	*(0.02916)*	*(0.02996)*	*(0.02982)*	*(0.02954)*	*(0.03070)*	*(0.03006)*
% Population over 65	-0.05144	-0.06985*	-0.05347	-0.04624	-0.08938**	-0.05148	-0.04769	-0.04884	-0.06852
	(0.04026)	*(0.04082)*	*(0.04091)*	*(0.04040)*	*(0.04273)*	*(0.04064)*	*(0.04043)*	*(0.04199)*	*(0.04237)*
% Population under 19	-0.06296	-0.09203	-0.05304	-0.05795	-0.0785	-0.05615	-0.06568	-0.07705	-0.09379
	(0.06916)	*(0.07176)*	*(0.07186)*	*(0.06896)*	*(0.07101)*	*(0.06900)*	*(0.06943)*	*(0.07161)*	*(0.07156)*
Average commute time	-0.00046	-0.00024	-0.0005	-0.00049	-0.00047	-0.00059	-0.00045	-0.00064	-0.00049
	(0.00051)	*(0.00052)*	*(0.00051)*	*(0.00051)*	*(0.00052)*	*(0.00052)*	*(0.00051)*	*(0.00052)*	*(0.00052)*
Nº. Neglected Buildings in 200 m	-0.00286***	-0.00279***	-0.00235**	-0.00276***	-0.00271***	-0.00215**	-0.00288***	-0.00176*	-0.00245**
	(0.00092)	*(0.00094)*	*(0.00094)*	*(0.00092)*	*(0.00095)*	*(0.00099)*	*(0.00093)*	*(0.00095)*	*(0.00097)*
Nº. Neighbourhood Monuments	0.00216	0.0018	0.0019	0.00201	0.00206	0.00167	0.00152	0.00119	0.00082
	(0.00132)	*(0.00133)*	*(0.00137)*	*(0.00142)*	*(0.00154)*	*(0.00144)*	*(0.00144)*	*(0.00148)*	*(0.00155)*
Accessibility Characteristics									
ln(Distance to airport)	-0.01055	-0.00912	-0.00905	-0.00972	-0.02298*	-0.00823	-0.00987	-0.01008	-0.02722**
	(0.01276)	*(0.01331)*	*(0.01285)*	*(0.01283)*	*(0.01389)*	*(0.01295)*	*(0.01282)*	*(0.01332)*	*(0.01370)*
ln(Distance to stadium)	-0.00027	-0.00213	-0.00256	-0.00087	-0.00075	-0.00213	-0.00122	-0.00449	-0.00507
	(0.00661)	*(0.00665)*	*(0.00664)*	*(0.00668)*	*(0.00673)*	*(0.00671)*	*(0.00665)*	*(0.00680)*	*(0.00673)*
ln(Distance to supermarket)	0.00163	0.00017	-0.00185	0.00187	0.00284	0.00015	0.00029	0.00403	-0.00451
	(0.00606)	*(0.00613)*	*(0.00617)*	*(0.00605)*	*(0.00609)*	*(0.00610)*	*(0.00633)*	*(0.00649)*	*(0.00648)*
ln(Distance to cultural amenity)	-0.00074	-0.001	0.00066	-0.00018	0.00123	0.00257	-0.00061	0.00624	0.00285
	(0.00760)	*(0.00769)*	*(0.00761)*	*(0.00764)*	*(0.00770)*	*(0.00771)*	*(0.00761)*	*(0.00773)*	*(0.00769)*
ln(Distance to art gallery)	-0.03101***	-0.03281***	-0.02755***	-0.03053***	-0.02390***	-0.02458***	-0.02982***	-0.01821**	-0.02075**
	(0.00805)	*(0.00819)*	*(0.00835)*	*(0.00808)*	*(0.00842)*	*(0.00838)*	*(0.00809)*	*(0.00839)*	*(0.00863)*
ln(Distance to museum)	0.00253	0.00324	0.00068	0.00206	-0.00191	0.00103	0.00139	0.00333	0.00585
	(0.00865)	*(0.00868)*	*(0.00871)*	*(0.00868)*	*(0.00878)*	*(0.00873)*	*(0.00871)*	*(0.00877)*	*(0.00890)*
ln(Distance to pharmacy)	0.00316	0.00231	0.00291	0.00344	0.00187	0.00363	0.00457	-0.00392	0.00326
	(0.00611)	*(0.00616)*	*(0.00611)*	*(0.00614)*	*(0.00618)*	*(0.00614)*	*(0.00627)*	*(0.00642)*	*(0.00636)*
ln(Distance to parking)	0.00219	0.00367	0.00102	0.00084	0.00231	-0.00179	0.00124	0.0013	0.0007
	(0.00516)	*(0.00531)*	*(0.00522)*	*(0.00522)*	*(0.00534)*	*(0.00530)*	*(0.00517)*	*(0.00522)*	*(0.00522)*
ln(Distance to train)	-0.01690***	-0.01620***	-0.01286**	-0.01790***	-0.01704***	-0.01816***	-0.01762***	-0.01111*	-0.00773
	(0.00616)	*(0.00627)*	*(0.00637)*	*(0.00629)*	*(0.00634)*	*(0.00647)*	*(0.00619)*	*(0.00629)*	*(0.00653)*
ln(Distance to metro)	-0.01943***	-0.01914***	-0.01698***	-0.01935***	-0.01749***	-0.01766***	-0.01879***	-0.01335**	-0.01854***
	(0.00542)	*(0.00545)*	*(0.00548)*	*(0.00544)*	*(0.00556)*	*(0.00549)*	*(0.00545)*	*(0.00555)*	*(0.00583)*
ln(Distance to high school)	-0.00955	-0.00496	-0.00916	-0.00888	-0.0088	-0.01142	-0.00809	-0.0031	-0.01119
	(0.00789)	*(0.00808)*	*(0.00797)*	*(0.00791)*	*(0.00797)*	*(0.00797)*	*(0.00794)*	*(0.00821)*	*(0.00804)*
ln(Distance to university)	0.01496**	0.01657**	0.01282*	0.01447**	0.01392**	0.01375**	0.01481**	0.0016	0.01363**
	(0.00651)	*(0.00675)*	*(0.00678)*	*(0.00650)*	*(0.00666)*	*(0.00675)*	*(0.00651)*	*(0.00721)*	*(0.00682)*

ln(Distance to hospital)	0.01907***	0.01836***	0.01635**	0.01867***	0.01654**	0.01395**	0.01891***	0.01266*	0.00636
	(0.00684)	*(0.00692)*	*(0.00700)*	*(0.00684)*	*(0.00696)*	*(0.00699)*	*(0.00685)*	*(0.00705)*	*(0.00734)*
ln(Distance to mall)	-0.00573	-0.00509	-0.00574	-0.00518	-0.00952	-0.00703	-0.00646	-0.00566	-0.00125
	(0.00686)	*(0.00711)*	*(0.00709)*	*(0.00688)*	*(0.00723)*	*(0.00725)*	*(0.00709)*	*(0.00751)*	*(0.00739)*
ln(Distance to police)	0.01334**	0.01293**	0.01133*	0.01297**	0.01247**	0.01116*	0.01348**	0.01139*	0.01229**
	(0.00604)	*(0.00611)*	*(0.00612)*	*(0.00608)*	*(0.00622)*	*(0.00624)*	*(0.00610)*	*(0.00669)*	*(0.00626)*
ln(Distance to fire station)	0.01904**	0.01916**	0.01798**	0.02052***	0.01870**	0.02254***	0.01866**	0.01101	0.00545
	(0.00783)	*(0.00791)*	*(0.00790)*	*(0.00771)*	*(0.00788)*	*(0.00795)*	*(0.00782)*	*(0.00794)*	*(0.00828)*
Local Environmental Characteristics									
ln(Distance to Tagus)	-0.01114*	-0.01244**	-0.00763	-0.01068*	-0.01580**	-0.00847	-0.01141*	0.00193	0.00895
	(0.00582)	*(0.00606)*	*(0.00603)*	*(0.00582)*	*(0.00615)*	*(0.00609)*	*(0.00583)*	*(0.00716)*	*(0.00753)*
ln(Distance to freeway)	0.01438*	0.01545**	0.02774***	0.01508**	0.01390*	0.02164*	0.01383*	0.01309*	0.04429***
	(0.00746)	*(0.00759)*	*(0.00971)*	*(0.00746)*	*(0.00773)*	*(0.01149)*	*(0.00753)*	*(0.00763)*	*(0.01300)*
Flood risk dummy	-0.02579**	-0.02017	-0.06991***	-0.02551**	-0.03261**	-0.05494***	-0.02570**	-0.02999**	-0.08334***
	(0.01298)	*(0.01322)*	*(0.02115)*	*(0.01299)*	*(0.01324)*	*(0.01808)*	*(0.01300)*	*(0.01316)*	*(0.02479)*
Conservation zone dummy	-0.01511	-0.01702	-0.07279***	-0.01515	-0.01782	-0.09139***	-0.01705	0.01464	-0.06881**
	(0.01362)	*(0.01421)*	*(0.02088)*	*(0.01363)*	*(0.01383)*	*(0.02638)*	*(0.01371)*	*(0.01470)*	*(0.02726)*
View of the Tagus river	0.06451***	0.06391***	0.06378***	0.06454***	0.06424***	0.06384***	0.06450***	0.06384***	0.06193***
	(0.00864)	*(0.00864)*	*(0.00863)*	*(0.00864)*	*(0.00864)*	*(0.00863)*	*(0.00864)*	*(0.00861)*	*(0.00863)*
Open Space Accessibility									
ln(Distance to football field)	-0.01581**	-0.01639**	-0.01611**	-0.01547**	-0.01893***	-0.01814***	-0.01596**	-0.02034***	-0.01704***
	(0.00636)	*(0.00643)*	*(0.00642)*	*(0.00635)*	*(0.00651)*	*(0.00643)*	*(0.00636)*	*(0.00646)*	*(0.00641)*
ln(Distance to playground)	0.00409	-0.01099	0.00306	0.00303	0.00273	0.00169	0.00423	-0.00461	0.0037
	(0.00669)	*(0.00793)*	*(0.00670)*	*(0.00669)*	*(0.00716)*	*(0.00671)*	*(0.00669)*	*(0.01136)*	*(0.00677)*
ln(Distance to forest)	-0.02777***	-0.03119***	-0.01997***	-0.02723***	-0.02988***	-0.02227***	-0.02755***	0.02276	-0.02455***
	(0.00703)	*(0.01142)*	*(0.00729)*	*(0.00702)*	*(0.01123)*	*(0.00724)*	*(0.00703)*	*(0.01471)*	*(0.00729)*
ln(Distance to park)	0.04753***	0.05519***	0.03834***	0.04687***	0.06436***	0.03227**	0.04664***	0.02165	0.02767**
	(0.01337)	*(0.01467)*	*(0.01356)*	*(0.01336)*	*(0.01550)*	*(0.01399)*	*(0.01338)*	*(0.01587)*	*(0.01386)*
ln(Distance to cemetery)	-0.00073	-0.00512	0.01273	-0.00019	-0.01061	0.00427	0.00156	-0.00894	0.00975
	(0.01799)	*(0.01808)*	*(0.01954)*	*(0.01808)*	*(0.01880)*	*(0.01981)*	*(0.01811)*	*(0.01925)*	*(0.02028)*
Open Space Heterogeneity									
ln(Forest size)	-0.00404	-0.00597**	-0.00356	-0.00383	-0.00313	-0.00305	-0.00424	-0.00740**	-0.00486*
	(0.00280)	*(0.00287)*	*(0.00280)*	*(0.00279)*	*(0.00284)*	*(0.00283)*	*(0.00281)*	*(0.00288)*	*(0.00286)*
ln(Park size)	0.01504***	0.01530***	0.01339***	0.01521***	0.01628***	0.01316***	0.01486***	0.02181***	0.01543***
	(0.00446)	*(0.00452)*	*(0.00449)*	*(0.00445)*	*(0.00459)*	*(0.00458)*	*(0.00446)*	*(0.00484)*	*(0.00452)*
log(Cemetery size)	0.00381	0.0033	0.00229	0.00371	0.00326	0.00183	0.00416	0.00185	0.00167
	(0.00461)	*(0.00464)*	*(0.00465)*	*(0.00460)*	*(0.00468)*	*(0.00473)*	*(0.00462)*	*(0.00468)*	*(0.00469)*
ln(Distance to forest) × ln(Forest size)	-0.00927***	-0.00987***	-0.00989***	-0.00914***	-0.00882***	-0.00915***	-0.00924***	-0.0017	-0.01195***
	(0.00194)	*(0.00212)*	*(0.00196)*	*(0.00194)*	*(0.00201)*	*(0.00199)*	*(0.00194)*	*(0.00282)*	*(0.00207)*
ln(Distance to park) × ln(Park size)	0.00958***	0.00996***	0.00762***	0.00956***	0.01058***	0.00708***	0.00936***	0.01188***	0.00774***
	(0.00262)	*(0.00265)*	*(0.00265)*	*(0.00262)*	*(0.00272)*	*(0.00272)*	*(0.00263)*	*(0.00280)*	*(0.00271)*
ln(Distance to cemetery) × log(Cemetery size)	-0.0058	-0.00747	-0.00396	-0.00579	-0.00666	-0.00529	-0.00568	-0.00489	-0.00395
	(0.00550)	*(0.00552)*	*(0.00567)*	*(0.00552)*	*(0.00558)*	*(0.00566)*	*(0.00553)*	*(0.00591)*	*(0.00578)*

ln(Distance to forest) × Forest NDVI	0.00578	0.01857	-0.0557	0.00276	0.00962	-0.03595	0.0044	-0.005	-0.03998
	(0.04413)	*(0.05083)*	*(0.04647)*	*(0.04404)*	*(0.05085)*	*(0.04617)*	*(0.04409)*	*(0.04555)*	*(0.04640)*
ln(Distance to park) × Park NDVI	-0.07294	-0.0593	-0.09950**	-0.07358	-0.07292	-0.06579	-0.07725*	-0.08751*	-0.06237
	(0.04498)	*(0.04526)*	*(0.04628)*	*(0.04514)*	*(0.04581)*	*(0.04671)*	*(0.04517)*	*(0.04772)*	*(0.04814)*
ln(Distance to cemetery) × Cemetery NDVI	-0.10822	-0.10045	-0.17059**	-0.11079	-0.05305	-0.13696*	-0.11922	-0.02961	-0.14569*
	(0.07535)	*(0.07595)*	*(0.08095)*	*(0.07540)*	*(0.07935)*	*(0.08324)*	*(0.07577)*	*(0.07816)*	*(0.08540)*
Measures of Urban Greenness (Measured at the Freguesia Scale)									
% Canopy Coverage × ln(*Freguesia* size)	0.14721*	0.19216**	0.20385*						
	(0.08935)	*(0.09108)*	*(0.10630)*						
Mean NDVI × ln(*Freguesia* size)				0.10897	0.04931	0.16251**			
				(0.07728)	*(0.08449)*	*(0.08117)*			
% NDVI[0.2,0.5] Coverage × ln(*Freguesia* size)							0.0338	0.05355	0.09129*
							(0.03826)	*(0.04116)*	*(0.04775)*
% NDVI[0.5,1] Coverage × ln(*Freguesia* size)							0.53408	0.06843	-0.19199
							(0.47173)	*(0.48321)*	*(0.52411)*
Local Urban Greenness Interactions (Local Urban Greenness Measured at the Collection of Census Tracts Scale)									
	Greenness: % Canopy Coverage			**Greenness: Mean NDVI Value**			**Greenness: % NDVI in [0.2, 0.5]**		
Greenness × ln(Distance to forest)		0.02695			0.02592			-0.30215***	
		(0.11200)			*(0.08138)*			*(0.06937)*	
Greenness × ln(Distance to park)		-0.24100*			-0.25484***			0.22178***	
		(0.13648)			*(0.08606)*			*(0.08477)*	
Greenness × ln(Distance to playground)		0.55430***			0.15198*			0.07369	
		(0.15777)			*(0.08605)*			*(0.11518)*	
Greenness × ln(Distance to freeway)			-0.24641			-0.08031			-0.21557**
			(0.15948)			*(0.14477)*			*(0.09100)*
Greenness × Flood risk dummy			0.97734**			0.46593*			0.21301
			(0.40832)			*(0.26270)*			*(0.23989)*
Greenness × Conservation zone dummy			0.80115***			0.88043***			-0.10131
			(0.28112)			*(0.29276)*			*(0.15086)*
Greenness × ln(Distance to Tagus)		0.08932	-0.0598		-0.10774	-0.08709		-0.17067**	-0.35203***
		(0.11203)	*(0.09418)*		*(0.06609)*	*(0.06262)*		*(0.07209)*	*(0.06188)*
							Greenness: % NDVI > 0.5		
Greenness × ln(Distance to forest)								0.26095	
								(0.48057)	
Greenness × ln(Distance to park)								-0.49237	
								(0.44844)	
Greenness × ln(Distance to playground)								0.78442	
								(0.67818)	
Greenness × ln(Distance to freeway)									-0.22102
									(0.63224)
Greenness × Flood risk dummy									1.47264
									(1.49857)

Greenness × Conservation zone dummy									3.29077*** *(1.09931)*
Greenness × ln(Distance to Tagus)								0.3716 *(0.37735)*	0.98485*** *(0.29024)*
Freguesia F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Breusch-Pagan	480.78***	487.67***	481.66***	479.07***	482.55***	477.36***	479.05***	491.24***	483.4***
Breusch-Godfrey	23.99***	23.25***	22.05***	24.02***	23.61***	22.64***	23.94***	19.64***	20.74***
Durbin-Watson	1.91***	1.91***	1.91***	1.91***	1.91***	1.91***	1.91***	1.91***	1.92***
AIC	-3056.10	-3060.90	-3070.70	-3055.40	-3058.20	-3064.80	-3055.40	-3120.30	-3097.60
SSE	512.59	512.03	511.60	512.62	512.15	511.86	512.53	508.98	509.97
Observations	11,617	11,617	11,617	11,617	11,617	11,617	11,617	11,617	11,617
Adjusted R^2	0.6663	0.6665	0.6668	0.6662	0.6664	0.6666	0.6663	0.6683	0.6677
Residual Std. Error	0.21110	0.21102	0.21093	0.21110	0.21104	0.21098	0.21109	0.21043	0.21064
F Statistic	206.25***	199.45***	199.70***	206.23***	199.38***	199.55***	204.45***	192.91***	192.35***

*Notes: ***Significance at 1 % level; **Significance at 5 % level; *Significance at 10 % level.*
Heteroskedastic Consistent Errors

CHAPTER 2:
Housing Price Boundary Effects From Flooding and Seismic Risk Zones

This chapter considers the capitalization of urban hazard information via the residential real estate market. A spatial hedonic framework with geographic regression discontinuities is used to estimate the impact of being located across areas with different levels of urban flood risks or seismic hazards. Special attention is given to capturing heterogeneity via varying local amenities and characteristics within and between the different zones, or differing effects along the distribution of housing prices.

2 – 1. Introduction

As population increasingly take up residency in urban areas and create denser living areas, a growing number of individuals are exposed to a variety of environmental hazard risks on a daily basis. Earthquakes, floods, landslides, avalanches and tsunamis are all examples of such hazards impacting cities across the globe and presenting severe threats to humans, property and the natural or built environment. Landslides can be caused by any number of factors including heavy rainfall and floods, earthquakes, or human activities, and are the most common natural hazards on land. In terms of casualties however, earthquakes and floods are often considered to be among the most significant. Tsunamis, many times a by-product of seismic events, are another type of geohazard that are relatively rare but as the 2004 deadly Indian Ocean tsunami tragically illustrated, their impacts can be devastating.[1]

In 2002, across the globe over 500 natural disasters were recorded with a total estimated direct damage of $55 billion and $13 billion in insurance losses, further killing 10,000 people and impacting 600 million more (United Nations 2004). The risk of being exposed to environmental hazards vary greatly according to location, climate, topography and the built environment. Cities must obey the particularities of their urban risks as they vary across the globe and consider these factors as they design and implement resilient policies, assess costs

[1] Geohazards are conditions relating to geology that have the potential to cause harm and damage, often involving some form of ground motion or instability. Examples include earthquakes, volcanic eruptions, landslides, flooding and tsunamis.

and support the local community. Each time these natural hazard events and their devastating losses occur, however, the same questions arise about the necessity of better managing urban development in areas prone to natural hazard.

The need to account for geological factors in land-use planning has often been urged by the United Nations. This is especially the case in coastal lowlands and more so for urban centers, most of which are located in earthquake zones or other hazard prone areas. Today, as more than half of the world's population live in urban areas, and coupled with the impacts of climate change, risk reduction strategies in urban areas are key to building resilient communities. The reach and potential impacts of natural hazards increase significantly in denser and growing urban areas and have important consequences for public policies supporting infrastructure, safety, mitigation strategies, cleanup or rebuilding (Lall and Deichmann 2010, Gencer 2013). Furthermore, many urban areas are confronted with a multitude of natural hazard risks and anthropogenic hazards such as accidents, pollution, explosions and fire. For instance, using a sample of 52 European cities representing 15% of the EU population, the PanGeo-project shows that an average sized European city could have four different types of geohazards covering an area of 186 km^2 and, exposing 626,000 people. Compressible-ground was identified by the PanGEo-project as the largest urban geohazard, by area, affecting the sampled European cities.[2] This is not surprising considering that most European cities have grown near rivers or coasts where compressible sediments and alluvium often accumulate. Therefore, identifying the spatial distribution and concentration of hazard risks in urban areas is crucial to understanding where and how preventative and corrective actions can reduce levels of vulnerability and exposure of urban inhabitants.

[2] Certain types of ground contain layers of very soft materials like peat and clays. These layers are often likely to compress if they are loaded by overlying structures, or if the groundwater level changes around them. This compression may result in depressions appearing in the ground surface or under structures, potentially damaging foundations and infrastructure. There are a number of problems that may affect properties built in such types of ground, including structural damage to foundations and to the fabric of the building, strains or break in service connections to water, gas and electricity or, cracks in walls, floors or ceilings of a building.

The goal of this chapter is to study the effects of localized natural hazard vulnerability on housing prices. In particular, to examine the direct effects that hazard risk zones (flooding or seismic) have on prices in Lisbon, Portugal, and further whether there exists any interactions with local amenities or spillover effects along the distribution of prices.

While many direct costs can be measured in relation to cleanup efforts, the estimated economic impact of urban hazard events often omit important effects such as persistent influences on the real estate market. When the location and occurrences of events can be predicted and made available to the general population through hazard risk maps or the media, residents can supposedly react accordingly and these risks may illicit important behavioral responses with impacts on the real estate market. In general, the market absorbs these behavioral responses through the price that households and firms are willing to pay for real estate in a particular location. Existing housing price studies on urban natural hazards usually deal with the effect of a single hazard and show that the residential real estate market responds to a natural hazard with depressed property prices in flood zones (Bin and Landry 2008, Bin and Landry 2013, Rajapaksa et al. 2016, Rajapaksa et al. 2017a, Rajapaksa et al. 2017b) and in areas with high seismic risk (Naoi et al. 2009, Naoi et al. 2010, Hidano et al. 2015).

Yet, little research has examined the spillover effects of high-risk areas on nearby real estate values nor the price impacts of multiple natural hazard risks and their interactions with local amenities and the built environment. This is surprising as many urban properties are vulnerable to multiple hazards and the continuous nature of spatial interactions yield contagion effects from high-risk areas to low-risk areas within close proximity. In addition, the variability of local amenities within hazard zones may mitigate or exacerbate the effects of their risks, translating to effects in their property value.

Finally, the degree of capitalization into property prices of different and multiple types of natural hazards can reveal not only the residents' risk beliefs but also the perceived potential

private property damage of these dis-amenities. Even if hazard risk maps are publicly available and media mentions occur, residents may still be apt to underestimate or even ignore these risks. For instance, people can feel either apathetic or optimistic about their risk of death, injury, or property damage due to a natural hazard (Bakkensen and Barrage 2018), or act as if a very low probability of an extreme event is zero by believing that it will not harm them or their property (Lindell 1997). Moreover, residents may expect they will be provided with disaster relief from governments and nonprofit organizations in case a natural disaster occurs (Burby 1998). Therefore, understanding the capitalization into housing prices of natural hazard risks can help local governments to be more aware of their residents' risk beliefs and behavior with regards to urban natural hazards and how mitigation measures may contribute to housing prices. It can also provide valuable information to value insurance contracts, to design resilient development strategies and to determine future urban development locations.

Lisbon provides an interesting context for studying the impact of natural hazard risks in an urban setting given its topography, coastal location and climate. Flooding occurrences are yearly events in the city and with its varied topography, with many valleys and hills, the areas of high flooding risk in Lisbon are not constrained to the riverfront like in many cities. On the other hand, seismic events are far less common, though the risk is still present and supposedly known to many residents. The country's history inflects on the date of the devastating category M9 Great Lisbon Earthquake of 1755, and this event remains much ingrained in modern culture and urban planning.

As such, the within city spatial heterogeneity of seismic and flooding risks is expected to be capitalized into Lisbon's residential property prices. A spatial hedonic framework is used to decompose the price of a residential dwelling into its value bearing attributes, paying particular attention to location within the city and relative to amenities and important areas of the city. The variability of georeferenced dwellings across zones of different risks in the city

allows for the estimated impact of being located in areas of a municipality exposed to greater hazards. If the residential real estate market capitalizes on these risks, then we would expect negative impacts on prices for dwellings located within or near such zones.

The empirical specification is chosen based on spatial diagnostics from the Moran's I and Lagrange multiplier statistics which test the spatial heterogeneity of the dependent variable and spatial autocorrelation of the error term. Results indicate the presence of significant spatial influences which may impact standard OLS results, and thus a spatial error (SEM) specification is employed to mitigate this potential bias.

The analysis further take spatial influences and locational spillovers into consideration in a number of ways. First, in constructing measures of location and neighbourhoods we forgo the use of pre-defined administrative boundaries and construct proper measures based on distances reflective of how residents perceive their neighbourhood. This mitigates potential biases from the modifiable aerial unit problem which may arise by using inconsistently sized administrative boundaries to represent neighbourhood and locational realities which may be delineated according to political or topographic considerations.

Second, to ensure that the estimated effects are not driven by underlying locational features, a geographic regression discontinuity (GRD) framework is used whereby the boundaries of areas with greater natural hazard risks are used as a geographic threshold. Results on the inside of the zones can thus be considered as treatments, while those located nearby on the other side of the boundary and in a non-hazard zone may be considered valid controls. A propensity score matching is also used to find valid control properties conditional on key locational features, and shows that results are robust when considering a range of potential underlying locational mechanisms which may be driving estimates.

Estimates indicate that the residential real estate market in Lisbon responds negatively to areas of increased natural hazard risk. Being located in areas with very high flooding

potential or damage from seismic risk yields a reduction in dwelling prices on average of 3.5% and 1.1%, respectively. Evidence also suggests that residents respond to the severity of the potential impacts with larger estimates for very high risk areas compared to high risk areas. While the potential damage effect of seismic events is significantly greater than flooding events, the estimated housing price impacts tends to be smaller. This signals that the real estate market underestimates the potential risk of this type of geohazard event, likely due to their scarcity relative to flooding which are seasonal occurrences. Further, given the topography of the city it is possible to have areas of both high seismic and high flooding risks. As expected, these joint hazard zones have stronger negative impacts on dwelling prices in the range of 3.8%. Using a quantile regression shows that the impact of flood risk is the largest for higher priced dwellings. Dwellings priced above the 70^{th} percentile are more negatively impacted by hazard risks, with those located at the 80^{th} percentile having negative impacts around 4.2%.

Another contribution of this work is related to capturing the heterogeneity of hazard zones, not only conditional on the variability of local amenities within the zones which may mitigate or exacerbate their impact, but also conditional on the relative location between zones across the city. Results show the impact of proximity to flood zones yield negative effects on housing prices which is mitigated when dwellings have greater accessibility to local urban green spaces. Urban green infrastructure has important implications for a city's storm water and flood management as they help absorb rainfall and localized riverine floods, preventing water from overwhelming pipes and pooling in streets or basements. The negative impact of being in flood zones is compounded by nearby lakes and increased impervious surfaces in the neighbourhood. Being in seismic risk zones on the other hand has marginally less of a negative impact conditional on the built and demographic characteristics of the neighbourhood, with mitigating effects coming from having more low-rise buildings and more owners or educated residents, which may serve as a signal for how well taken care the dwellings in the zone are.

Not only does the within variation of local amenities influence the relative effect of hazard risks on housing prices, but it is important to consider the location of the hazard zones across the city and their spillover effects to nearby areas. While being located in flooding zones yield negative average price effects, this is dependent on location in the city. Coastal properties have a net positive impact, even after controlling for the multitude of flood zones in these areas. This positive effect of being located near the Tagus river suggests that residents believe the amenity value of riverfront proximity outweighs potential flood risk and associated damages.

The structure of this chapter is as follows. Section 2 explores the previous literature of urban natural hazards with an emphasis on flooding and seismic events or risks. Section 3 outlines urban hazards in the context of Lisbon, Portugal, and presents the data. Section 4 describes the empirical strategy emphasizing the importance of considering spatial influences in the variable measurement and estimation. Results are presented in section 5 with final conclusions presented in section 6.

2 – 2. Literature Review

There is a range of literature on the impact that urban hazards and information plays into dwelling prices and the decision making process of residents. This can broadly be categorized into two types, some focusing on the impact of individual geohazard events and others accounting for the information regarding geohazard risk zones. The valuation of specific seismic or flooding events are common, however these likely underestimate the true impact if the potential reoccurrence of these events cause behavioral changes in other markets that are not accounted for. In general, results indicate that the real estate market responds negatively to potential risk due to urban natural hazards and prices are depressed in such zones of higher hazard risk, and can thus be used to inform on true economic impacts of these event risks.

Given the prominence of flooding events across the globe, a number of works have focused on the impact that flooding risk, flooding occurrences and flash floods have on the real

estate market. Many works value the *ex-ante* effect of how flooding risk information is taken into account and influences property values, however, there are further many *ex-post* studies which focus on the economic impact due to occurred flooding events.

In North Carolina, Bin and Landry (2008) and Bin and Landry (2013) highlight both the estimation of impacts from flood risk information and the impact of specific events. The average effect of being in a flood plain is an approximate discount of 7.3% when evaluated using a spatial error hedonic specification (Bin and Landry 2008). Using the events of major hurricanes in the area however, Bin and Landry (2013) estimate the impact of being located in flooding plains while controlling for spatial influences from positive amenity values related to water proximity. In particular, the authors use distance measures of proximity to the water in their estimations. Following recent hurricane and flooding events, the authors estimate that prices decreased 5.7% following Hurricane Fran and 8.8% following Hurricane Floyd.

Atreya and Czajkowski (2016) disentangle the countervailing impacts of flood risk and water-related amenities by interacting distance to the nearest coastline and flood risk to account for these impacts acting jointly on housing sale prices in Galveston County, Texas. They further vary flood return periods to allow for an interaction between negative and positive amenities related to proximity to water. The study shows that properties located in high-risk areas command a price premium up to 146% for up to nearly a quarter mile from the nearest coastline and the expected distance effects vary by flood risk type. In particular, housing premiums to higher risk homes decay at a faster rate the further one moves away from the water.

Rajapaksa et al. (2016) employ a difference-in-difference methodology to identify the impact of flood risk information and flood occurrences on housing prices in Brisbane, Australia. The authors conduct both OLS and spatial maximum likelihood estimation which accounts for significant spatial dependence in housing prices, making use of temporal variation in regards to when flood hazard information was made available and when actual floods

occurred. This work uses the impact of the release of flood risk map information in 2009 and the impact of actual flooding events which occurred in 2011, comparing the sale time of dwellings and their spatial exposure to each. Estimates indicate that the impact of public flood information being released led to price depressions in flood zones between 1 to 4% while actual flooding occurrences had impacts on prices by detracting between 18 and 19%.

Rajapaksa et al. (2017a) further explore whether the impact due to flood risk areas are conditional on locational or sub-market attributes. Using a spatial quantile regression under the difference-in-difference framework, the variation of risk impact conditional on areas of high value or low valued homes is explored. Flood risk is found to have the largest impacts on high valued property sub-markets in the range of a 4% to 8% decrease in prices while little to no significance was found to indicate that flood risk impacted lower value properties.

Also under a spatial quantile regression framework, Zhang (2016) find that in the North Dakota and Minnesota area, the average impact of being in a flood zone on dwelling prices is around 6% while lower-valued properties are more impacted than higher valued properties. Using the time of flooding events and time of sale, the author further concludes that the impact of a flood event depresses dwelling prices, but that this effect dissipates over time.

Along the lines of heterogeneity, Rajapaksa et al. (2017b) highlight the importance of considering proximity to major waterways in terms of flood risk capitalization. Using a semi-parametric model allows the authors to capture the non-linearity of impacts over space as they relate to proximity to the river. The estimated average impact of flooding zones is a depression of 5% of prices in these areas, however the benefits of being closer to the river outweighs potential risks of urban hazards and the effect is non-linearly related to proximity to the river.

Using similar methodologies the impact of seismic risk zones has been studied by exploiting discontinuities in time or over space as they pertain to known hazard zones, seismic activities and the impact on dwelling prices. While regular seismic activity is relatively

uncommon in any municipal area, unlike flooding occurrences, the housing market in the few areas with significant risk capitalizes on living in zones with higher potential damage.

With regular seismic activity and occurrences in Japan, much research has been dedicated to understanding the influence of hazard information on property values in this region. Naoi et al. (2009) find that the price impact of seismic zones on dwellings is larger following a large event and thus that the real estate market potentially undervalues the impact of these events until they occur. While seismic activity is negatively capitalized, the effect is heterogeneous according to local characteristics. The negative impact of seismic risk is influenced by dwelling characteristics, age and the local built environment (Naoi et al. 2010).

Using a two-dimensional spatial regression discontinuity, Hidano et al. (2015) value the difference between areas with high seismic risk or high risk of building collapse and their respective low risk areas. Under such approach, the authors find a dwelling premium for those that are located in low risk areas ranging from ¥13,970 to ¥17,380.

While natural seismic activity can be rare, new on land technologies have been introduced in recent decades that have exacerbated seismic events, primarily to facilitate oil and natural gas extraction. These induced seismic activities have become a standard by-produce of human ground interventions, and have important consequences on the real estate market in line with natural seismic occurrences. Metz et al. (2017) use the spatial and temporal variation of these events to identify the impact of drilling-induced seismic activity in Oklahoma. The effect translates to an approximate decrease in prices of 3.09% in affected areas, and the authors show that this result is robust using a number of spatial sub-setting of the data conditional on important locational characteristics. In the Netherlands, Koster and van Ommeren (2015) compare earthquakes felt by residents with those which have not been felt by residents to identify the impact of induced seismic activity. Dwellings which have experienced a noticeable earthquake sold on average at a price reduced by 1.9%.

2 – 3. Urban Hazard Risks and Measurement

2 – 3.1. Seismic and Flooding History of Lisbon

The name *Baixa Pombalina* is in reference to the *Pombaline* style of architecture and urban design, widely introduced in the area during the rebuilding which followed the Great Earthquake of 1755. Previously narrow medieval streets were rebuilt as open avenues, and buildings were constructed in a specific design to allow more flexibility in the case of a repeated seismic event. This caged-style building design is one of the first widespread architectural seismic policy initiative in a major urban area. Though a vast majority of the city was destroyed during the earthquake and subsequent tsunami, Portuguese history emphasizes this period due to the events importance as a catalyst to implement a wide range of frontier urban designs from the time in redesigning the city.

The devastation of the 1755 Earthquake is still evident in modern Lisbon through the stock of historically significant buildings, churches, palaces and amenities which have survived and remain a testimony to the past.[3] Estimates of the impact of this singular event on Portuguese history is massive, with a value ranging from 32% to 48% of GDP at the time, and a lower bound estimate of almost 23,000 completely destroyed or substantially damaged buildings (nearly 70% of Lisbon's dwellings at the time) and 30,000 – 40,000 lives lost by the combined effects of the quakes, fire and tsunami (Pereira 2009). Under current housing stock and urban design, an earthquake of that magnitude and characteristic would have an impact of approximately €11.4B, or 8% of GDP (Tang et al. 2012).

The 1755 earthquake was not, however, the first devastating earthquake to hit the city. The Tagus river follows a fault line, and large earthquakes can and do occur along it. In 1531, the city was hit hard when an earthquake along this fault struck the center of Portugal, northeast

[3] The Carmo Convent (located in *Baixa Pombalina*) for example now houses an archeological museum and stands out as a defining feature of the city skyline. The main drawing point of this site however is the partially destroyed arches and surviving pillars of the ancient church caused by the earthquake, which serves as a reminder to locals and tourists alike of the events destruction.

of Lisbon, with an estimated magnitude of M7. There was also a severe earthquake in 1321, again with widespread destruction, and other significant quakes occurring 1147 (also leaving the city, just captured from the Moors, in ruins), 1334 (destroying the cathedral roof) and 1356. In 1909 a quake hit north-east of Lisbon with a magnitude of M6.5, however since then the region has been seismically calm. There have been, however, occasional earthquakes of much lower magnitude briefly felt in the city. Recent examples include the M4.3 quake in the region of Sobral de Monte Agraço on August 17th, 2017, and the M6 earthquake located near the St. Vicent Canyon, offshore to the south-west coast of Portugal on December 17th, 2009.

While there have been many discussions of the chances of a re-occurrence of the 1755 event, such an event may only occur every 5,000 years or so, leaving imminent danger out of the mindset of the local population. However, a more immediate danger is a potential repeat of those medieval quakes in the near future. In addition, the more critical source of concern is the Lower Tagus Valley region which could produce an M6 to M7 earthquake with a return period as short as 150 to 200 years. This proximate seismic zone, combined with the city's large number of old masonry buildings and a fraction of reinforced concrete frames designed with limited lateral resistance, presents the most significant potential for large scale loss.

The city's intimate history with earthquakes, and awareness of their significance and destructive capacity, has thus become ingrained in the current culture. The residential population is therefore aware of the inherent risks that come with living in certain areas of the city with higher potential hazard occurrences and subsequently larger damages. While seismic activity tends to be few and far between, flooding is a regular occurrence in the city.[4] Yearly, during the rainy winter months, many parts of Lisbon experience sometimes severe flooding.

[4] Since 2007 Portugal has experienced twelve seismic events of varying magnitude (four in the Lisbon region), while in 2014 alone there were a total of 1,336 flooding incidents reported throughout the city.

The rugged topography of the city means that these flooding occurrences are not limited to the riverfront and we thus observe heterogeneity of this urban risk across the city.

Catastrophic flooding events are not frequent, however, from time to time Lisbon suffers extreme meteorological events such as heavy rains causing flash floods and landslides. The two most well-known extreme rainfalls include the flooding of November 25th, 1967, and February 18th, 2008. In particular, the floods of 1967 claimed 464 lives, making it the worst natural disaster to hit Portugal since the earthquake of 1755 and the fourth deadliest flash flood in world history. The severity of flooding events occurring in the city is expected to escalate with the rising of sea level and more severe rainfall patterns due to climate change.

The risk of these hazard events therefore are likely to be realized by residents at some point in their lifetime and act as dis-amenities to specific areas of the city where the risks and potential for damages are greatest. The value of these urban dis-amenities can be estimated by geo-locating dwellings inside of these zones and comparing prices while controlling for other important locational features which vary across the city.

Figure 5. Urban Hazard Risks in Lisbon, Portugal

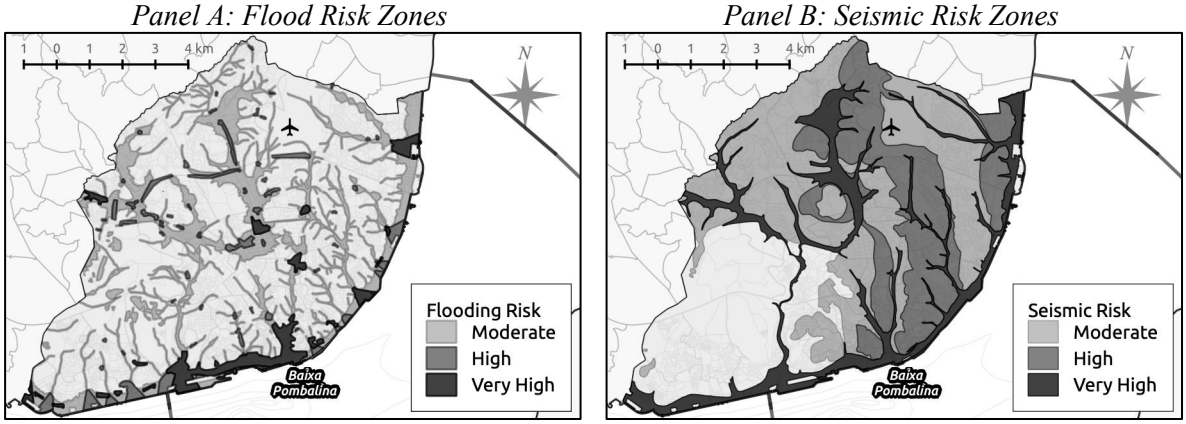


Figure 5 shows Lisbon's flood and seismic risk zones maps. Much of the flooding and seismic risk zones appear near the Tagus riverfront or are linked to the city's topography, and thus distance to the river and elevation are important aspects to take into consideration along with urban hazards. About 6% of the city is zoned under the high or very high flood risk categories, while 24% of the city falls into the high or very high seismic risk categories. While

different properties across an earthquake prone area are equally likely to experience the same large earthquake, there is significant local variation in the likely damage that would result from that earthquake over small distances due to variation in soil geology. In terms of the number of dwellings, around 45% of the homes in the data set are in high and very-high seismic risk zones while 14% of the homes are in high and very-high flood risk zones. There are 11% of the homes located within areas with high risk of both hazards.

2 – 3.2. Data and Sources

Residential property data from 2002 to 2007 is obtained from *Confidencial Imobiliário*.[5] While year fixed effects are included to capture variability in housing prices over time, it is important to note that the data is cross-sectional and without repeated observations. The database contains list price, structural characteristics and location identifiers for 32,420 dwellings which allow for the assignment to any neighbourhood or hazard zone.

The municipality of Lisbon, *Câmara Municipal de Lisboa*, maintains a wide range of publicly available geo-referenced data regarding local amenities, ecological characteristics, transportation infrastructure, urban hazards and the city's built environment. The municipal urban planning strategy (*Plano Diretor Municipal*) provides the size and location of urban hazard zones in the city, which are classified hierarchically based on the potential risk to residents as seen in figure 5. Areas of the city can be categorized according to the severity of risk (moderate, high or very high) of flooding or seismic risk.

In the case of seismic risk, studies from the city highlight these areas according to soil quality and type, fault lines, topography and the potential for damages in terms of the built and population density. Seismic studies led to the development of the risk map in 2001, informing the public of these locations (Instituto Superior Técnico 2005). Given the regular flooding,

[5] 32,420 observations 65.9%, or 21,353 observations, are from 2007 with 0.3%, 0.9%, 2.4%, 7.5% and 23.1% of the data from 2002 to 2006 respectively.

areas of flooding risk are easily determined and heavily influenced by the slope and elevation. Online platforms and active local authorities make it easy for residents to report flooding occurrences with the municipality keeping a logged archive of these floods for recent years.[6]

The locations of urban amenities in Lisbon come from the city's open data platform. Using these geospatial databases, we can measure proximity to employment centers and other areas of the city such as the riverfront, determine proximity to and neighbourhood concentrations of transportation infrastructure and spaces such as parks or forests. Census 2011 data further captures neighbourhood level socio-demographic and building stock variables.

Table 4. Key Descriptive Statistics

	N	Mean	St. Dev.	Min	Max
Price (€)	32,420	243,500 €	148,503	25,000	2,500,000
Locational Characteristics					
Located 3 km from the Riverfront	32,420	0.480	0.500	0	1
Located 100 m from Tagus Riverfront	32,420	0.007	0.084	0	1
Located 500 m from Tagus Riverfront	32,420	0.116	0.320	0	1
Located 500 m outside Very High Seismic Risk Areas	32,420	0.633	0.482	0	1
Located 500 m outside Very High Flood Risk Areas	32,420	0.408	0.491	0	1
Open Spaces and Ecological Urban Hazards					
Located in High or Very High Seismic Risk Area	32,420	0.454	0.498	0	1
Located in High or Very High Flood Risk Area	32,420	0.140	0.347	0	1
Located in Very High Seismic Risk Area	32,420	0.196	0.397	0	1
Located in Very High Flood Risk Area	32,420	0.062	0.242	0	1

The *Agência Portuguesa do Ambiente* maintains data on the elevation and topographical profile of the city. The measure of elevation also conveys flood risk by measuring altitude relative to the sea. Data on greenness, including tree canopy and normalized difference vegetation index (NDVI) is obtained from chapter 1. Descriptive statistics for key variables are presented in table 4 with all variables presented in table A2 – 1 of the appendix.

2 – 4. Empirical Analysis

2 – 4.1. Spatial Hedonic Specification

The empirical specification is built on the seminal work of hedonic valuation from Rosen (1974). If a household has preferences across local amenities, the implicit value of these

[6] Geo-referenced data on reported flood locations is only available for recent years since 2011, and thus cannot be merged with dwelling observations from 2007.

amenities can be imbedded in real estate prices. A hedonic estimation is able to decompose the price of a dwelling into its value bearing attributes. Using a range of geospatial databases, the standard set of covariates for each dwelling are enhanced with neighbourhood amenities, locational attributes, and the areas ecological and topographical profile.

The baseline hedonic model is extended to consider potential important underlying spatial dependence. Such dependence in the data may influence the estimation associated with dwelling prices which are closely related to their neighbors or commonly influenced by omitted neighbourhood characteristics. The most general form of the framework decomposes housing prices into its value bearing attributes as follows:

$$P_{it} = X_i\beta_1 + H_i\beta_2 + S_i\beta_3 + T_t\beta_4 + D_H \cdot A_i\beta_5 + H_i \cdot L_i\beta_6 + \rho\mathbb{W}P_{it} + u_{it}$$
$$u_{it} = \lambda\mathbb{W}u_{it} + \varepsilon_{it} \;\; ; \qquad \varepsilon_{it} \sim iid(\mathbf{0}, \sigma^2 \mathrm{I_n}) \qquad \text{Eq. 2 – 1}$$

where log price, P_{it}, for an observation at location i and time t is decomposed into a vector of time-invariant covariates, X_i, including the constant and a range of structural characteristics or neighbourhood attributes. When appropriate, concentrations of local neighbourhood attributes are calculated using individual buffer radii surrounding each dwelling so as to limit any influences from the modifiable aerial unit problem. Census tract neighbourhood characteristics are measured as the area weighted concentration within 500 meters of a dwelling. The variable of interest is captured in the vector H_i which includes a dummy variable for whether a dwelling is located in an area of respective flood, seismic or a jointly hazardous zone, as well as spillover effects occurring to directly adjacent dwellings.

Fixed effects for both year, T_t, and space, S_i, are included. S_i is used to mitigate potential biases due to omitted locational factors resulting from time-invariant unobserved neighbourhood characteristics that contribute to dwelling prices, however, it is important to consider the scale of these units in order to appropriately capture these underlying influences. The introduction of spatial fixed effects using administrative boundaries (e.g. civil parishes or

freguesias) may provide an inconsistent definition of one's neighbourhood and inaccurately capture the influence from the concentration of local amenities or the locational realities at an observation point. The size of administrative *freguesia* boundaries varies greatly and using these units as controls for a location in the city may not be refined enough to be effective. Moreover, if dwellings are on the edge of the spatial unit, they may further receive some spillover from a neighboring unit's unobserved characteristics. As in chapter 2 and Franco and Macdonald (2018), spatial fixed effects S_i are introduced according to a dwellings location in a constructed 500 meter by 500 meter grid superimposed over the city. This improves the model specification by more accurately capturing very localized potentially omitted spatial influences while addressing the modifiable areal unit problem.

Including a large range of spatial fixed effects at such detailed resolution is feasible given the size of the data, and further relegates neighbourhood and locational characteristics to these spatial controls. This limits the need to control for an abundance of locational characteristics such as distance to all types of local urban amenities or proximity to business districts. An additional benefit of this methodology is the reduction of multicollinearity which may come from controlling, for example, for distance to the main city center and important urban amenities which may be located in this area such as the river.[7]

Although spatial fixed effects capture omitted location influences across the city, spatial dependence in prices or the error term of the models may have significant effects if the chosen fixed effect units do not accurately reflect or align with the underlying data generating process (Anselin and Arribas-Bel 2013). It is therefore important to test and incorporate, where necessary, spatial dependence in the form of either the spatially lagged dependent variable, $\mathbb{W}P_{it}$, with coefficient ρ, or modeling the error u_{it} as an autoregressive error term accounting for spatial correlation, $\mathbb{W}u_{it}$, with coefficient λ.

[7] To ensure no multicollinearity, the variance inflation factor (VIF) for the estimates of interest is below 10.

The general econometric specification in equation 2 – 1 nests multiple spatial models where the chosen empirical models are decided on via the results of the spatial diagnostics. From this specification, when $\rho = 0$ we have a spatial error model and with $\lambda = 0$ a spatial autoregressive model. $\mathbb{W}$ represents the $n \times n$ weight matrix defining the extent and strength of spatial spillovers between dwellings. Six weights are used, ranging from quite local to more spatially broad, to ensure estimates are not the product of the chosen matrix. These spatial weights include an inverse distance weight for all neighbors within 500 meters (SW1); an inverse squared distance weight for all neighbors within 500 meters (SW2); a binary weighting schemes to indicate all neighbors within 500 meters (SW3); a binary weighting scheme to indicate all neighbors within 100 meters (SW4); neighbors based on the 100 nearest dwellings (SW5); and neighbors based on the 10 nearest dwellings (SW6). Table A2 – 2 of the appendix summarizes the properties of these weights.

2 – 4.2. Identification and Robustness of Results

Although the estimation of spatial hedonic models may alleviate estimation biases, it does not address concerns regarding the identification of impacts. The locations of hazard zones are exogenously determined via ecological and topographical processes, and thus not conditional on dwelling prices, however there may exist some significant underlying locational influences driving the estimated impacts near these areas.

The robustness of the estimates are checked via a geographic regression discontinuity (GRD) framework, a type of regression discontinuity with a geographic treatment assignment comparing treated (hazard prone) properties to valid control properties nearby but not in a hazard risk zone.[8] The methodology behind the GRD framework and choosing valid treated

[8] The GRD methodology has previously been employed to study area wide impacts from media market zones on political turnout (Keel and Titiunik 2015), police surveillance zones on crime (MacDonald et al. 2016) and historic conservation areas on property values (Franco and Macdonald 2018). One of the prerequisites for a GRD is to identify the geographic boundary where a discontinuity exists in how the treatment is assigned. We use municipal defined boundaries representing the locations of high flood risk or high seismic risk as our regression

and control properties in nearby geographic sub-sets is highlighted in figure A2 – 1 of the appendix for the example of flood risk areas.

The data is homogenized by sub-setting around hazard zone boundaries. In the relevant geographic subset, all other locational attributes such as accessibility to the CBD, accessibility to the riverfront, and other natural amenities are likely to be similar and thus provides a quasi-experimental design. The locational similarity of properties in all other aspects will be greater for bordering dwellings located at the geographic boundary that separates the two zones.

We test the use of a variety of spatial sub-setting thresholds, conditional on the hazard zone of interest which include sub-setting to all properties within 500 meters from very high risk flood zones and 500 meters from very high risk seismic zones. We further subset the data conditional on being located within 3 kilometers of the Tagus riverfront, and estimate our parametric GRD hedonic pricing model again to ensure that the impacts from these hazard zones for example are not being driven by properties being located near to the Tagus river.

While this methodology ensures that the estimated impacts come from a relatively homogeneous set of observations from which potentially important neighbourhood effects are not driving the results, it may be the case that relative location along the hazard zone boundary also has significant variability. To compensate, a propensity score match is used to examine distance as defined by covariates and homogenize the data by finding controls conditional on important locational characteristics.

Figure A2 – 2 of the appendix highlights the geographic discontinuity in prices occurring at either hazard boundary. These figures suggest that dwellings within hazard zones sell at a lower price with dwelling prices increasing as we move from these boundaries. The discontinuity at the flood risk boundary suggest some indication of higher prices the further

discontinuities, where the treatment jumps discontinuously along these geographic borders. Another prerequisite is that we compare similar properties in the control and treatment groups on either side of the geographic boundary and that enough variability along that boundary exists.

inside very high risk flood zones. Since the majority of the very high risk flood zones are located along the Tagus river however, this figure highlights the importance of distinguishing between the negative zonal effect of being in an area of higher flood risk and the positive amenity value of being closer to the riverfront.

Placebo Tests of GRD Hazard Boundary Effects

A set of placebo tests are used to make sure the identification strategy is capturing the true effects reported. If the hypothesis is that hazard zone boundaries have significant impacts on the local real estate, a placebo test applies this hypothesis on a subset of data where this treatment boundary effect is known to be zero. We follow the methodology of Rischard et al. (2018) to identify placebo treatment and control observations within this data's setting.

The first requirement is to consider observations which are located in high risk zones and those which are not separately. By estimating a boundary treatment effect on observations from only high risk properties or from only non-risk properties, we eliminate heterogeneity in this respect and thus would expect the placebo results to return null effects. For robustness, we consider all subsets of the dataset according to any combination of high or very high seismic or flooding risk. Once the data is split into high risk and non-risk properties, a number of constructed placebo boundaries are introduced to test whether the GRD framework captures other underlying locational effects.

The placebo threshold boundaries are drawn to mimic a random pattern splitting an area into a new treated and non-treated set. The boundaries are drawn in a zig-zag pattern to increase the randomness and distribution of properties into our placebo treated and non-treated categories. If straight lines were used there is the risk that this would align with topographical aspects of the city, such as the location of valleys or other clearly defined sub-regions, and thus the placebo boundaries would capture this effect. Figure A2 – 3 of the appendix shows a map of Lisbon with various of the constructed placebo boundary thresholds introduced.

This exercise is repeated various times, bisecting the study region with constructed zig-zag boundary delineations to define placebo treated and non-treated observations in different ways. With the placebo boundary introduced, a GRD is built around the boundary using cut-off thresholds of 500 meters, 1 kilometer and 3 kilometers. We should not expect the GRD methodology to pick up significant effects across the introduced boundary, given the randomness in the assignment of local placebo treated and non-treated observations.

We estimate the effect under various GRD specifications with differing thresholds and placebo boundaries. All specifications include the same structural characteristics, year fixed effects and location fixed effects and covariates as in the fully estimated models. Table A2 – 2 of the appendix shows that no significant estimated effect is observed using the constructed thresholds, highlighting three of the placebo boundaries that bisect the study region and cut the municipality in half. Results from different boundaries and subsets are all in line and show no significant estimated effect. This indicates that our GRD framework, focusing on properties straddling the boundary of either type of hazard risk zone, is not identifying spurious effects of the zone's impact on housing prices.

2 – 4.3. Within and Between Heterogeneity of Urban Hazard Zones

This research concerns itself with capturing not only the direct effect of how urban hazard zones impact residential property prices, but further how this effect is conditioned on the heterogeneity of local amenities found across the different zones, the relative location of these hazard zones across the city, and how the impact of these zones may depend on the distribution of dwelling prices.

The baseline empirical model is thus estimated with and without the inclusion of the terms $D_H \cdot A_i$ which represents the within variation, and $H_i \cdot L_i$ which represents the between variation. Here D_H measures the accessibility to a hazard zone as the distance to the border of

the nearest hazard risk, while A_i represents a range of local amenities, and L_i represents a dummy for being located in notable areas of the city.

Since we measure the impact from flooding risk and seismic risk through proximity of the dwelling to the boundary of hazard zones, dwellings located inside these zones and in particular if located at the center of the zone, are exposed to larger external effects. However, as we move towards the boundary (if located inside) or away from the boundary (if located outside) we expect the negative effects of these dis-amenities to decline. Hence the measure D_H captures the relative strength of being located further inside either of these hazard areas. In particular, dwellings located outside of a hazard area take a positive geographic distance value while those located inside a hazard area take a negative geographic distance value.

On the other hand, the within variation is interpreted relative to the global average impact of being in a hazard zone, H_i. We estimate the average marginal impact of being closer to a hazard zone border as conditioned by local amenities. A positive estimate of β_5 would therefore signal marginally higher house prices further away from a hazard zone conditional on higher levels of the local amenities, and thus an exacerbating effect, while negative values would indicate a mitigating influence of local amenities with marginally higher prices for dwellings closer to hazard zones conditional on higher levels of the local amenities. If residents trade off the urban dis-amenity value from hazard zones according to benefits coming from other local urban amenities, we expect there to be variation in the estimate of β_5 according to the type of local amenities considered.

This measure of variation in H_i focuses on how the heterogeneity of a dwelling's local amenities can mitigate or exacerbate the impact of proximity to these zones. We focus on how the interaction between proximity to hazard zones and urban green areas, the local built environment or neighbourhood crime levels influence a dwellings price. If these local amenities or dis-amenities can be used to make an otherwise risky zone marginally better, or

worse, then we expect this interaction to capture this effect and the residential real estate market to respond.

The between variation captures how the relative location of these hazard zones in a city can influence their impact on housing prices. If housing prices are representative of a bundle of attributes, then residents may give more or less importance to urban hazard zones if the local area compensates in other aspects. While the location of flood and seismic risks are in many cases concentrated along the river and in the downtown core, the amenity value of proximity to the river may outweigh the negative dis-amenity of being in a flood zone. Interacting dummy variables to represent whether a dwelling is located in a hazard zone while simultaneously in direct proximity to the riverfront would thus capture this potential impact representing the relative costs and benefits that residents accept by living in certain areas of the city.

This hazard indicator is further interacted with dwellings which are simultaneously located in historically preserved conservation areas which are also located along the riverfront and in historically important areas of the city. These historic areas of the city are preserved for their historic charm and the combination of aesthetically pleasing buildings, open spaces and neighbourhood allure. If these areas are more preserved relative to other areas nearby, then the benefit of living here may outweigh the cost of being in a zone of increased urban risk. In the case of these between hazard zone effects, it is important to consider the net effect coming from $\beta_2 + \beta_6$ when $L_i = 1$.

2 – 5. Results

The baseline results focus on three categories of models of urban hazards for flooding risk, seismic risk and jointly hazardous areas (simultaneously in a zone of flooding and seismic risk), each introduced separately so as not to introduce conflicting impacts in the effects on housing prices. All models include structural characteristics with magnitudes in line with previous literature and all providing positive price effects on dwellings, the largest impact

coming from its size with a price effect of 0.86% per percentage increase in square meterage, and more luxury type amenities such as pools, air conditioning or a view of the Tagus river drawing premiums of 15.9%, 12.5% and 5.8% respectively.

All models are tested for spatial dependence of the dependent variable and spatial autocorrelation of the error term. Diagnostic results from the spatial tests are presented in table A2 – 3 of the appendix. Global results on the Moran's I statistic indicate significant spatial dependence influencing the estimates which should be accounted for. Using the Lagrange multiplier (LM) tests to better identify the source of the spatial influences indicate that a SEM is appropriate to control for the underlying spatial autocorrelation in the error term, while no significant spatial influence is found from the spatially lagged dependent price variable. Therefore, according to our empirical model specification presented in section 4, we conclude that $\rho = 0$ and there is no significant spatial lag effect, while λ is a significant parameter and a SEM specification should be estimated. Results from a spatial Breusch-Pagan test indicate the presence of heteroscedasticity, and thus robust standard errors are presented.

Using the AIC model selection criteria, SEM models outperform their OLS counterpart and further all have reduced sum of squared errors (SSE). Diagnostics suggest, based on a combination of the AIC, SSE, robust LM tests and variable significance, the preferred model is the SEM with weight matrix using all properties within 100 meters as neighbors (SW4), and subsequent analyses use this specification. We note also that SEM coefficients of hazard covariates are smaller than in OLS models, showing the bias induced by not controlling for spatial autocorrelation.

2 – 5.1. Flood and Seismic Risks Average Price Impacts

Spatial hedonic results presented in table 5 indicate that the residential real estate market negatively capitalizes on hazard risks. All specifications include an interaction effect between elevation and distance to the riverfront so as to ensure that the price impact for location

in a hazard risk zone is not biased by these important locational features. Often homes located in high risk zones are also the most desirable in terms of their proximity to the water. Thus it is important to allow for the joint impact of the potential negative housing price effects of higher risk (whether flood, seismic or both) and the potential positive housing price effects of living close to the water. Moreover, there is an inherent variation in flood risk within a given region that can be associated to the elevation in relation to the sea. This measure of elevation has its zero value at the level to which stormwater flows, and from where water would pool. Thus, elevation is used in relation to the sea-level in addition to a flood risk indicator variable to account for the spatially inherent variation of topography within a risk zone.

While dwelling prices are positively influenced by being at a higher elevation, this effect is stronger the closer a dwelling is to the riverfront. The negative coefficient estimate on the interaction term indicates that dwelling prices increase as the distance to the river decreases, and this effect is stronger for dwellings at higher elevations. This may be associated with better access to water-related amenities and views while having a lower flood risk from being located at higher elevation in relation to the base flood elevation level. The model further controls for average distance to all parks and gardens in the city as a form of concentration of green spaces, as well as the number of urban forests nearby, with results indicating that these green amenities are also positively valued by residents.

The per dwelling price impact of being located in a designated very high risk flood zone (model 1) is a decrease of approximately 3.5%. For an average priced dwelling, this corresponds to an approximate price discount of €8,500. The price impact due to flooding risk is the largest taking into consideration the other different urban hazards studied. This is likely due to the fact that flooding is a common occurrence in the city and happens yearly. Residents looking to purchase or sell their dwelling are well aware of the flooding risk of their neighbourhood, given the nearby slopes, elevation, whether it is in a valley or near the river.

Table 5. Estimated Impacts of Urban Hazard

Dep. Variable: ln(Price)	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Nº. of Urban Forests in 500 m	0.01540**	0.01520*	0.01535**	0.01587**	0.01590**	0.01548**
	(0.00781)	*(0.00780)*	*(0.00780)*	*(0.00781)*	*(0.00781)*	*(0.00781)*
ln(Average Distance to all Parks)	-1.19597***	-1.17027***	-1.18900***	-1.15547***	-1.18652***	-1.16523***
	(0.16812)	*(0.16747)*	*(0.16861)*	*(0.16752)*	*(0.16805)*	*(0.16744)*
Elevation	0.00130***	0.00139***	0.00149***	0.00156***	0.00143***	0.00142***
	(0.00028)	*(0.00028)*	*(0.00027)*	*(0.00027)*	*(0.00027)*	*(0.00027)*
ln(Distance to Tagus River)	-0.03037*	-0.02392	-0.02499	-0.02109	-0.03000*	-0.02578
	(0.01769)	*(0.01736)*	*(0.01767)*	*(0.01733)*	*(0.01771)*	*(0.01732)*
Elevation × ln(Distance to Tagus River)	-0.00054**	-0.00062**	-0.00064**	-0.00067***	-0.00056**	-0.00062**
	(0.00026)	*(0.00025)*	*(0.00025)*	*(0.00025)*	*(0.00026)*	*(0.00025)*
	Flood Risk		***Seismic Risk***		***Joint Hazards***	
	Very High	***High***	***Very High***	***High***	***Very High***	***High***
Urban Geohazard Risk	-0.03513***	-0.01612**	-0.01114*	-0.01118*	-0.03786***	-0.02474***
	(0.01128)	*(0.00731)*	*(0.00647)*	*(0.00631)*	*(0.01419)*	*(0.00802)*
Lambda	0.15503***	0.15801***	0.16021***	0.16128***	0.15532***	0.15600***
	(0.03744)	*(0.04114)*	*(0.01360)*	*(0.02077)*	*(0.03940)*	*(0.03389)*
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
500 m F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Structural Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
AIC	-3990.9	-3984.6	-3982.7	-3982.9	-3987.6	-3989.0
$AIC_{SEM} \div AIC_{OLS}$	*1.00500*	*1.00520*	*1.00540*	*1.00550*	*1.00500*	*1.00510*
Log Likelihood	2241.5	2238.3	2237.3	2237.4	2239.8	2240.5
$L.L._{SEM} \div L.L._{OLS}$	*1.00490*	*1.00510*	*1.00520*	*1.00530*	*1.00490*	*1.00500*
SSE	1652.5	1652.8	1652.9	1652.9	1652.7	1652.6
$SSE_{SEM} \div SSE_{OLS}$	*0.99897*	*0.99897*	*0.99891*	*0.99891*	*0.99903*	*0.99897*
Residual Std. Error	0.22577	0.22579	0.22580	0.22580	0.22578	0.22578
$Res.\ Error_{SEM} \div Res.\ Error_{OLS}$	*0.99572*	*0.99572*	*0.99572*	*0.99572*	*0.99572*	*0.99572*
Adj. VIF for Hazard Variable	1.99	2.01	2.07	2.54	2.10	2.04
Spatial Breusch-Pagan	1208.6***	1205.2***	1206.9***	1208.1***	1207.5***	1208.3***
Wald Test	17.15***	14.75***	138.87***	60.28***	15.54***	21.18***
Likelihood Ratio Test	21.70***	22.70***	23.28***	23.65***	21.78***	22.07***
Observations	32,420	32,420	32,420	32,420	32,420	32,420

*Notes: ***Significance at 1 % level; **Significance at 5 % level; *Significance at 10 % level.*
Heteroskedastic consistent errors

The impact on housing prices for being located in designated flood risk zones is sensitive to the strength of the risk, with very high flood risk zones yielding significantly larger impacts than the more dispersed combination of high or very high flood risk zones, with negative impacts of 3.5% and 1.6% respectively (model 1 and model 2). This suggest that on average residential prices reflect differently to the relative variability and strength of flooding risk areas across the city.

Seismic risk on the other hand, yields smaller magnitude price discounts in the order of 1.1% (model 3) with little difference between the impact of being located within a designated very high seismic risk zone relative to being located within a high risk or very high seismic risk combined zone. Seismic activity in Lisbon is quite rare. Even if a property is located near a fault line and the potential for damage is catastrophic, the low magnitude of the estimate

likely stems from the undervaluation of seismic risk that residents have given the scarcity of these events and low chances of any occurrence on the same magnitude as in the past, even if located in a seismic zone. Evaluated at an average priced dwelling, the impact of being within designated seismic risk zones on property values detracts from prices by approximately €2,700.

Although these price estimates are relatively conservative, given that they are per-dwelling effects and a potentially large number of dwellings are located in these areas, the total effect of urban hazards on the residential market of Lisbon is potentially quite large. Within our sample, 6.2% of the homes are located in an area of very high flooding risk while 19.6% of homes are located in areas of very high seismic risk. The aggregate effect across all dwellings exposed to these risks therefore is large.

Given the heterogeneity and overlap of these urban risks, it is possible to examine the impact of being jointly in designated areas of flooding risk and seismic risk. Dwellings located in both types of very high risk zones have a negative impact on prices on the order of 3.8%, or around €9,250 evaluated at the value of an average priced dwelling (model 5). While the market responds to urban natural hazard risks, there seems to be heterogeneity across how the risk from different types of natural hazards are capitalized into dwelling prices.

Hazard Risk Zone Spillover Effects

The spillover impacts of hazard zones are presented in table 6. While there is a negative price impact of being located in a hazard zone, this effect is not restricted to the boundaries of the hazard zone itself. Results suggest that the negative effect of flood risk zones extend beyond the boundary of the zone and impact properties adjacent and within 50 or 100 meters of the boundary as well. For dwellings located just outside a very high risk zone (whether in terms of flood, seismic or both), there is a negative effect on price of approximately 1.5%. This effect is driven by properties that are located adjacent to very high risk flood zones and simultaneously located in a non-risk area (model 7 and 8).

This suggests that nearby dwellings capitalize on the dis-amenity value of being located near to high risk flood zones, with no significant impact coming from seismic zones. Flooding is regular in the city and the path of water runoff is not limited to any boundaries, and we would expect these negative direct spillovers to occur for flooding events and not for seismic events.

Table 6. Hazard Zone Spillover Effects

Dep. Variable: ln(Price)	**Model 7**	**Model 8**	**Model 9**	**Model 10**	**Model 11**	**Model 12**	**Model 13**
In V. High Risk Flood Zone	-0.04424***	-0.05174***	-0.04533***	-0.05279***			
	(0.01259)	*(0.01344)*	*(0.01258)*	*(0.01326)*			
50 m from V. High Risk Flood Zone	-0.01422*	-0.01132	-0.01414*	-0.01197			
	(0.00806)	*(0.00820)*	*(0.00805)*	*(0.00810)*			
In V. High Risk Seismic Zone					-0.01048	-0.00588	-0.0115
					(0.00911)	*(0.00933)*	*(0.00912)*
100 m from V. High Risk Seismic Zone					0.00072	0.00836	0.00500
					(0.00709)	*(0.00799)*	*(0.00742)*
100 m from V. High Risk Flood Zone × In a Non-Flooding Risk Zone		-0.01496*					
		(0.00889)					
100 m from V. High Risk Flood Zone × In High Risk Flood Zone			-0.13299*				
			(0.07532)				
100 m from V. High Risk Flood Zone × In High Risk Seismic Zone				-0.04200**			
				(0.01842)			
100 m from V. High Risk Seismic Zone × In a Non-Flooding Risk Zone						-0.01593**	
						(0.00803)	
100 m from V. High Risk Seismic Zone × In High Risk Seismic Zone							-0.02150*
							(0.01098)
Lambda	0.15404***	0.15312***	0.15280***	0.15429***	0.16035***	0.16116***	0.15891
	(0.02346)	*(0.03819)*	*(0.02619)*	*(0.02090)*	*(0.03373)*	*(0.02915)*	*(0.10038)*
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
500 m F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AIC	-3991.9	-3992.8	-3991.6	-3995.9	-3980.7	-3982.2	-3982.3
$AIC_{SEM} \div AIC_{OLS}$	*1.00490*	*1.00480*	*1.00480*	*1.00490*	*1.00540*	*1.00550*	*1.00530*
Log Likelihood	2243.0	2244.4	2243.8	2246.0	2237.4	2239.1	2239.1
$L.L._{SEM} \div L.L._{OLS}$	*1.00480*	*1.00470*	*1.00470*	*1.00480*	*1.00520*	*1.00530*	*1.00520*
SSE	1652.4	1652.2	1652.3	1652.1	1652.9	1652.7	1652.7
$SSE_{SEM} \div SSE_{OLS}$	*0.99903*	*0.99903*	*0.99903*	*0.99903*	*0.99891*	*0.99891*	*0.99891*
Residual Std. Error	0.22576	0.22575	0.22576	0.22574	0.2258	0.22578	0.22578
$Res.\ Error_{SEM} \div Res.\ Error_{OLS}$	*0.99572*	*0.99572*	*0.99572*	*0.99572*	*0.99572*	*0.99563*	*0.99568*
Mean Adj. VIF for Hazard Variables	1.82	1.83	1.73	1.81	2.54	2.38	2.33
Max Adj. VIF for Hazard Variables	2.23	2.39	2.24	2.33	2.94	3.04	2.94
Spatial Breusch-Pagan	1212.2	1212.5	1212	1212.3	1207.2	1210.5	1206.7
Wald Test	43.11***	16.07***	34.03***	54.51***	22.60***	30.57***	2.51
Likelihood Ratio Test	21.35***	21.12***	21.18***	21.43***	23.30***	23.62***	23.00***
Observations	32,420	32,420	32,420	32,420	32,420	32,420	32,420

*Notes: ***Significance at 1 % level; **Significance at 5 % level; *Significance at 10 % level.*
Heteroskedastic consistent errors

These spatial spillover results also reveal that hazard zones compound each other and that being located in areas of high concentrations of either types of hazards impact housing prices. When a dwelling is located in a high flood risk zone which is adjacent to a very high flood risk zone, it is surrounded by these risk areas and residents perceive the combined effect of these zones together. Prices in high risk zones that are directly adjacent to very high risk

zones are detracted by approximately 13.3% (model 9), indicating that being in and around many flood-prone areas provides even higher dis-amenity values. These compounded spillovers also occur between flood risk and seismic risk zones with negative impacts of 4.2% for dwellings adjacent to very high risk flood zones which are simultaneously located in high risk seismic zones (model 10).

The direct spillover effect from seismic zones is less pronounced than flood risk zones, yet suggests that being adjacent to very high risk seismic areas while simultaneously in high risk seismic areas has a compounded negative price effect of 2.2% (model 13). Being adjacent to a seismic zone which has no risk of flooding further indicates a negative price impact of 1.6% highlighting that these dis-amenity values are not constrained directly to the boundaries of the zones and, dwellings located nearby, even without and direct risk themselves, are further subject to the impact stemming from natural hazards.

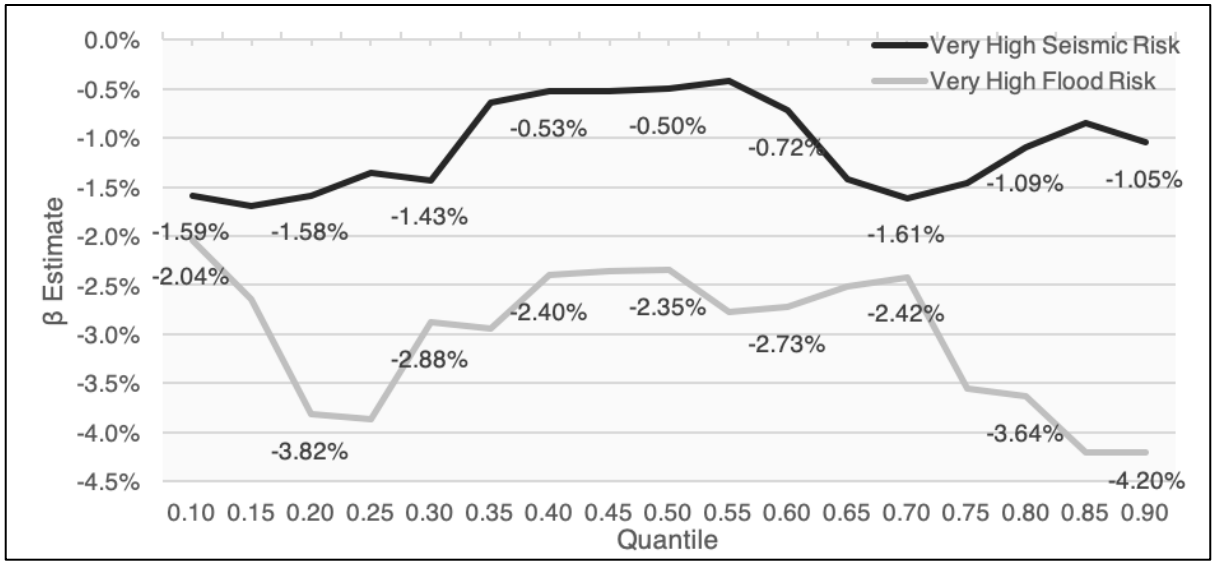
2 – 5.2. Hazard Risks Quantile Price Effects

To capture the potential impact of natural hazards conditional on the distribution of dwelling prices, a quantile regression for model 1 and model 3 is estimated with results plotted in figure 6. This highlights whether specific portions of the distribution of prices are more or less impacted from the very high risk of hazards. Dwellings at different points in the distribution of housing prices may have coefficient values which vary from the average if properties are inherently more susceptible to these risks or sensitive to the dis-amenity value.

Results for very high flood risk areas suggest that dwellings at the higher end of the distribution, above the 70^{th} percentile, are more negatively impacted by these hazard risks. For these priced dwellings, large floods have the potential to have more relative damaging costs and residents in such properties capitalize more on these perceived risks and costs. For a resident in a higher priced dwelling located in a very high flood risk, their potential for loss is greater than for cheaper dwellings. Above the 85^{th} percentile, the impact on dwelling prices

increases significantly to approximately 4.2%, compared to the average value estimated under model 1 at 3.5%. At the 10th percentile of dwelling prices on the other hand, the impact of flood risk reduces to 2.0%. While flood risk impacts differently dwelling prices at different ends of the distribution, we find little evidence that seismic risk has similar heterogeneous effects.

Figure 6. Very High Flood and Seismic Risk Quantile Estimates



2 – 5.3. House Price Response to Hazard Risks Conditional on Other Urban Features

Within Variation of Hazard Risks

If residents value differently hazard zones conditional on local amenities, then we would expect there to be within variation in very high flood risk or very high seismic risk areas. It is important to capture this heterogeneity across areas conditional on their local context to better understand the interaction of hazard zones with broader municipal infrastructure and amenities. Model 1 and model 3 are estimated with a range of local dwelling and neighbourhood attributes which relate to each risk with results presented in tables 7a and 7b.

In general, results indicate that local green infrastructure plays an important role in mitigating the dis-amenity value of being located in high risk flood zones. The average marginal impact of being closer to very high risk flood zones is mitigated if a dwelling has a higher concentration of urban forests nearby or a higher average level of neighbourhood greenery as determined by the NDVI (model 15 and 16). Urban forests and large tree stands

are important aspects of a city in terms of storm water runoff management and flood mitigation strategies, and general vegetation also plays a similar role. These urban green amenities provide ample pervious surfaces that allow excess water to drain off easily and thus cause less flood dis-amenity to residents.

Table 7a. Flood Risk Interaction Effects

Dep. Variable: ln(Price)	**Model 14**	**Model 15**	**Model 16**	**Model 17**	**Model 18**
Located in V. High Risk Flood Zone	-0.03489***	-0.03507***	-0.03569***	-0.03601***	-0.03208***
	(0.01128)	*(0.01128)*	*(0.01129)*	*(0.01128)*	*(0.01129)*
Distance to V. High Flood Risk Zone × Nº. of Lakes in 100 m	0.01008**				
	(0.00435)				
Distance to V. High Flood Risk Zone × Nº. of Urban Forests in 100 m		-0.02492**			
		(0.01144)			
Distance to V. High Flood Risk Zone × Average NDVI in 100 m			-0.14395*		
			(0.08410)		
Distance to V. High Flood Risk Zone × ln(Average Slope in 100 m)				-0.00598***	
				(0.00228)	
Distance to V. High Flood Risk Zone × ln(Length of Roads in 100 m)					0.00551***
					(0.00152)
Lambda	0.15388***	0.15431***	0.15489***	0.15299***	0.15238***
	(0.02309)	*(0.03935)*	*(0.06793)*	*(0.02869)*	*(0.02569)*
Year F.E.	Yes	Yes	Yes	Yes	Yes
500 m F.E.	Yes	Yes	Yes	Yes	Yes
AIC	-3993.5	-3991.0	-3992.5	-3996.3	-3999.7
$AIC_{SEM} \div AIC_{OLS}$	*1.00490*	*1.00490*	*1.00490*	*1.00480*	*1.00480*
Log Likelihood	2243.7	2242.5	2243.3	2245.2	2246.9
$L.L._{SEM} \div L.L._{OLS}$	*1.00480*	*1.00480*	*1.00480*	*1.00470*	*1.00470*
SSE	1652.3	1652.4	1652.3	1652.2	1652.0
$SSE_{SEM} \div SSE_{OLS}$	*0.99903*	*0.99903*	*0.99897*	*0.99903*	*0.99903*
Residual Std. Error	0.22576	0.22576	0.22576	0.22575	0.22573
$Res.\ Error_{SEM} \div Res.\ Error_{OLS}$	*0.99577*	*0.99572*	*0.99572*	*0.99577*	*0.99572*
Mean Adj. VIF for Hazard Variables	2.39	1.92	3.85	2.66	4.79
Max Adj. VIF for Hazard Variables	2.79	1.99	5.71	3.33	7.58
Spatial Breusch-Pagan	1208.6***	1208.9***	1210.8***	1207.7***	1204.0***
Wald Test	44.42***	15.38***	5.20**	28.44***	35.19***
Likelihood Ratio Test	21.37***	21.58***	21.65***	21.14***	20.91***
Observations	32,420	32,420	32,420	32,420	32,420

*Notes: ***Significance at 1 % level; **Significance at 5 % level; *Significance at 10 % level.*
Heteroskedastic consistent errors

While urban green infrastructure can mitigate some of the negative dis-amenity values associated with being in an area of high flooding risk, compounding negative effects come from being located nearby to lakes and impervious surfaces (model 14 and model 18). The pooling of storm water can be significant in these areas and result in high surface runoff and reduction in lag time.[9] Dwellings with higher concentrations of lakes or a denser road network

[9] Surface runoff is water, from rain, snowmelt, or other sources, that flows over the land surface, and is a major component of the water cycle. Lag time is defined as the time difference between peak runoff and the mass center of rainfall excess.

nearby therefore have higher prices for being located further away from flood risk zones with prices increasing with higher concentrations of these amenities further away from the hazard.

The built environment can therefore have profound influence on how residents value the relative impact of these urban hazards, and thus has important implications for municipalities and developers to create more amenities which may mitigate the negative impact of urban natural hazards. If neighbourhoods in risky areas of a city can be developed in such a way as to provide residents with these mitigating amenities, then the results suggest that this is captured and capitalized by the residential real estate market.

Table 7b. Seismic Risk Interaction Effects

Dep. Variable: ln(Price)	**Model 19**	**Model 20**	**Model 21**
Located in Very High Risk Seismic Zone	-0.00044*	-0.00058**	-0.00074***
	(0.00026)	*(0.00026)*	*(0.00026)*
Distance to Very High Seismic Risk Zone × % Neighbourhood Property Owners	-0.26412***		
	(0.06317)		
Distance to Very High Seismic Risk Zone × % Neighbourhood Educated		-0.18500**	
		(0.09163)	
Distance to Very High Seismic Risk Zone × % Buildings with 1 or 2 Stories			-0.33073***
			(0.11074)
Lambda	0.16013***	0.16114***	0.15865***
	(0.04145)	*(0.01857)*	*(0.01717)*
Year F.E.	Yes	Yes	Yes
500 m F.E.	Yes	Yes	Yes
AIC	-4001.1	-3985.1	-3991.3
$AIC_{SEM} \div AIC_{OLS}$	*1.00540*	*1.00540*	*1.00520*
Log Likelihood	2247.6	2239.6	2242.7
$L.L._{SEM} \div L.L._{OLS}$	*1.00520*	*1.00530*	*1.00510*
SSE	1651.9	1652.7	1652.4
$SSE_{SEM} \div SSE_{OLS}$	*0.99897*	*0.99897*	*0.99879*
Residual Std. Error	0.22573	0.22578	0.22576
$Res.\ Std.\ Error_{SEM} \div Res.\ Std.\ Error_{OLS}$	*0.99572*	*0.99568*	*0.99563*
Mean Adj. VIF for Hazard Variables	4.22	4.11	6.00
Max Adj. VIF for Hazard Variables	6.29	6.07	9.90
Spatial Breusch-Pagan	1215.4***	1218.1***	1218.7***
Wald Test	14.92***	75.34***	85.37***
Likelihood Ratio Test	23.30***	23.58***	22.82***
Observations	32,420	32,420	32,420

In terms of seismic risk zones, neighbourhood characteristics have important mitigating behaviors. In neighbourhoods where there is a higher percentage of owner-occupiers or educated individuals (model 19 and model 20) the dis-amenity value of being located nearer to seismic risk areas is attenuated. These indicators may serve as a proxy to indicate how well homes in an area are maintained with property owners specifically having a larger incentive to provide protection for their properties and for themselves and their relatives against such risk.

Moreover, in general educated people tend also to be better informed about general topics including urban hazards and thus, be potentially more engaged in preparedness activities such as collecting survival items such as food and water, undertaking mitigation actions such as retrofitting buildings, securing household items, making a household emergency plan or simply learning survival skills.[10]

In terms of the built environment, our results seem to suggest that there is a price premium for being located closer to seismic risk zones in which there are higher percentages of low-lying buildings with one or two stories (model 21). This result should be interpreted with caution as it may be related to the residents' perception that in the case of an earthquake, these buildings tend to be the most stable with higher risks coming from larger structures or high-rise buildings. However, damages during an earthquake results from several factors including strength and length of the shaking, type of soil and type of building. Buildings of different heights tend to respond differently in an earthquake. Aside from architectural constraints (i.e., how well built the structure is) the particular resonance of an earthquake can knock down a small building and spare the skyscraper.[11] Small building are more affected, or shaken, by high-frequency waves (short and frequent). On the other hand, large structures or high rise buildings are more affected by long period, or slow shaking.

Between Variation of Hazard Risks

While the average price impact of hazard zones is negative, residents may trade off this risk if other aspects of their location have benefits which outweigh these risks. In table 8,

[10] It should be noted that the infrequent nature of seismic hazard events means that people often also lack personal experience of such a hazard (Becker et al. 2017). They will, however, have indirect experience (e.g. experience of small seismic events that did not impact them directly), vicarious experience (e.g., media reports of national or international events, accounts of prior events from relatives), and challenging life event experience (e.g., of accidents, crime etc.), all of which could play independent and interdependent roles in future preparedness decision making and actions.

[11] The resonance is the oscillation (up-and-down or back-and-forth motion) caused by a seismic wave. During an earthquake, buildings oscillate. If the frequency of this oscillation is close to the natural frequency of the building, resonance may cause severe damage.

dwellings which are located in very high risk areas are interacted with indicators to represent being located in attractive areas of the city as determined by proximity to the riverfront or historic conservation areas. If the benefits of these zones outweigh the costs, then we would expect a mitigating effect on the price impact risk of urban hazards.

Table 8. Locational Interaction Effects

Dep. Variable: ln(Price)	**Model 22**	**Model 23**	**Model 24**	**Model 25**
Located in Very High Risk Flood Zone	-0.04069*** *(0.01173)*	-0.03680*** *(0.01132)*	-0.05462*** *(0.01604)*	
Located in Very High Risk Seismic Zone				-0.01509** *(0.00673)*
Located in Very High Risk Flood Zone × Located in Conservation Area	0.04246* *(0.02537)*			
Located in Very High Risk Flood Zone × Located 100 m from Tagus		0.09123* *(0.05540)*		
Located in Very High Risk Flood Zone × Located 500 m from Tagus			0.03798* *(0.02141)*	
Located in Very High Risk Seismic Zone × Located in Conservation Area				0.03581* *(0.02005)*
Lambda	0.15216*** *(0.05108)*	0.15407*** *(0.04272)*	0.15421*** *(0.05458)*	0.15806*** *(0.01841)*
Year F.E.	Yes	Yes	Yes	Yes
500 m F.E.	Yes	Yes	Yes	Yes
AIC	-3993.0	-3992.9	-3992.5	-3984.9
$AIC_{SEM} \div AIC_{OLS}$	*1.00470*	*1.00490*	*1.00490*	*1.00520*
Log Likelihood	2243.5	2243.5	2243.3	2239.5
$L.L._{SEM} \div L.L._{OLS}$	*1.00470*	*1.00480*	*1.00480*	*1.00510*
SSE	1652.3	1652.3	1652.3	1652.7
$SSE_{SEM} \div SSE_{OLS}$	*0.99903*	*0.99903*	*0.99897*	*0.99897*
Residual Std. Error	0.22576	0.22576	0.22576	0.22578
$Res.\ Std.\ Error_{SEM} \div Res.\ Std.\ Error_{OLS}$	*0.99577*	*0.99572*	*0.99572*	*0.99568*
Mean Adj. VIF for Hazard Variables	1.66	1.69	2.51	1.86
Max Adj. VIF for Hazard Variables	2.06	2.00	2.79	2.16
Spatial Breusch-Pagan	1212.6	1211.1	1210.0	1209.1
Wald Test	8.88***	13.00***	7.98***	73.72***
Likelihood Ratio Test	20.87***	21.43***	21.44***	22.63***
Observations	32,420	32,420	32,420	32,420

*Notes: ***Significance at 1 % level; **Significance at 5 % level; *Significance at 10 % level.*
Heteroskedastic consistent errors

Conservation areas are shown to positively mitigate the dis-amenity value of both flooding and seismic risk. These areas are maintained by the municipality in order to preserve their charm and character, and thus are likely to be more prepared for the eventual floods which occur each year and with priority clean ups occurring after significant events. Even with these significant hazard risks, the net effect of being located in a simultaneous flood hazard and conservation area is 0.2%, or approximately €500 (model 22). This effect is more pronounced for seismic risk zones with a net effect of approximately 2.1% (model 25).

While flooding risk is highest near the riverfront, it is important to disentangle the impact due to the dis-amenity value of the urban hazard risk and the amenity value of the riverfront, which is sought after by residents. If the benefits of being located nearer to the riverfront, for recreational or aesthetic purposes, outweigh the costs of being in a risk zone, then we would expect this trade-off to be capture in dwelling prices.

Very localized effects are found coming from dwellings directly at the riverfront and located within 100 meters of the Tagus. The positive impact of being on the riverfront outweighs the negative cost associated to being in a very high risk hazard zone with a net benefit of approximately 5.4% (model 23). Residents therefore capitalize on direct proximity to the riverfront even if these areas have inherently large risk, a result consistent with the previous literature. These results however appear to be fairly localized with net negative impacts still occurring if a dwelling is only located within 500 meters of the riverfront (model 24). This suggests that the amenity value of the riverfront is strongest for those directly in the line of sight, and residents are willing to trade-off the risk of flooding to be in this zone.

2 – 5.4. Geographic Regression Discontinuity Robustness

From the baseline results, we check the robustness of estimates by considering spatial subsets around each type of hazard zone boundary. The GRD estimates are presented in table A2 – 4 of the appendix and show that the estimated price impacts of being located within a designated flood risk or seismic risk zone is consistent and robust to a variety of spatial subsets.

We consider the effects from very high hazard risks (Models 1, 3 and 5 in table 5), and consider a subset of properties 500 meters outside of the respective hazards geographic boundary as the control group. Nearby properties should have similar local amenities and underlying influences, and by removing properties located at some farther distance of the geographic boundary we remove potential locational influences which may be driving the results. We further directly consider the clustering of hazard zones near the river, by showing

that estimated effects are robust to considering a subset of dwellings 3 kilometers from the riverfront. This, along with explicitly including covariates measuring proximity to the river and its interaction with elevation ensures that the price impacts of flood risk are not driven by significant locational characteristics which may be attributed to proximity to the Tagus.

The choice of distance outside a boundary to consider in a GRD however may be subjective, and so further models draw control properties using a propensity score matching process to match properties located in hazard zones to those located outside of these areas conditional on important locational influences. Flood prone properties are matched to non-flood prone properties based on their distance to the nearest urban green infrastructure in the form of urban forests, on neighbourhood population density, and on the amount of impervious road surface within 100 meters of an observation. Seismic risk properties on the other hand are matched conditional on the average slope within 100 meter of a dwelling. By comparing similar properties in these respects, we are removing potential mechanisms which may be related to and influence the estimated price impact of being in such hazard zones.

2 – 6. Conclusions

This chapter investigates the capitalization of urban hazards on residential property values in Lisbon, Portugal, with specific emphasis on spillover effects and the heterogeneity within and between areas of urban hazards. Results indicate that housing prices are negatively impacted by being located in areas of very high flooding risk or very high seismic risk, however these results may be mitigated or exacerbated conditional on a dwellings local environment.

While location in a flood zone detracts from housing prices, this effect is found to be mitigated by proximity to urban green spaces and greenery and exacerbated by nearby lakes and impervious surfaces. Seismic risk on the other hand is significantly mitigated by characteristics of the neighbourhood in terms of more owner-occupiers and educated individuals, and more low-lying buildings.

Further, although being located in an urban hazard zone has negative impacts results suggest that not only are there negative spillover effects to nearby non-risk areas, but residents may trade off the dis-amenity value from flood zones for the benefit of being located in desirable areas of the city, namely at the riverfront or in historically protected zones.

Although the location of hazard risk zones in the city are not driven by housing prices, there may be some underlying influencing impacting the estimates. In implementing a GRD design, we ensure that the estimated impacts of hazard zones are not driven by significant locational differences. Sub-setting the data around hazard borders removes locational or neighbourhood differences that could be a driving mechanism from which the estimates are obtained. The GRD design shows that results are robust and not driven by this heterogeneity.

These results have important policy implications for municipalities. Not only does it provide a value for how the risk of these events impact residential real estate markets, and subsequently property tax collection, but further provides an indication as to what amenities and neighbourhood characteristics either attenuate or compound the negative effects of natural hazard risks. As the risk of these hazard events capture their persistence and local residents' exposure, better understanding the true value of their impacts is important.

By showing that the impact of flooding zones is conditional on urban green infrastructure, these results provide an indication that the variability in flood risk zone is conditional on local green amenities. Such amenities could thus be implemented in high risk flooding zones to attenuate the negative effects experienced by residents. Similarly, considering the types of buildings in high risk seismic areas should be a priority for developers and the municipalities' point of view, with low-lying structures not only safer in the event of seismic activity, but also valued by local residents.

The estimates suggest that the per dwelling average price impact of being located in a flood risk zone or a seismic risk zone is €8,500 and €2,700 respectively. This effect however,

especially for flood risk areas, is not the same across the distribution with higher priced dwellings having a stronger impact. These properties appear to react to the threat of flood risk more than seismic risk, and higher valued properties may be relatively more damaged with greater price influences in the case of floods. Aggregated across all residents exposed to such risk suggests that the overall impact of these natural hazard risks are quite large. By understand the impacts that these hazards can have on the real estate market, the municipality and planners are better able to prepare and plan for the occurrences of these hazards and better respond to the needs of residents.

2 – References

Anselin, Luc, and Daniel Arribas-Bel. 2013. "Spatial fixed effects and spatial dependence in a single cross-section." *Papers in Regional Science* 92, no. 1: 3-17.

Atreya, Ajita, and Jeffrey Czajkowski. 2016. "Graduated Flood Risks and Property Prices in Gaveston County." *Real Estate Economics:* 1-38.

Bakkensen, Laura A., and Lint Barrage. 2018. "Flood Risk Belief Heterogeneity and Coastal Home Price Dynamics: Going Under Water?" *NBER Working Paper 23854.*

Becker, Julia, Douglas Paton, David M. Johnston, Kevin R. Ronan, and John McClure. 2017. "The role of prior experience in informing and motivating earthquake preparedness." *International Journal of Disaster Risk Reduction* 22: 179-193.

Bin, Okmyung, and Craig E. Landry. 2008. "Flood Hazards, Insurance Rates, and Amenities: Evidence from the Coastal Housing Market." *The Journal of Risk and Insurance* 75, no. 1: 63-82.

Bin, Okmyung, and Craig E. Landry. 2013. "Changes in implicit flood risk premiums: Empirical evidence from the housing market." *Journal of Environmental Economics and Management* 65: 361-376.

Burby, Raymond J. 1998. *Cooperating with Nature: Confronting Natural Hazards with Land-Use Planning for Sustainable Communities.* Joseph Henry Press.

Franco, Sofia F., and Jacob L. Macdonald. 2018. "The effects of cultural heritage on residential property values: Evidence from Lisbon, Portugal." *Regional Science and Urban Economics* 70: 35-56.

Gencer, EA. 2013. *Natural disasters, urban vulnerability, and risk management: a theoretical overview*. In: *The interplay between urban development, vulnerability, and risk management a case study of the Istanbul metropolitan area*, Mediterranean Studies 7. Berlin: 7-43.

Hidano, Noboru, Tadao Hoshino, and Ayako Sugiura. 2015 “The effect of seismic hazard risk information on property prices: Evidence from a spatial regression discontinuity design.” *Regional Science and Urban Economics* 55: 113-122.

Instituto Superior Técnico. 2005. “Estudo Sectorial sobre Risco Sísmico.” *Instituto de Engenharia de Estruturas, Território e Construção.*

Keel, Luke J., and Rocio Titiunik. 2015. “Geographic Boundaries as Regression Discontinuities.” *Political Analysis* 23: 127-155.

Koster, Hans R. A., and Jos van Ommeren. 2015. “A shaky business: Natural gas extraction, earthquakes and house prices.” *European Economic Review* 80: 120-139.

Lall, Somik V., and Uwe Deichmann. “Density and Disasters: Economics of Urban Hazard Risk.” *World Bank Observer* 27 (July 2010): 74-105.

Lindell, Michael K. 1997. “Adoption and Implementation of Hazard Adjustments.” *International Journal of Mass Emergencies and Disasters* 15, no. 3: 327-414.

MacDonald, John M., Jonathan Klick, and Ben Grunwald. 2016. “The effect of private police on crime: evidence from a geographic regression discontinuity design.” *Journal of the Royal Statistical Society Series A: Statistics in Society* 179 (3): 831-846.

Metz, Neil E., Travis Roach, and Jordan A. Williams. 2017. “The costs of induced seismicity: A hedonic analysis.” *Economic Letters* 160: 86-90.

Naoi, Michio, Kazuto Sumita, and Miki Seko. 2010. “Estimating Consumer Valuation of Earthquake Risk: Evidence from Japanese Housing Markets.” *International Real Estate Review* 13, no. 2: 117-133.

Naoi, Michio, Miki Seko, and Kazuto Sumita. 2009. “Earthquake risk and housing prices in Japan: Evidence from before and after massive earthquakes.” *Regional Science and Urban Economics* 39: 658-669.

Pereira, Alvaro S. 2009. “The Opportunity of a Disaster: The Economic Impact of the 1755 Lisbon Earthquake.” Centre for Historical Economics and Related Research at York Discussion Paper.

Rajapaksa, Darshana, Clevo Wilson, Shunsuke Managi, Vincent Hoang, and Boon Lee. 2016. “Flood Risk Information, Actual Floods and Property Values: A Quasi-Experimental Analysis.” *Economic Record* 92, no. S1: 52-67.

Rajapaksa, Darshana, Clevo Wilson, Viet-Ngu Hoang, Boon Lee, and Shunsuke Managi. 2017a. “Who responds more to environmental amenities and dis-amenities?” *Land Use Policy* 62: 151-158.

Rajapaksa, Darshana, Min Zhu, Boon Lee, Viet-Ngu Hoang Clevo Wilson, and Shunsuke Managi. 2017b. “The impact of flood dynamics on property values” *Land Use Policy* 69: 317-325.

Rischard, Maxime, Zach Branson, Luke Miratrix, and Luke Bornn. 2018. “A Bayesian Nonparametric Approach to Geographic Discontinuity Designs: Do School Districts Affect NYC House Prices?” Working Paper.

Rosen, Sherwin. 1974. “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition.” *Journal of Political Economy* 82, no. 1: 34-55.

Tang, Y, Y. Yin, K. Hill, V. Katiyar, A. Nasseri, and T. Lai. 2012. “Seismic Risk Assessment of Lisbon Metropolitan Area under a Recurrence of the 1755 Earthquake with Tsunami Inundation.” 15th World Conference on Earthquake Engineering.

United Nations. *Living with Risk: A Global Review of Disaster Reduction Initiatives*. 2 vols. 2004, Geneva.

Zhang, Lei. 2016. “Flood hazards impact on neighbourhood house prices: A spatial quantile regression analysis.” *Regional Science and Urban Economics* 60: 12-19.

2 – Appendix

Table A2 – 1. Descriptive Statistics

	N	*Mean*	*St. Dev.*	*Min*	*Max*
Price	32,420	243,500	148,503	25,000	2,500,000
Structural Characteristics					
Sq. Meters	32,420	98.900	45.970	15	420
New Construction	32,420	0.200	0.400	0	1
View of Tagus River	32,420	0.060	0.230	0	1
Swimming Pool	32,420	0.010	0.100	0	1
Parking Spaces	32,420	0.120	0.330	0	1
Fireplace	32,420	0.040	0.190	0	1
Double Windows	32,420	0.210	0.410	0	1
Air Conditioning	32,420	0.130	0.340	0	1
Elevator	32,420	0.230	0.420	0	1
Locational Characteristics					
Located within 3 km of the Riverfront	32,420	0.480	0.500	0	1
Located 100 m from Tagus Riverfront	32,420	0.007	0.084	0	1
Located 500 m from Tagus Riverfront	32,420	0.116	0.320	0	1
Located 500 m outside of an Area of Very High Seismic Risk	32,420	0.633	0.482	0	1
Located 500 m outside of an Area of Very High Flooding Risk	32,420	0.408	0.491	0	1
Located in Conservation Area	32,420	0.182	0.386	0	1
Neighbourhood Characteristics					
% Neighbourhood Property Owners	32,420	0.506	0.163	0.117	0.849
% Neighbourhood Educated	32,420	0.300	0.107	0.012	0.571
% Buildings with 1 or 2 Stories	32,420	0.283	0.175	0.006	0.884
Length of Roads in 100 m	32,420	648.000	325.24	0	2071
Average Slope within 100 m	32,420	8.176	4.240	1.500	26.224
Neighbourhood Crimes per Person	32,420	0.030	0.031	0.005	0.421
Neighbourhood Thefts per Person	32,420	0.023	0.026	0.005	0.376
Open Spaces and Ecological Urban Hazards					
Elevation	32,420	66.900	29.710	0	145
Distance to Tagus Riverfront	32,420	2.590	1.975	0.009	7.43
No. of Urban Forests in 500 m	32,420	0.367	0.609	0	3
Average Distance (km) to Parks	32,420	4.713	0.883	3.504	7.429
Nº. of Lakes in 100 m	32,420	0.133	0.524	0	10
Average NDVI in 100 m	32,420	0.061	0.056	-0.064	0.284
Located in an Area of High or Very High Seismic Risk	32,420	0.454	0.498	0	1
Located in an Area of High or Very High Flooding Risk	32,420	0.140	0.347	0	1
Located in an Area of Very High Seismic Risk	32,420	0.196	0.397	0	1
Located in an Area of Very High Flooding Risk	32,420	0.062	0.242	0	1

Table A2 – 2. Spatial Weight Properties

	Description	Nº. Locations	Nº. Non-zero Links	% Non-zero Links	Avg. Nº. of Links	Locations Without Links
SW1	Inverse distance of all properties in 500 meters	32,420	32,175,894	3.06	992.47	5
SW2	Inverse sq. distance of all properties in 500 meters	32,420	32,175,894	3.06	992.47	5
SW3	All properties in 500 meters	32,420	32,175,894	3.06	992.47	5
SW4	All properties in 100 meters	32,420	22,211,752	2.11	685.13	115
SW5	100 nearest neighbors	32,420	3,242,000	0.31	100.00	0
SW6	10 nearest neighbors	32,420	324,200	0.03	10.00	0

Figure A2 – 1. Geographic Regression Discontinuity: Flood Risk Zones

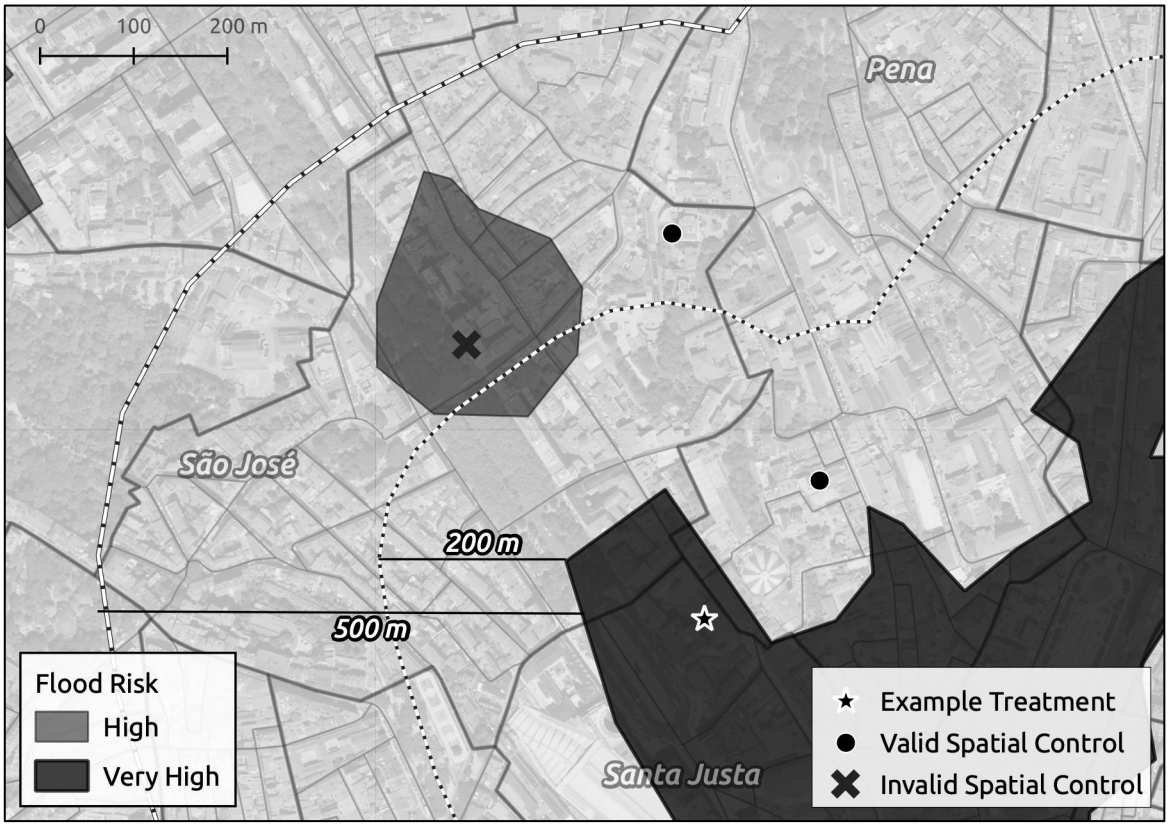


Figure A2 – 2. Price Discontinuity at Urban Hazard Boundaries

Panel A: Flood Risk Zones

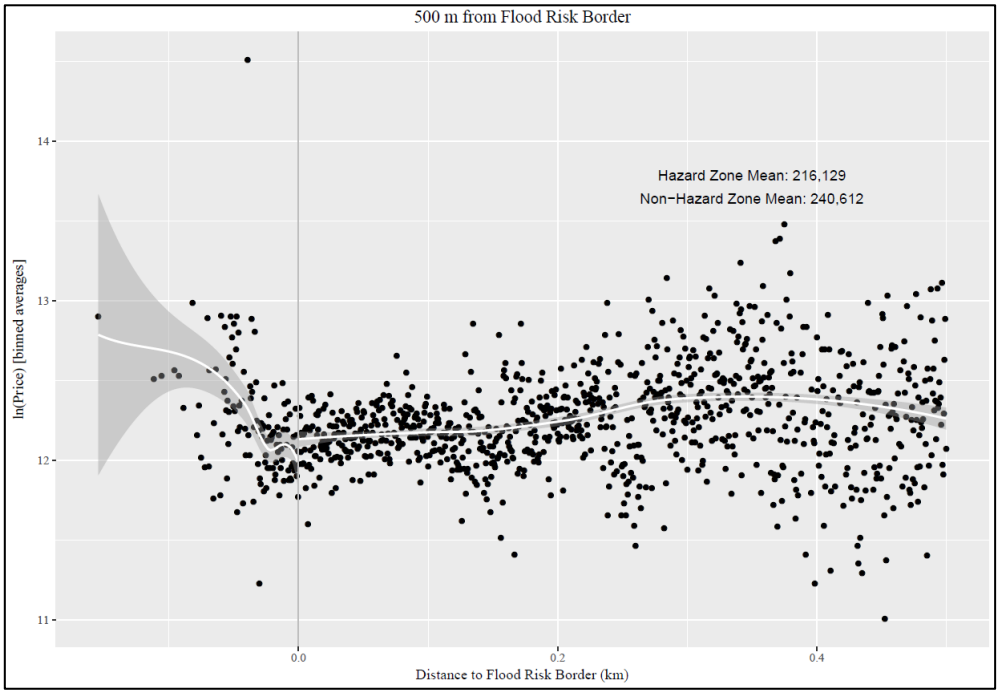


Figure A2 – 2. Price Discontinuity at Urban Hazard Boundaries

Panel B: Seismic Risk Zones

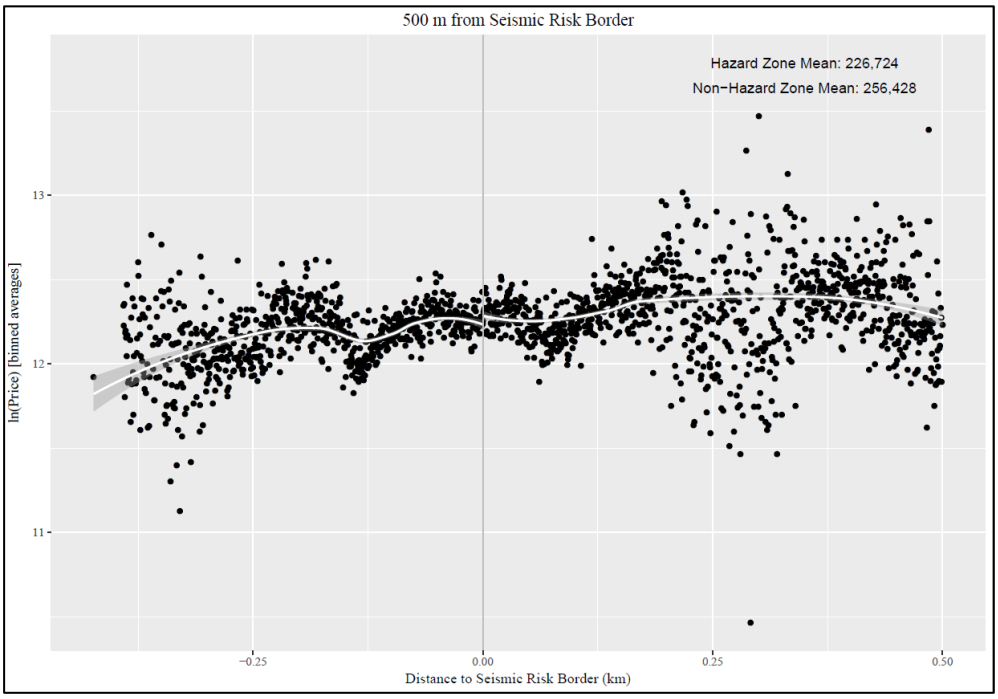


Figure A2 – 3. Various Placebo Threshold Boundaries

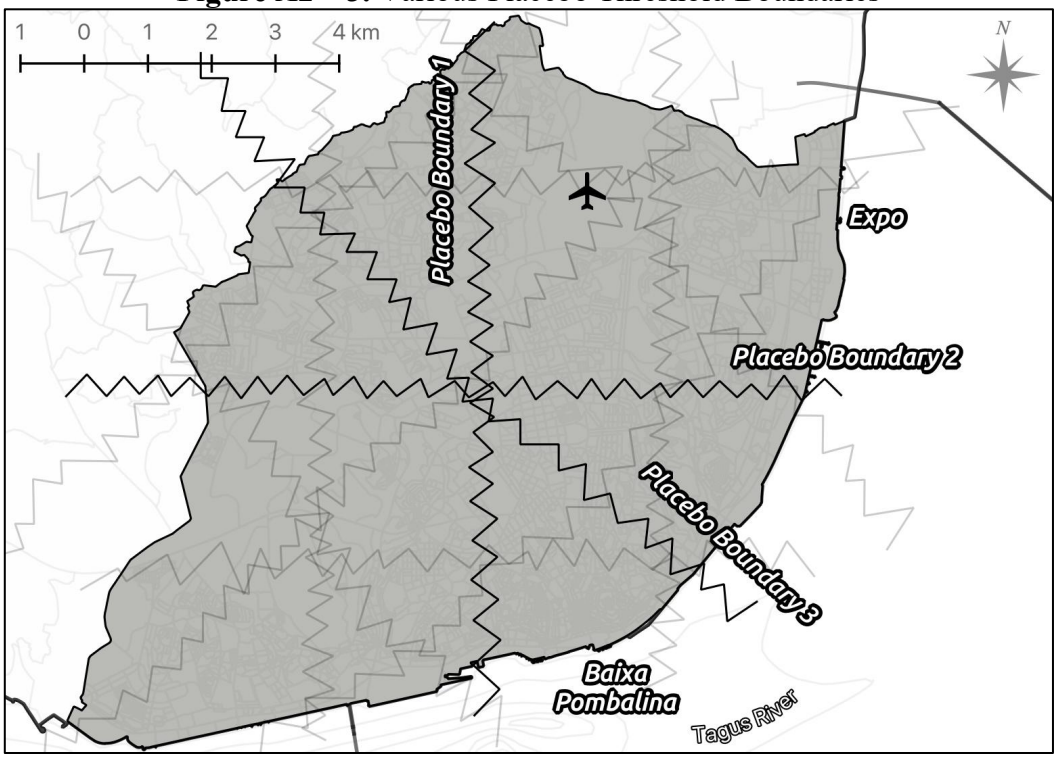


Table A2 – 3. Various Estimated Placebo Threshold Effect

Data Subset:	**GRD Threshold of 500 meters**				**GRD Threshold of 1 kilometer**				**GRD Threshold of 3 kilometers**			
	Flood Risk	***No Flood Risk***	***Seismic Risk***	***No Seismic Risk***	***Flood Risk***	***No Flood Risk***	***Seismic Risk***	***No Seismic Risk***	***Flood Risk***	***No Flood Risk***	***Seismic Risk***	***No Seismic Risk***
Placebo Threshold Boundary # 1												
Estimated	-0.02105	0.01884	-0.01486	0.0082	-0.02381	-0.00961	-0.02182	-0.01546	-0.02296	-0.00559	-0.0215	0.00729
Threshold Effect	*(0.02451)*	*(0.07935)*	*(0.02562)*	*(0.05461)*	*(0.02422)*	*(0.08805)*	*(0.02512)*	*(0.06517)*	*(0.02344)*	*(0.09204)*	*(0.02497)*	*(0.06845)*
Observations	4,709	2,311	3,540	3,480	7,966	2,629	6,399	4,196	20,036	4,320	11,007	13,349
Adjusted R^2	0.79248	0.77563	0.81883	0.73349	0.80147	0.77491	0.81758	0.7459	0.80287	0.79727	0.81035	0.77742
Placebo Threshold Boundary # 2												
Estimated	0.04046	0.05159	-0.00298	-0.0081	0.0287	0.03241	-0.02077	-0.00629	0.01658	0.05035	-0.02623	0.02293
Threshold Effect	*(0.03167)*	*(0.06221)*	*(0.03941)*	*(0.05917)*	*(0.03123)*	*(0.05866)*	*(0.03923)*	*(0.05792)*	*(0.03116)*	*(0.06604)*	*(0.04034)*	*(0.05793)*
Observations	4,137	510	2,858	1,789	7,230	1,075	5,867	2,438	20,596	3,105	11,784	11,917
Adjusted R^2	0.80466	0.82646	0.80576	0.8225	0.79794	0.82671	0.7985	0.81603	0.79542	0.78337	0.79336	0.79153
Placebo Threshold Boundary # 3												
Estimated	0.00038	0.10721	0.0074	-0.01789	-0.00643	0.09999	0.02522	-0.02728	-0.00656	0.02899	0.02425	-0.0219
Threshold Effect	*(0.01651)*	*(0.09986)*	*(0.02572)*	*(0.02095)*	*(0.01568)*	*(0.07973)*	*(0.02521)*	*(0.01881)*	*(0.01546)*	*(0.07705)*	*(0.02458)*	*(0.01917)*
Observations	6,572	800	2,637	4,735	11,409	1,046	5,014	7,441	21,024	3,651	12,435	12,240
Adjusted R^2	0.78786	0.81821	0.80507	0.74863	0.80271	0.84996	0.80385	0.77328	0.79299	0.79287	0.79587	0.77958

Table A2 – 4. Spatial Diagnostics

	Global Moran's I (Dep.)	Z-Value (Dep.)	Global Moran's I (Res.)	Z-Value (Res.)	LM SEM	LM SAR	Rob. LM SEM	Rob. LM SAR
Model 1: Very High Flood Risk								
SW1	0.1292***	49.50	0.0048**	2.18	3.38*	0.83	2.57	0.02
SW2	0.0971***	244.10	0.0015***	6.04	13.41***	0.14	13.27***	0.00
SW3	0.1192***	164.90	0.0030***	5.34	15.99	1.20	14.99***	0.20
SW4	0.1292***	145.10	0.0042***	6.02	22.72***	0.21	22.56***	0.05
SW5	0.1213***	162.10	0.0013***	2.85	2.85*	2.42	1.67	1.23
SW6	0.1300***	54.80	0.0033**	1.78	1.99	0.33	3.988**	2.33
Model 3: Very High Seismic Risk								
SW1	0.1292***	49.50	0.0050**	2.25	3.645*	1.00	2.684	0.04
SW2	0.0971***	244.10	0.0015***	6.25	14.78***	0.12	14.66***	0.00
SW3	0.1192***	164.90	0.0030***	5.51	17.32	1.27	16.26	0.21
SW4	0.1292***	145.10	0.0044***	6.20	24.43***	0.23	24.25***	0.06
SW5	0.1213***	162.10	0.0014***	3.02	3.451*	2.64	2.092	1.28
SW6	0.1300***	54.80	0.0035**	1.84	2.169	0.21	3.997**	2.04
Model 5: Very High Joint Hazards								
SW1	0.1292***	49.50	0.004837**	2.20	3.435*	0.91	2.56	0.03
SW2	0.0971***	244.10	0.001506***	6.16	14.16***	0.13	14.03***	0.00
SW3	0.1192***	164.90	0.0029***	5.36	16.17	1.22	15.16	0.21
SW4	0.1292***	145.10	0.004242***	6.02	22.78***	0.24	22.6***	0.07
SW5	0.1213***	162.10	0.001307***	2.90	3.039*	2.54	1.79	1.29
SW6	0.1300***	54.80	0.003315**	1.77	1.948	0.27	3.796*	2.12

Table A2 – 5. Full Estimates and GRD Robustness

	Model 1			**Model 3**			**Model 5**		
	Full Sample	*500m from VH Flood Zone*	*3km from Tagus Riverfront*	*Full Sample*	*500m from VH Seismic Zone*	*3km from Tagus Riverfront*	*Full Sample*	*500m from VH Flood Zone*	*500m from VH Seismic Zone*
Elevation	0.00130***	0.00168***	0.00111***	0.00149***	0.00042	0.00126***	0.00143***	0.00181***	0.00037
	(0.00028)	*(0.00040)*	*(0.00037)*	*(0.00027)*	*(0.00061)*	*(0.00036)*	*(0.00027)*	*(0.00040)*	*(0.00061)*
ln(Distance to Tagus river)	-0.03037*	-0.02449	-0.03913*	-0.02499	-0.04817*	-0.04155*	-0.03000*	-0.02407	-0.05193**
	(0.01769)	*(0.02191)*	*(0.02190)*	*(0.01767)*	*(0.02542)*	*(0.02251)*	*(0.01771)*	*(0.02191)*	*(0.02570)*
Elevation × ln(Distance to Tagus river)	-0.00054**	-0.00046	-0.00061	-0.00064**	0.00019	-0.00073	-0.00056**	-0.00043	0.00022
	(0.00026)	*(0.00038)*	*(0.00049)*	*(0.00025)*	*(0.00050)*	*(0.00049)*	*(0.00026)*	*(0.00038)*	*(0.00050)*
	Flood Risk			***Seismic Risk***			***Joint Hazards***		
	Very High			***Very High***			***Very High***		
Urban Geohazard Risk	-0.03513***	-0.02567**	-0.02900**	-0.01114*	-0.01323*	-0.01830**	-0.03786***	-0.02825*	-0.02897*
	(0.01128)	*(0.01270)*	*(0.01290)*	*(0.00647)*	*(0.00758)*	*(0.00912)*	*(0.01419)*	*(0.01534)*	*(0.01580)*
Lambda	0.15503***	0.30865***	0.31957***	0.16021***	0.23800***	0.32509***	0.15532***	0.31096***	0.23108***
	(0.03744)	*(0.04320)*	*(0.03810)*	*(0.01360)*	*(0.02723)*	*(0.03068)*	*(0.03940)*	*(0.04364)*	*(0.04463)*
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
500 m F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Structural Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Green Amenities	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	32,420	14,717	21,955	32,420	18,980	21,955	32,420	14,717	18,980

CHAPTER 3:
Metro Stations, Low Emission Zones and the Spatial-Temporal Dynamics of Air Pollution

This chapter uses geostatistical interpolation and clustering to process high-frequency and high-resolution open source pollution data. A longitudinal neighbourhood-scale database is generated to evaluate monthly pollution levels and the mitigating impacts of urban public transportation infrastructure. Short and long-run localized reductions around the city are estimated and attributed to the expansion of the underground metro system of Lisbon, Portugal, and the introduction of a targeted low emission zone in the city centre.

3 – 1. Introduction

Broad and efficient transportation networks and infrastructure are core characteristics of an attractive and liveable city. Public transit initiatives, such as the opening of new underground metro stations or the introduction of zonal traffic restrictions, not only influence the daily movement of residents, workers and visitors, but indirectly have important spill-over effects by impacting spatial and temporal patterns of urban air pollutants. In 2010, transportation accounted for almost a quarter of all emissions generated across the globe, of which 40% was from urban transportation specifically (Sims et al. 2014). Local urban transport policies and best practices can therefore yield important contributions to larger scale pollution mitigation and abatement.

Estimating how the introduction of various transit initiatives influence an area's pollution dynamics enables planners and local authorities to evaluate non-monetary environmental benefits and further enact best practices. Challenges exists however in studying such spatially and temporally granular urban dynamics using available open source data.

This chapter uses geostatistical methods to process high-frequency and high-resolution open source pollution measures to value various transit initiatives in terms of their contribution to the reduction of airborne pollutants. A large focus is on how fixed-point measures can be interpolated and aggregated across space and time. Under hyperparameter optimization, enhanced by the inclusion of temporal lags of predicted pollution, Kriging and Inverse Distance

Weight (IDW) families of interpolation are used to generate neighbourhood level monthly pollution concentrations. A variety of parameter, variable and specification choices for each model are compared to ensure the strongest prediction, and generalized sets of diagnostics and algorithms are used to select the best, most consistent, model for each.

The spatial and temporal nature of pollution monitoring is used to generate a neighbourhood level monthly longitudinal database to explore how neighbourhood pollution concentration has been affected by transportation policy over the past two decades. This high-dimensional database is used to estimate the pollution abating impacts of urban public transportation infrastructure. In particular, the short and long-run localized pollution reductions surrounding the expansion of the underground metro system of Lisbon, Portugal, and the introduction, of a series of traffic based configurations and a targeted low emission zone (LEZ) in the city centre aimed at limiting congested flows of high polluting vehicles.

Long-run effects are estimated under a spatial-temporal difference-in-difference strategy to obtain the average treatment effect of a transit intervention on neighbourhoods in key areas of the city. With limited observations in shorter-run time spans surrounding an intervention, bootstrapping is conducted to provide valid difference-in-difference estimates for immediate effects. This allows for the estimation of month-to-month decaying impacts following the introduction of transit initiatives and further highlights this behaviour over space.

Results indicate that the expansion of new metro stations have decreased pollution primarily in the city centre and around newly opened stations. Short run localized reductions of PM_{10} immediately following the opening range up to 2% with longer run reductions of 0.18%. Metro openings had a particularly large impact on decreasing nitrogen-based combustion emission along the riverfront with short-run reductions of 20%, dissipating over time and space.

Interventions surrounding the LEZ by comparison decreased pollution in and around its boundary and into the city centre with immediate reductions in subsequent months up to 4% of PM_{10} and long-run impacts of 0.43%. Some evidence suggests however that the introduction of the LEZ may have shifted pollution patterns with increases just outside its boundary and along the Tagus river. The LEZ had significant reductions on SO_2 where metro openings did not, capturing the policy's aim of reducing the heaviest polluting vehicles, often running on diesel fuel.

Granular data, from a spatial and temporal dimension, will increasingly continue to shape the valuation of detailed urban-environmental processes. This chapter develops a set of generalizable criteria and diagnostic selections from which sophisticated geostatistical and temporal methods are used to generate measures of the urban environment. Increasingly detailed data yields increasingly detailed applications, and this work highlights the benefit, in the resolution of spatial and temporal impact evaluations, that can be had by leveraging the use of geostatistical methods in the urban context.

3 – 2. Literature Review

Underground metro systems, and regional public transit in general, have a wide range of impacts and spill-over effects as accessibility increases. Not only do network expansions have positive impacts on ridership (Baum-Snow and Kahn 2000, Baum-Snow and Kahn 2005, Goetzke 2008, Zhang et al. 2017) and congestion (Anderson 2014, Adler and van Ommeren 2016), but additional influences are often felt in other markets. This could include impacts to local property prices (Bowes and Ihlanfeldt 2001, Martinez and Viegas 2009, Mohammad et al. 2013, Mulley and Tsai 2016, Li 2018, Mulley et al. 2018), land use and spatial distributions (Cervero and Kang 2011, Roukouni et al. 2012, Gonzalez-Navarro and Turner 2018), or even local labour markets (Sanchez 1999, Kawabata and Shen 2007).

From an environmental perspective, one of the most important spill-over impacts of public transit accessibility is change to local pollution. The evaluation of these impacts at a detailed spatial and temporal resolution however is challenged by the limitations of accurate and timely data. Many works have examined this relationship at varying scales ranging from in-situ sampling (Meinardi et al. 2008, Pereira et al. 2013) or more model-based framework (Anselin and Gallo 2006, Chen and Whalley 2012, Bertazzon et al. 2015).

A common theme across the literature is the identification of shocks to the transit system, often in the form of station openings, in order to estimate the resulting change in pollution as a proxy for the contribution of public transit. Gendron-Carrier et al. (2018) exploit cross-city variation in subway systems to estimate the impacts on particulate matter one year and a half before and after the opening of respective stations. Across cities, results indicate average reductions of 4% extending 10-kilometres from the city centre. In Granada, the expansion and restructuring of the public transportation network reduced PM_{10} concentrations by up to 33% (Titos et al. 2015). In terms of opening of new stations in an urban setting, Zheng et al. (2019) estimate a difference-in-difference reduction effect on carbon monoxide in the areas surrounding the new subway lines.

One of the most common municipal transport policies used to mitigate pollution is the introduction of zonal based traffic restrictions in key congested areas of a city. Currently across Europe there are total of 264 LEZ's varying in scope and extent (Santos et al. 2019). These zones are geographically delineated areas with targeted enforcement focused on restricting heavy polluting vehicles. Different versions of this type of policy have been enacted in different contexts and could vary based the scope (e.g. restriction based on time of day, year or type of car, or licence plate number) or based on the manner of enforcement (e.g. ticket citations, camera detection, seriousness of fine) (Wolff and Perry 2010, Holman et al. 2015, Zhang et al. 2017).

In general, the introduction of LEZ's in different contexts has led to reductions in local pollution levels. However, the outcome and mechanisms through which pollution is potentially changed is not straightforward. While we would expect direct pollution abatement due to the restriction of vehicles in these zones, the introduction of regulations could have unintended consequences altering commuter or broader transport networks and resulting in behavioural change which can be difficult to capture with available data.

Many studies focus on comparing monitored values inside and outside of LEZ boundaries to estimate differences in pollution (Nunes da Silva et al. 2014). Following the five years after the implementation of the LEZ in London, Ellison et al. (2013) estimate average reductions from 2.46% to 3.07% using point estimates from four monitoring stations comparing those inside and out of the boundary. Results suggest that effects may be temporary with concentrations reverting towards original levels after some period.

Complementing many studies of LEZ is in-situ measurement of vehicle fleet data to better link underlying pollution reductions to specific mechanism (Ellison et al. 2013, Ferreira et al. 2012). In the Lisbon context, Ferreira et al. (2012) estimate the impact of the LEZ on PM_{10} and NO_2 between 2011 and 2013 by comparing observed effects from the *Avenida da Liberdade* (inside), *Entrecampos* (boundary) and *Olivais* (outside) stations. The reduction in pollution is linked to observed vehicle distributions and ages at different points in time in or around the LEZ. The primary mechanism through which pollution is influenced is more from changes to the traffic composition, removing old fuel-inefficient cars, rather than reductions in traffic volume.

Further behaviour style changes have been noted after the introduction of a LEZ which can be linked to urban pollution. Across Germany, Wolff (2014) estimate an average decline in pollution of around 9% in urban areas, primarily attributed to shifting to greener and less

polluting transport modes. Xu et al. (2015) meanwhile highlight the substitution behaviour as people may switch to public transit following the introduction of private driving restrictions.

At a more macro-scale, Boogaard et al. (2012) estimate the average impact of LEZ's introduced in five Dutch cities targeting heavy-duty trucks by comparing urban and suburban monitoring station values pre and post introduction. While estimated effects show a general overall decline in pollution, considering specifically the urban and suburban stations showed no significant differences. Viard and Fu (2015) use a multivariate regression to model aggregate city-level pollution in Beijing controlling for temporal breaks, weekends, holidays, weather patterns and different transport policies introduced. The authors estimate a decrease in average pollution around 21% from the introduction of a one-day-a-week driving restriction.

Santos et al. (2019) look at relative reductions in pollution levels in Lisbon following the introduction of the LEZ via a multi-dimensional factor design considering a temporal and a spatial dynamic to compare areas before and after the introduction. Estimates indicate that pollution levels from *Avenida da Liberdade* and *Entrecampos* experienced significantly larger declines in ambient pollution between 2009 to 2012. This was estimated via a treatment interaction on zone and time effect, with estimated impacts ranging between a reduction of 22% to 25% for PM_{10} concentration, yet no discernible impacts for NO_2 or NO_X levels.

While many studies provide estimates of pollution impacts from mass public transit and traffic restricting zones, the majority are based on city-level averages or simple mean differences between fixed-point stations. These methods are thus unable to estimate average treatment effects as they may vary across locations in an urban area, and further any differences at the neighbourhood level over time. This lack of heterogeneity can be addressed by making use of more geostatistical interpolation and spatial-temporal modelling.

Anselin and Gallo (2006) incorporate spatial heterogeneity and autocorrelation into a hedonic model of housing prices as influenced by local levels of pollution. The authors

highlight the importance of robust interpolation and diagnostics of pollution data prior to any modelling and, in the context of Southern California, conclude that Kriging interpolation techniques provides the best fit for eventual spatial econometric modelling. Significant bias can be introduced into econometric specifications by not using the most appropriate spatial interpolation to generate data (Anselin and Lozano-Garcia 2008).

In an urban setting, limited data availability and detail make comparable transit intervention analyses difficult. It is in this context that this chapter estimates the within-city spatial and temporal decaying patterns of urban air pollution. Often, the spatial detail comes at the expense of the temporal detail but using high-dimensional data and differencing estimation this work contributes to better understanding the high-resolution and high-frequency dynamics and trajectories of pollution levels following transit interventions.

The analysis merges robust geostatistical frameworks and diagnostics on interpolated air pollution with spatially detailed urban transportation policy. This allows for a focus on more than just city-wide impacts of various policies and evaluates any potential neighbourhood level disparities and environmental inequalities which may occur following changes to public transit or traffic limiting features.

Further, the discussion is built entirely on open sourced georeferenced data. With current computational abilities it is feasible to do spatially and temporally detailed, robust urban policy analyses. As municipalities are at the forefront of climate change and pollution mitigation, detailed evaluations of transportation interventions can be used to implement best practices to maximize positive spill-over benefits.

3 – 3. Pollution Monitoring and Transport Infrastructure

As the capital city and economic hub of the country, the city of Lisbon is densely populated and busy with people, traffic and commerce. In 2017 there were 384,535 firms in the capital, representing around 30% of all those in the country. These businesses attract many

workers and visitors with daily increases to the city population of around 70% coming from those commuting into the city centre. The economy is service oriented and heavily driven by accommodation and transport, retail and trade, technology and communication (*Câmara Muncipal de Lisboa* 2018a).

Local economic activity and further, environment patterns, are driven by the city's important location along the Tagus river, with many ports facilitating the trade and transports of goods and people. While the city had a history of manufacturing, agricultural and industrial practices, particularly along the riverfront, in modern times the bulk of these firms have all but moved out of the urban area and have been replaced by increasingly service and technology-oriented industries. The clustered density around economic hubs of the city, in terms of commerce, population, buildings and traffic, mean that key areas of the city can at times become highly congested leading to high levels of airborne pollutants.

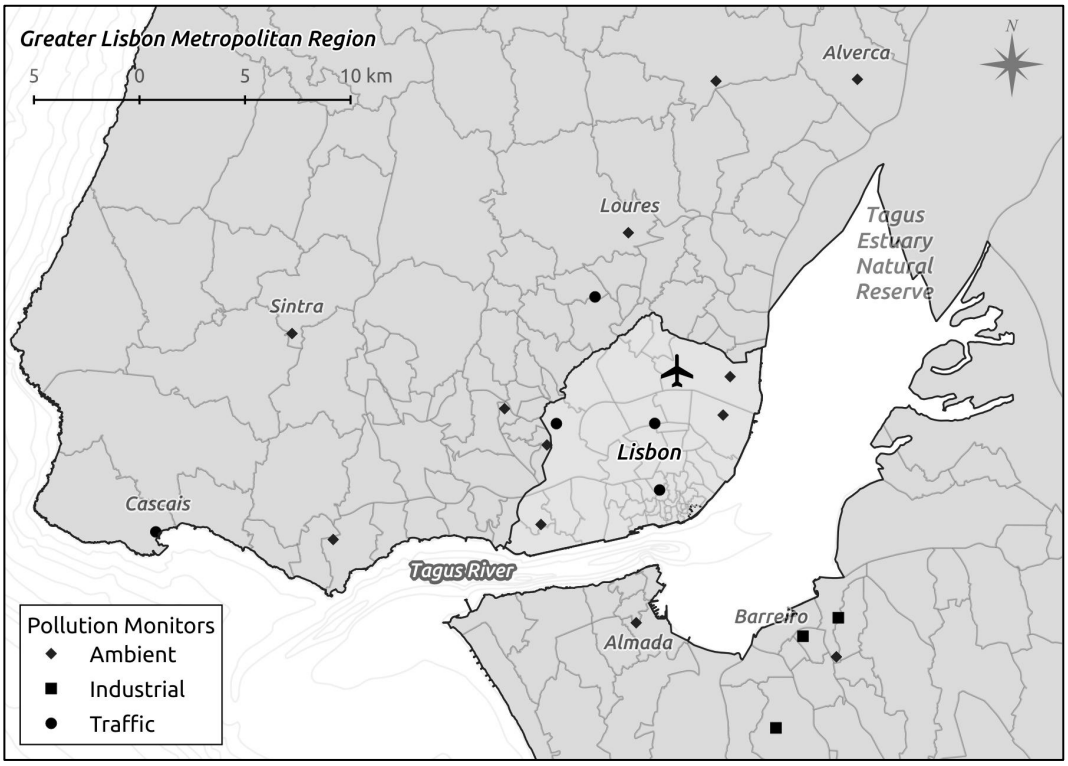
3 – 3.1. Local Pollutants and Trends

The intermittent monitoring of various pollutants across the region began in 1995, tracking high-frequency concentrations of common air and ground level pollution. The collection of pollution data in Portugal is managed by *QualAr*, maintained within the *Agência Portuguesa do Ambiente*. There are 68 monitoring stations across the country of which 20 are located in the immediate vicinity of the greater Lisbon region, as seen in figure 7. This grouping of stations represents a density of approximately one station for every 40 km^2 in the greater metro region and one for every 15 km^2 considering only stations located within the municipal boundaries.

There is variation in the timing of when stations began tracking different pollutants, but full coverage daily measures are in general available for six pollutants over 15 years. These include particulate matter (PM_{10}) since 2002, nitrogen emissions (NO, NO_2, NO_X) since 2000,

1999 and 2004 respectively, carbon monoxide (CO) since 2000 and sulfur dioxide (SO_2) since 2001.[1]

Figure 7. Air Pollution Monitoring Stations in Greater Lisbon, Portugal



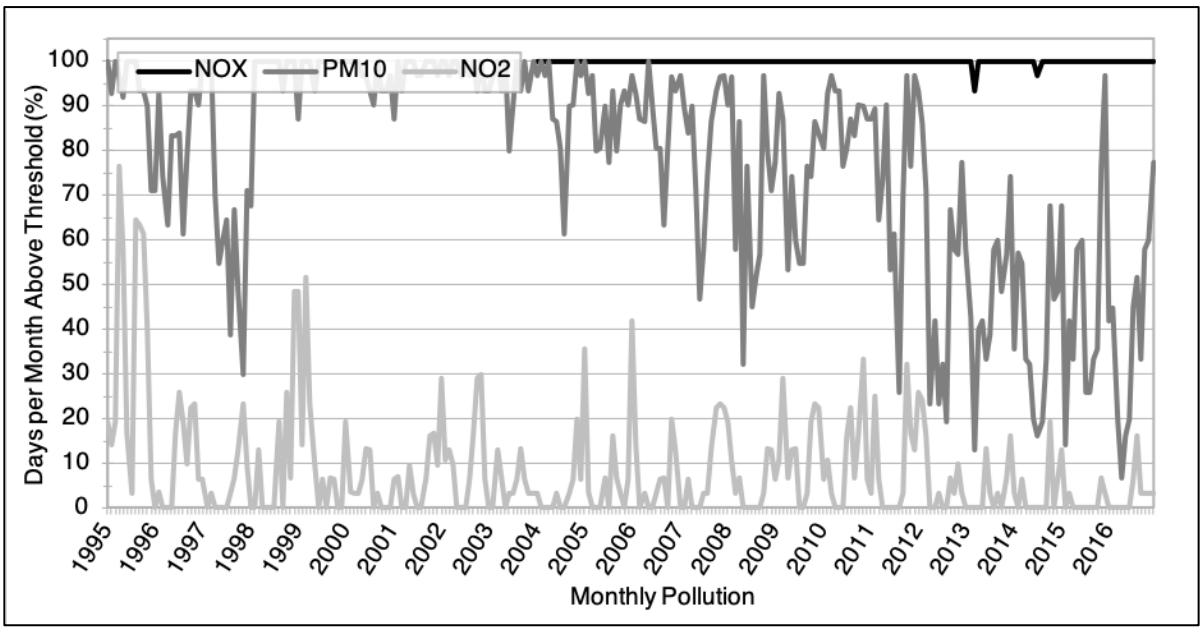
The European Commission has developed a set of air quality standards and regulations which are based on health related research on pollution impacts and are legally binding for member countries (Directive 96/62/EC and subsequent daughter directives). In the case of failing to meet targets, local authorities are then responsible for developing and implementing air quality management plans. Since monitoring began, readings indicated that the Lisbon area consistently exceeded threshold limits for particulate matter and combustion emissions, with high concentrations of PM_{10} across the city and NO_2 particularly to the north. This has led to poor rankings in pollution planning and outcomes relative to other large European cities,

[1] Measures of different pollutants may be available for earlier years (starting in 1995) at some stations. To ensure a sufficient base upon which to interpolate, the empirical analysis interpolates only when there is a minimum of six active monitoring stations.

however, has also led to a significant investment and focus into local municipal environmental concerns.[2]

Figure 8 shows the proportion of days per month that various pollutants have exceeded their respective regulated maximum threshold limit. We see both particulate matter (PM_{10}) and combustion-based pollution (especially NO_X) consistently fail to meet regulated limits, however significant improvements in particulate matter are seen over time. Decreases in PM_{10} correspond to the development of the 2006 Air Quality Action Plan for Lisbon and Tagus Valley which had a goal of ensuring the compliance of legal limits of air pollution set out by regulatory authorities.

Figure 8. Proportion of Days per Month Exceeding Pollution Threshold Limits



Pollution from PM_{10} is a concern for the region given its serious health implications, especially with regards to respiratory health. While there are no safe levels, a daily maximum threshold of 50 μg/m3 is deemed to represent the limit of harmful concentration not to be exceeded 35 days out of the year. High concentrations of PM_{10} is a common problem,

[2] European City Ranking 2015: Best practices for clean air in urban transport
http://www.sootfreecities.eu/sootfreecities.eu/public/

particularly in the main transport corridor leading to the primary business and historic district, *Avenida da Liberdade*.

Particulate pollution however can be highly influenced by regional and larger scale continental trends. Estimated decompositions of local pollution in 2009 indicate that around half of the PM_{10} concentration that year was attributed to external forces. Specifically, in the Lisbon context, particulate levels can be driven by large air masses coming from North African deserts. In 2015, when the region experienced 48 days of atmospheric intrusion by African air masses, there was a significant spike in the monitored values of PM_{10} (*Câmara Muncipal de Lisboa* 2018b).

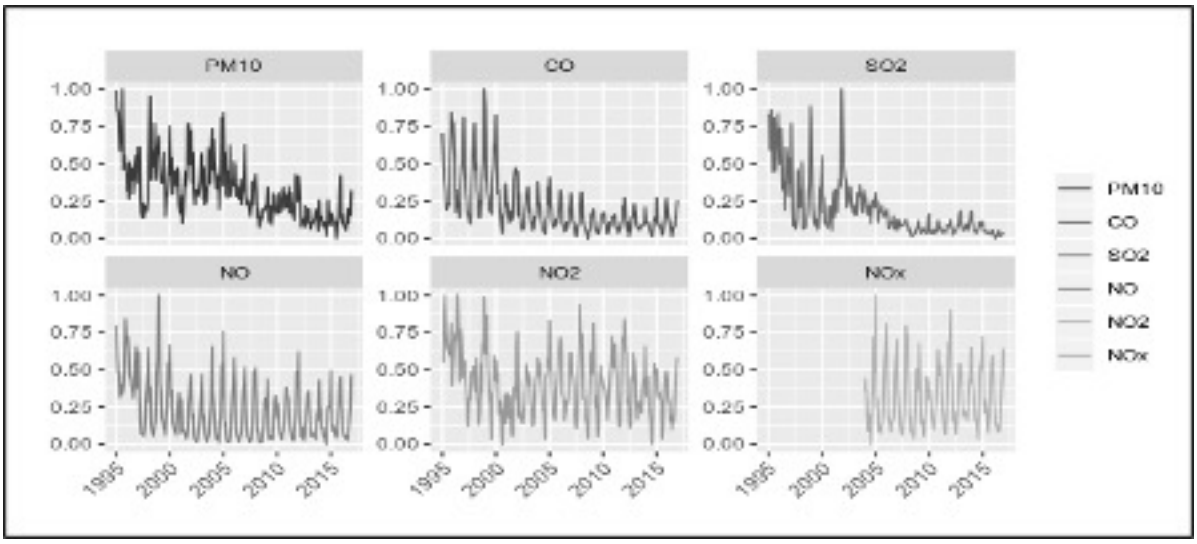
The other family of pollutants common across Lisbon are nitrogen based and primarily attributed to the combustion of fossil fuels and transportation. These include NO and NO_2 which can be more generally measured as NO_X. While these pollutants could be driven by natural combustion forces, at ground level their concentration is attributed to man-made processes and transportation. There are thresholds set for the level of NO_2 mandating that daily maximum levels not to exceed 200 µg/m3 more than 18 times per year.

As the largest city in the country through which much trade and transport occurs, many heavy vehicles and marine transport pass through Lisbon. These types of transportation often use higher amounts of diesel fuel. Additionally, industrial processes related to manufacturing and trade in the city can be large contributors to SO_2 which is commonly associated with acid rain. Both heavy and light vehicle transportation further contribute to levels of CO emissions in the region.

Figure 9 shows the normalized monthly trends of all pollutants in Lisbon using 2000 as a base year for indexing. In 2016 the average concentration of PM_{10} was 24.96 compared to a low (high) of 23.15 (60.99) in 2014 (1995); NO was 21.76 compared to a low (high) of 18.41 (57.89) in 2013 (1996); NO_2 was 33.48, its lowest value compared to a high of 60.90 (1996);

NO_X was 66.25, its lowest compared to a high of 85.08 in 2007; SO_2 was 1.23, its lowest value compared to a high of 14.46 in 1995; and CO was 0.27 compared to a low (high) of 0.25 (0.89) in 2014 (1995).

Figure 9. Standardized Monthly Air Pollution Levels in Lisbon



Partially in response to high levels for some pollutants local environmental quality has become a priority for the municipality. In recent years a significant focus has been dedicated to municipal environmental improvements across many fronts. The city has undertaken ambitious projects in developing urban greenery in the form of tree planting and the provision and maintenance of open spaces, among them the Tagus riverfront running along the South-East edge of the city. Large planned green corridors further aim to link the entire city in an eco-friendly way.

Traffic measures directed specifically at pollution include the introduction of the LEZ around 2012, road restrictions and other residential traffic limitations, the promotion of eco-driving, cycling and public transport. The city has outlined a comprehensive air quality management plan in recent years and based on improvements and planned strategies for the future, Lisbon and the municipal authority, *Câmara Muncipal de Lisboa*, was awarded the European Green Capital 2020.

3 – 3.2. Municipal Transportation Infrastructure and Interventions

Two different transportation initiatives are studied in this work, the opening of new metro stations and the introduction of Lisbon's LEZ in the city centre. While the first looks at the marginal expansion of a public transport network, the second is a traffic limiting policy targeting private ridership and aims to decrease high polluting vehicles in the city centre. Both have the potential to influence the behaviour of drivers in the city by switching to alternative modes of transport or changing their commuting patterns and further, both have the implicit aim of improving transit flows and local pollution levels. Although general comparisons between the resulting pollution reductions can be made, it is important to note how different both types of transit interventions are in terms of their costs, purpose, planning, administration and function.

The Lisbon metro was inaugurated in 1959 with eleven stations running North to South from the historic central business district. Since its inception, stations have been adorned with local art and designed with culturally significant tiles and statues. Different stations, with widely different themes and aesthetics, are known around the world for their uniqueness and attractiveness, and especially for showcasing renowned Portuguese artists and craftspeople.

Construction on the metro continued with nine additional stations opening between 1963 and 1972 after which no new stations were built for almost two decades. Following the political revolution in 1974, the transit system was nationalized in 1975 and has since been run as a public institution. After the nationalization, the metro experienced a revival towards the late 1980's with the construction of new stations and significant expansions to the existing network.

Since 2000, 14 stations have been added to the network, most recently focusing on connecting the international airport and urban peripheries. Currently, the metro consists of 56 stations and 44.5 km of track divided among four lines, with the construction of two additional

stations underway for inauguration in the early 2020's. The system services an average of 600,000 riders per day and in 2018 total ridership was 169 million (Grupo ML 2019).

As with most underground systems Lisbon's metro is electric and the carriages themselves do not release any direct combustion emission. Further, as this system is removed from the aboveground road network, congestion is less of an issue. This is compared to other forms of mass public transit like busses which release exhaust as they travel the city, contributing to local pollution levels. Pollution reductions from alterations to the metro network are thus less related to direct changes to public transit exhaust and more related to spill-over and structural changes caused by increasing accessibility and decreasing aboveground private ridership.

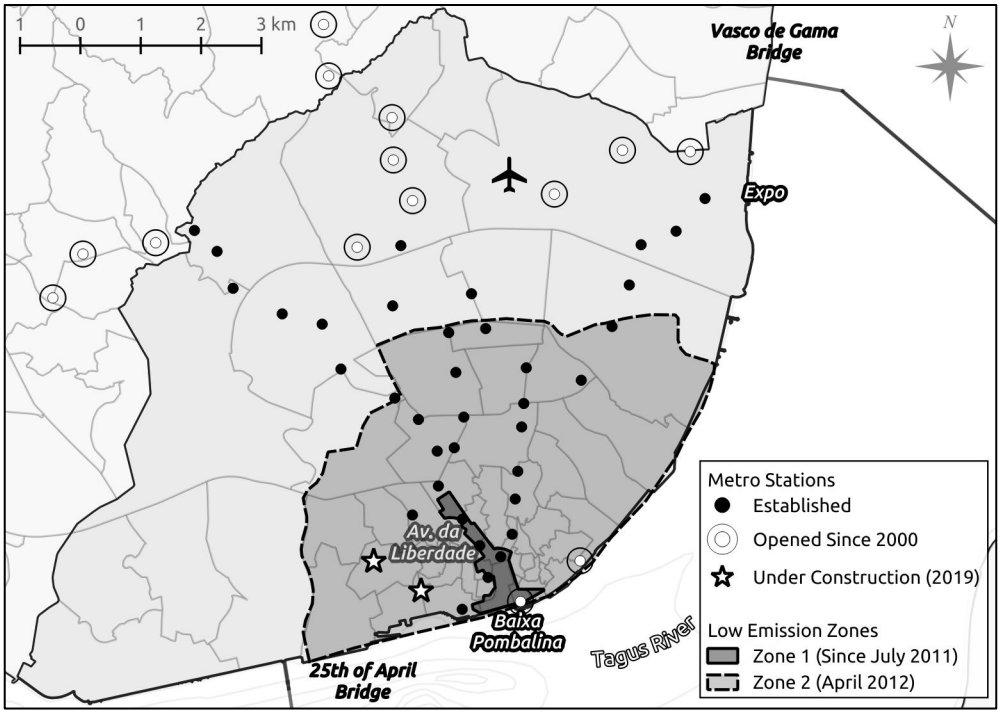
Commuters in the city predominantly use private vehicles for transportation, around 47.7% of residents compared to 19.4% who use busses and 11.63% who use the metro (INE 2011). This contributes significantly to pollution as motor vehicles are primary sources of CO, NO_X, SO_2 and PM_{10}, among other. Vehicle sales in Lisbon are not only much higher, around 71 new vehicle purchases per 1,000 inhabitants compared to the national of 25 per 1,000 in 2017, but further, since 2012, new sales in the capital have risen substantially faster than the national average (INE 2018).

In response to deteriorating pollution levels in the downtown city core, the municipal authority implemented a LEZ in the area with the aim of restricting the worst heavy-polluting transportation fleets. This was part of the larger Air Quality Action Plan introduced in 2006. The current LEZ boundaries cover approximately 30% of the city and are shown in figure 10.

The zone was implemented in three phases starting in 2011. From July of that year to March 2012, during phase one, vehicles from before 1992 were unable to enter an area concentrated around *Avenida da Liberdade*, the primary artery of the city (Zone 1 with a total area of approximately 0.7 km^2). From April 2012, and until January 2015, regulations were

strengthened. The original boundary now restricted vehicles manufactured prior to 1996, and an extended area around the city core was introduced, restricting vehicles from before 1992 (Zone 2 with a total area of approximately 25 km^2). Finally, the last alteration came in January 2015, only allowing vehicles from after 2000 in Zone 1 and 1996 in Zone 2.

Figure 10. Lisbon Transit Interventions: Low Emission Zones and Metro Stations



In terms of enforcement, the LEZ restrictions are upheld by the local traffic. The enforcement fine during the first phase amounted to between €25 and €125 for non-compliance. During the first phase, around 20 fines per month were recorded. Given the relatively *ad hoc* manual enforcement policies, plans were made during the last phase of the LEZ implementation to introduce a network of traffic cameras within Zone 2 with license plate reading capabilities in order to increase enforcement and compliance of the regulations (Gonçalves 2014).

Of particular importance for this study is the fact that Lisbon's LEZ was implemented in conjunction with other policy measures. As the LEZ was introduced, so to were other traffic changes, for example altering road axes to remove parking spaces or restructuring the main

roundabout (*Marques de Pombal*), to manage traffic volumes and congestion (occurring in September 2016). This implies that isolating the impact directly from the LEZ policy is difficult since the resulting effect could come from the variety of transport policies introduced around the same time. For this reason, any estimated effects are likely to be representative of the broader collection of transit policies introduced in and around the LEZ and surrounding areas.

3 – 4. Spatial Interpolation and Aggregation of Sequential Pollution Monitoring

Any analysis and discussion surrounding potential spill over environmental costs or benefits of an urban transport intervention must necessarily involve accurate, representative and adequately detailed data in terms of the spatial and temporal resolution, often one at the cost of another. While the temporal resolution of pollution data in Portugal is very detailed, in many instances down to the hour, geostatistical interpolation and aggregation is used to expand the spatial dimension for each time interval, allowing for detailed within-urban policy evaluation.

A relatively large body of work has been developed along the lines of statistical interpolation and modelling methods for environmental pollutants over space and time, predicting unknown pollution levels in areas and time where no data is observed. Sophisticated spatial, temporal and, or spatial-temporal interpolations have all been employed, in addition to more deterministic models conditional on external influences such as elevation, the built structure of the city or weather patterns, for example in a geographically weighted regression.

The primary goal of this work however is not the creation of multi-dimensional deterministic pollution models, but rather to employ a battery of diagnostics and selection criteria to those interpolation methods most commonly used in applied urban-scale environmental work. This gives rise to a set of generalizable criteria and algorithms for selecting the best method for the generation of a high-frequency and high-resolution longitudinal urban pollution database. With statistically robust and comprehensive pollution

data, it is then possible to evaluate the effects of transit interventions on neighbourhood pollution with sufficient spatial and temporal detail.

3 – 4.1. Interpolation Methods and Techniques

In its raw form the air pollution data represents the concentration of a pollutant (PM_{10}, NO, NO_2, NO_X, CO, SO_2) observed at 20 monitors from inside the study region, as well as those in the immediate proximity in all directions (figure 7). The chosen base temporal frequency to which hourly station level observations are first aggregated is at monthly intervals, balancing the granularity of capturing short-term spatially detailed effects while computationally feasible.

Spatial interpolations are conducted at every month rather than using spatial-temporal interpolations or deterministic modelling across multiple dimensions. This is in part due to the application of interest, using temporal cross sections to estimates changes before and after transit interventions. Increased modelling variables or sophisticated temporal dynamics may smooth the data series too much to observe effects, and the final longitudinal database implicitly accounts for the sequential and cross-sectional nature of pollution.

Each sequence of monthly interpolations is estimated varying the underlying parameters, specifications and variables, comparing diagnostics and practical concerns. The methods are compared based on their predictive power and the continuity and consistency of estimates over space and time. Subsequent aggregation to observational units for the econometric specification further yields diagnostic information regarding the interpolation fit for applied modelling.

Two commonly employed families of spatial interpolation methodologies are used to go from the fixed-point static measures of pollution to a continuous distribution over space for each month. Both the Inverse Distance Weight (IDW) and Kriging models are geo-spatial interpolations able to fill in spatially missing values using only the observed concentration of

pollution and relative distance between observations. Each family differs however in the underlying assumptions, statistical techniques and customization by the user.

The varying forms of all models are compared according to a range statistics and diagnostics evaluating the predictive power and fit of the interpolated estimates. The main diagnostic to compare models is the root mean square error (RMSE) of prediction obtained via a cross validation approach. More practical concerns focus on the distribution of the interpolated values relative to the observational unit size and further representing adequate heterogeneity across space and time.

Selection of Grid for Interpolation and Neighbourhood for Aggregation

It is important *a priori* to determine the base spatial resolutions onto which the interpolated and aggregated data will be projected. For interpolated data this includes the pixels and their sizes representing the continuity of ground truth. The choice of aggregating unit however is conditional on the research question and the model of interest. In this case, we are interested in evaluating how average neighbourhood level pollution levels have changed, and so the observational units should represent spatial neighbourhoods across the city.

A discrete grid of high-resolution pixels is used as a canvas when interpolating from the monitoring station locations. The choice of grid size must ensure an adequately continuous ground distribution of values which accurately represents marginal changes in pollution moving in any direction. The most important of the criteria is that the size of the interpolation grid is smaller than the observational neighbourhood units of interest, yet large enough for feasible computation. Pixels of 100 meters by 100 meters are chosen, and spatially detailed enough within the study region to capture continuous heterogeneity of pollution.[3]

[3] Lisbon has a total size of 100 km^2, and a pixel length of 100 m would represent approximately 1% of the city's horizontal or vertical distance.

For this study, 3,623 census enumeration tracts are used to represent the neighbourhood unit of analysis to which interpolated data is then aggregated and modelled. Census tracts in Lisbon are typically delineated along roadways and other natural barriers of the city and thus implicitly separate areas according to the built environment and natural boundaries. These are broadly aligned with the idea of a small-scale city-block neighbourhood capturing the spatial heterogeneity of pollution across the area.

The distribution, size and boundaries of neighbourhoods in an area may be endogenously determined. Any selection of neighbourhood extent however is based on some *a priori* assumptions conditioning the neighbourhood definition on the size of local population or building density, geodemographics, topography, geographic area, or based on the historic and cultural evolution of the city.

Table 9. Size and Density of Lisbon Neighbourhoods

	Min.	Median	Mean	Max.	St. Dev.
Size (km2)	0.00	0.01	0.02	3.70	0.09
Elevation (m)	0.00	70.38	64.51	201.60	31.95
Building Density (per km2)	0.00	1,199	2,052	20,976	2,478
Population Density (per km2)	0.00	12,644	15,214	99,456	13,148
Dist. to Nearest Metro Station (km)	0.02	0.70	1.30	6.70	2.43
Dist. to LEZ Boundary (km)	-0.18	3.08	3.25	7.97	2.22
Dist. to Baixa (km)	0.04	4.15	4.25	9.44	2.43
Dist. to Tagus (km)	0.00	1.75	2.46	8.08	2.06
Dist. to Nearest Freeway (km)	0.00	0.39	0.49	2.78	0.42

Census Tract Level Data (N = 3,623 neighbourhoods)

If neighbourhood units are not structurally dissimilar in terms of their relative size and density over space and time, then model estimates can capture average effects by controlling for these spatial and temporal differences. Table 9 gives some underlying statistics regarding the neighbourhood units chosen for the aggregation of pollution in terms of neighbourhood size, density and locational features.

Inverse Distance Weight Interpolation

The IDW family of interpolations predict pollution values at locations for which no observed measure exists based on weighing observed values from a location proportional to

the distance between the locations and some weighting parameter. The general formula for obtaining the interpolated values from the IDW is as follows, where the predicted interpolated value $\mathbb{P}$ at a location x (pixel) for which no data exists is determined according to:

$$\mathbb{P}(x) = \begin{cases} \frac{\sum_{i=1}^{N}[1/d(x,x_i)^{\rho}] \cdot \mathbb{P}(x_i)}{\sum_{i=1}^{N}[1/d(x,x_i)^{\rho}]} & \text{if } d(x,x_i) \neq 0 \\ \mathbb{P}(x_i) & \text{if } d(x,x_i) = 0 \end{cases} \qquad \text{Eq. 3 – 1}$$

If the distance between the pixel location and any of the monitoring stations i is equal to zero ($d(x,x_i) = 0$), then the pixel is exactly at a monitoring location and is assigned that value, $\mathbb{P}(x_i)$. Otherwise, if there is some positive distance between the observed value and the pixel, then the interpolated predicted value is conditional on two parameters: the choice of the number of nearest neighbours, N, and the inverse distance power, ρ. The value of pollution from neighbours (monitoring stations) is weighted according to the respective distance between the observed location and the interpolated pixel of interest.

Given the infinite possible combinations of N and ρ parameters, hyperparameter fine-tuning is done to select among different combinations of realistic parameter sets representing the most commonly used and extreme limit cases. Selecting the best model by varying parameters among a select few within a pre-determined group greatly improves the speed and efficiency in determining the best model. This fine-tuning is done by running all combinations of the interpolation (and respective aggregation) of values and comparing the predictive power among all choices of parameters. How each of the parameter combinations effect the different variants of pollutants over time can help in better understanding the dynamics of pollution in terms of its relative decay and continuity over space from the observed monitored locations.

Four choices for each parameter represent the *de facto* commonly used values and limit cases. The possible values of each parameter are given in table 10. The 16 combinations of all parameters capture some extreme cases, for example, using all monitoring stations N and no decay, $\rho = 0$, attributes to every pixel the average pollution level from across Lisbon. While a

combination of parameters like this is acceptable from a geostatistical point of view, this would not provide any spatial variability and would not be of any use for evaluating high-resolution and high-frequency urban policy initiatives.

Table 10. IDW Nearest Neighbour and Weight Parameter Combination

$N = 1$ The singular nearest monitoring station.	$\rho = 0$ A weight of zero indicating no decay in the value moving away from the observed location.
$N = 3$ The three nearest monitoring stations aimed at capturing a broad triangulation around the pixel.	$\rho = 1$ A linear decay proportional to the distance between the observed value and the prediction location.
$N = All$ All monitoring stations in the study region.	$\rho = 2$ Squared inverse distance weight assigning higher weight to closer observed values relative to the linear.
$N = CVOpt.$ A cross-validated determined optimal number of neighbours.	$\rho = CVOpt.$ A cross-validated determined optimal weight parameter.

Given the relatively small sample size of monitoring stations with observed values, the cross-validated (CV) optimal versions of each respective parameter are obtained using a leave one out cross-validation (LOOCV) approach. This method iteratively uses all the data, removing one observed value at a time and running the IDW model sequentially through each removal of an observation. For each sequence, an optimization function determines the values of N and ρ which minimize the RMSE.[4]

This generates a set of optimized values and RMSE criteria for each iteration (removal) of a monitoring station. For each pollution-month interpolation there is a vector of LOOCV RMSE estimates with length equal to the number of monitoring stations. The selected CV optimized set of parameters are the median within-value of all the N and ρ estimates which minimize their respective iteration of the RMSE. The median is chosen so that selected parameters are not sensitive or driven by outliers or any spurious training sample chosen. When comparing results, most frequently the minimum RMSE is attributed to the median parameter values.

[4] The optimization searches for the combination of N and ρ parameters which will minimize the RMSE function according to the Nelder–Mead methodology with initial parameter values of $N = 3$ and $\rho = 2$.

Kriging Interpolation

The Kriging family of models are similar to the IDW in that they both interpolate values at areas for which no observations are observed conditional on the locations from which there are measurements. Differences arise however in the underlying assumption regarding how to weight the contribution of measurements from different locations. While the IDW bases this weight on the relative distance between a location and its neighbours for which there are observations, the Kriging prediction bases this weight on a Gaussian process. The general format of the Kriging process in in line with the IDW process and can be represented according to equation 4.2.

$$\mathbb{P}(x) = \sum_{i=1}^{N} \omega_i \, \mathbb{P}(x_i) \qquad \text{Eq. 3 – 2}$$

where the inverse distance weights are replaced by a series of optimized weights, ω, minimizing the square deviation between predicted and observed values, much like a regression specification.

The underlying assumptions related to the structure of the expectation of the observed monitoring station values, $\mathbb{P}(x_i)$, will influence the choice of model specification and complexity. The Ordinary Kriging specification is used if the underlying pollution variable is assumed to come from a random data generating process where the mean is a constant unknown value with random disturbances, $\mathbb{P}(x_i) = \mu + \varepsilon_i$. This would imply that the underlying pollution data follows a spatially stationary process.

If the underlying process varies deterministically, then a trend component can be included. This assumption forms the basis of Universal Kriging specifications, where the expected value of pollution varies deterministically according to other processes, $\mathbb{P}(x_i) = \mu_i + \varepsilon_i$, where the expectation, μ_i, can be expressed in terms of covariates $\mu_i = \sum_k \beta_k z(x_i)$, for k potential predictors in vector z, conditional on location x_i.

The baseline Universal Kriging specifications here is the inclusion of latitude and longitude of the prediction locations as determinants of the predicted value. This would be an appropriate model if we expect some underlying trend where the average pollution varies deterministically across space – potentially due to some external, unobserved spatial factors such as the variability in wind speed and direction. So, while the set of Ordinary Kriging specifications use an underlying Gaussian process to estimate predicted values, the Universal specifications include a deterministic component. If we expect location to be significant, then the Universal Kriging models should outperform the Ordinary counterparts.

While high-dimensional Kriging specifications can be developed, the goal of this work is not the sophisticated modelling of interpolated air pollution. Additional auxiliary variables can be included to enhance the Universal Kriging model; however, their inclusion is outside the scope of this work. This would require that additional external variables are available at a complete spatial coverage such that each pixel of interest onto which we want to interpolate has underlying data from which to build a model.

There is however one exception which does not require any additional data beyond what is openly available. Given that a sequence of temporal interpolations is conducted at every month, important auxiliary influences can be included in the model in the form of the lagged values of pollution without the burden of having to obtain data on external factors. This could potentially result in significant prediction accuracy increases with minimal additional data processing.

At each time period both baseline Ordinary Kriging and Universal Kriging specifications are enhanced by the inclusion of up to two periods of lags of predicted values. If there is some residual temporal dynamics involved in the prediction of pollution, as would be expected given the continuous nature of these variables over time and space, then the inclusions of past prediction values could improve upon the baseline versions.

Thus, six specifications of Kriging are estimated. The standard Ordinary Kriging specification, which considers that pollution has a random unobserved mean, the Universal Kriging model which considers that pollution is deterministically influenced by the inherent location across space, and further two additions to each of these base specifications which look at the inclusions of one and two temporal lags.

3 – 4.2. Spatial Interpolation Diagnostics and Choice

Spatial interpolation and aggregation are statistical concepts and so the choice of model, sequences and methods to follow can vary widely in different contexts and with different data. There are however a series of diagnostic tools and practical ideas which can be implemented to ensure that the choice of interpolation produces the best and most robust statistical series for each pollutant. This further ensures that any discussions or inferences regarding transit interventions are not driven by the choice of model.

This section outlines the series of criteria and diagnostic inferences guiding the choice of interpolation for generating high-frequency spatial and temporal pollution measures. There is a separate series of interpolations conducted for every pollutant at each month interval since the beginning of their series. This includes 1,307 pollution-month combinations, omitting those with missing values or fewer than six active monitoring stations. Diagnostics are compared between all combinations of methods described, 16 variants of IDW and six variants of Kriging, proving a wealth of information from which to draw conclusions.

One additional point of note is the choice of variable on which to interpolate, namely whether to interpolate directly on the concentration of pollution or whether there are efficiency gains by interpolating directly on the log value of the pollution.[5] With certain interpolation

[5] Because the econometric variable of interest is the log of pollution (to estimate percentage change) there is no need to do any post-interpolation transformation and therefore is less alteration to the data. If the final variable of interest is not in log form, then additional cross-validation should be conducted on the post-transformation to check deviations from the original non-transformed ground truth.

methods, namely the Kriging specifications, relying on an assumption of Gaussian residuals, the pre-transformation of pollution concentrations to log form may provide a better fit and distribution of the observed values to interpolate.

Decision Criteria

Several different diagnostics are used to evaluate the overall model performance based on type, parameter selection and variable of interpolation. Given that a temporal sequence of spatially interpolated and aggregated values are generated, no one singular diagnostic value can adequately evaluate the overall model performance across the entire spectrum of interpolations. Diagnostics generated for every month, pollution and model combination will be evaluated using linear regressions to determine how different parameters or model selection influences predictive power.

The criteria for selecting the appropriate interpolation is based on several conditions, however the driving decision should be selecting models with the highest accuracy. The first step is disregarding any observations with inadequate representation, here by removing interpolations with less than six active monitoring stations or missing values.

The predictive accuracy is evaluated using the RMSE.[6] Given the relatively small number of monitoring stations, the overall RMSE for each model is estimated using a LOOCV approach, systematically removing one monitoring station at a time and evaluating the RMSE between the predicted value and the observed value left out. The average of these RMSE across all the sequentially left-out monitoring stations provides the index upon which we compare all model specifications. A low RMSE indicates that the model more accurately predicts the ground truths.

[6] The general formula for the RMSE evaluates the difference between some value of x and it's expected (or predicted) value: $RMSE = \sqrt{[avg((x - \bar{x})^2)]}$

The distribution of the interpolated values is also important to consider, specifically whether the distribution of values obtained correspond to the distribution of values following the aggregation to the neighbourhood (census tract) level units. This is important for considering whether the results from the interpolation have a spatial resolution which is detailed enough to accurately be aggregated without changing the underlying structure of the pollution values.

A mismatch between these distributions would signal an additional skew being added to the data in moving from the interpolated values to the aggregated values, potentially biasing any results. The is evaluated using the Kolmogorov-Smirnov test statistic to determine if the interpolated and the aggregated series are drawn from the same continuous empirical distribution.

The distributional concern should further be complemented by looking at the continuity and variation of interpolated values. As is particularly the case with IDW models with zero weights, the interpolated value could lack spatial variation if the predictions represent simple unweighted averages. This is evaluated by considering the number of unique values obtained from interpolation and ensuring this is at least as large as the number of observational units so as not to introduce further alterations to the data.

Finally, it is important to have temporal consistency in the choice of specification. Since different pollutants are entirely different series the choice of model can vary depending on pollution but within, there should be a consistent method for generating the data across the time span. Selecting different specifications at different intervals will change the underlying assumptions in the data series. Given the interest is in estimating temporal breaks, it is necessary to reduce any time inconsistencies from different methodologies.

Diagnostics and Final Selection of Interpolated Data

To guide the choice of interpolation model, a set of linear regressions are performed to see how different parameters, specifications, and variables influence the overall prediction accuracy and fit of the interpolated values. For every pollutant-month combination, this diagnostic information is available for all interpolations and further for comparing each specification using interpolation on the direct concentration and interpolation on log values.[7]

A regression model estimates the impact that model specification, number of active stations, time and distribution of value variability (as measured by the number of unique interpolated values) have on the RMSE of prediction. A pooled model first controls for these characteristics, and further for each pollutant and respective pollutant-specification interaction. Subsequent pollution-specific models are estimated to determine the best specification for each data series. Results on these diagnostics are presented in table 11.[8]

The full specification with control variables significantly determines the variability of the prediction power of models with an R^2 of 82.5%. As the number of monitoring stations increase, the variation in the RMSE decreases (t-value of -28.38), and this is also seen as we consider observations occurring in later years (decrease in t-value from -10.42 in to -39.43 in 2013-2016). While robust and accurate modelling can significantly enhance the prediction, these results show that inherently the power of any model is conditioned by the external availability of data.[9]

[7] This gives $1{,}307 \cdot (16 + 6) \cdot 2 = 57{,}508$ sets of parameter combinations and resulting accuracy measures.
[8] t-values estimate how many standard deviation reductions (or increase) each specification has on the RMSE from the study-wide average value. Direct effects shown only and full values for all diagnostics available upon request.
[9] While not shown here for brevity, similar results are found when considering as a dependent variable the Kolmogorov-Smirnov test statistic. The more monitoring stations and later collection time, for example, relatively reduces the KS statistic, indicating no significant difference between the interpolated and respectively aggregated distribution of values. This captures the idea that the interpolated values must be well suited for the observational units, and a lower KS statistic indicates a closer correspondence between the interpolated and aggregated values.

Table 11. Interpolation Model Diagnostics: Full and Sub Linear Regression t-Values

	Full (Model 1)		**PM_{10} (Model 2)**		**NO (Model 3)**		**NO_2 (Model 4)**		**NO_X (Model 5)**		**CO (Model 6)**		**SO_2 (Model 7)**	
Dep. Variable: RMSE of Prediction	*Level's*	*Log's*	*Level's*	*Log's*	*Level's*	*Log's*	*Level's*	*Log's*	*Level's*	*Log's*	*Level's*	*Log's*	*Level's*	*Log's*
Nearest Neighbour: 1; Weight: 0	0.00	-90.99	17.12	-40.90	7.89	-23.89	8.16	-39.36	10.00	-35.72	1.28	17.09	1.32	-12.62
Nearest Neighbour: 1; Weight: 1	0.00	-90.99	17.12	-40.90	7.89	-23.89	8.16	-39.36	10.00	-35.72	1.28	17.09	1.32	-12.62
Nearest Neighbour: 1; Weight: 2	0.00	-90.99	17.12	-40.90	7.89	-23.89	8.16	-39.36	10.00	-35.72	1.28	17.09	1.32	-12.62
Nearest Neighbour: 1; Weight: CV	0.00	-90.99	17.12	-40.90	7.89	-23.89	8.16	-39.36	10.00	-35.72	1.28	17.09	1.32	-12.62
Nearest Neighbour: 3; Weight: 0	-13.86	-91.24	7.11	-41.33	4.43	-24.31	1.30	-39.63	3.05	-35.83	0.64	14.91	-2.47	-13.16
Nearest Neighbour: 3; Weight: 1	-12.47	-71.08	-0.95	-16.88	0.12	-11.86	-1.09	-18.17	3.01	-4.86	1.69	8.17	-0.97	-6.06
Nearest Neighbour: 3; Weight: 2	-12.38	-71.07	-0.78	-16.87	-0.29	-11.85	-1.31	-18.17	3.02	-4.86	1.70	8.24	-0.55	-6.02
Nearest Neighbour: 3; Weight: CV	-12.28	-71.09	-1.26	-16.89	-0.27	-11.87	-1.40	-18.18	3.04	-4.86	1.68	8.00	-1.12	-6.11
Nearest Neighbour: All; Weight: 0	*Ref.*	-91.30	*Ref.*	-41.54	*Ref.*	-24.48	*Ref.*	-39.61	*Ref.*	-35.88	*Ref.*	12.49	*Ref.*	-13.09
Nearest Neighbour: All; Weight: 1	-15.77	-71.14	-2.83	-16.95	-1.01	-11.92	-1.36	-18.18	2.57	-4.87	1.50	7.30	-0.90	-6.09
Nearest Neighbour: All; Weight: 2	-15.73	-71.13	-2.09	-16.92	-1.13	-11.89	-1.82	-18.19	2.57	-4.87	1.57	7.56	-0.64	-6.07
Nearest Neighbour: All; Weight: CV	-15.41	-71.13	-3.16	-16.96	-1.40	-11.93	-1.36	-18.17	2.62	-4.87	1.54	7.26	-0.61	-6.13
Nearest Neighbour: CV; Weight: 0	-14.80	-91.25	6.15	-41.35	2.94	-24.35	0.38	-39.65	2.58	-35.83	0.51	14.60	-2.52	-13.23
Nearest Neighbour: CV; Weight: 1	-13.07	-71.09	-1.17	-16.89	-0.29	-12.00	-1.39	-18.38	2.93	-4.86	1.66	8.21	-1.00	-6.13
Nearest Neighbour: CV; Weight: 2	-12.82	-71.08	-0.92	-16.88	-0.50	-11.99	-1.50	-18.37	2.96	-4.86	1.68	8.30	-0.56	-6.08
Nearest Neighbour: CV; Weight: CV	-12.81	-71.10	-1.48	-16.90	-0.37	-12.01	-1.61	-18.38	2.96	-4.86	1.65	8.06	-1.21	-6.19
Ordinary Kriging	-17.63	-71.95	-3.32	-16.91	-1.07	-12.27	-1.79	-18.90	2.33	-5.08	1.51	7.52	-0.72	-7.23
Ordinary Kriging: 1 Temporal Lag	-46.68	-71.80	-8.15	-17.11	-7.03	-12.51	-8.88	-18.90	-1.60	-5.00	0.75	4.79	-3.70	-7.94
Ordinary Kriging: 2 Temporal Lags	-46.09	-71.65	-8.18	-17.08	-6.94	-12.46	-8.84	-18.84	-1.52	-4.97	0.81	4.88	-3.57	-7.85
Universal Kriging	-9.92	-71.00	-1.95	-17.94	0.25	-11.81	0.77	-18.06	3.35	-4.85	1.80	8.30	0.56	-5.74
Universal Kriging: 1 Temporal Lag	-45.27	-71.52	-8.11	-17.78	-6.74	-12.26	-7.75	-18.34	-1.40	-4.93	0.87	4.28	-2.58	-6.77
Universal Kriging: 2 Temporal Lags	-43.77	-71.40	-8.03	-17.69	-6.65	-12.24	-6.50	-18.27	-1.20	-4.92	1.00	4.55	-2.24	-6.62
Observations	56,996		7,864		8,656		9,292		6,808		8,300		8,212	
Pollution Controls	Yes		No		No		No		No		No		No	
Pollution-Specification Interactions	Yes		No		No		No		No		No		No	
Adjusted R^2	0.8241		0.8667		0.6449		0.7949		0.8206		0.4301		0.3636	
Residual Std. Error	3.2602		1.3260		4.5601		2.5763		6.4886		0.2555		1.6824	
F Statistic	853.85***		1,044.70***		321.79***		735.75***		636.44***		128.83***		96.72***	

The choice of model specification for each pollutant is done by estimating pollution-specific sub-models with relevant time and active station controls. The reference interpolation category used is all nearest neighbours with a weight of zero – or in other words the Lisbon level average of the respective pollution at any month in time. The negative t-value attributed to each model specification therefore represents the relative decrease in the average RMSE, resulting in a better prediction power for that give specification over the alternative simple average.

Clear gains are seen when using the log version in terms of reduced RMSE. This is particularly the case when comparing the Kriging specification of models which are conditional on an underlying Gaussian distribution. This suggests that, when possible, conducting pre-transformation of the data should be done to create a more standard distributed variable.

Overall, Kriging specifications are superior for most pollutants and the inclusion of one temporal lag provides the better fit. While the inclusion of a secondary temporal lag in the prediction yields strong results, compared to one temporal lag it appears that these models may be overfitting the data series. In terms of the IDW family of models, it is always best to include all observations in generating the predicted value rather than a localized subset of monitoring stations.

The RMSE of prediction for PM_{10} is well described by the choice of model, with an R^2 value of 86.67%. For the interpolation of PM_{10} values, similar gains in prediction power come from using all the available data in an IDW model and using the Kriging specification. Still, however, the Kriging family of models have a clear gain over the IDW with better fit and significantly larger reductions in the RMSE of prediction. The preferred PM_{10} specification is the Universal Kriging using one temporal lag. This specification is also the best choice for modelling of SO_2, consistently better than any of the IDW models. Suggesting a spatial deterministic component in the distribution of these series.

In terms of combustion-based pollutants, NO and NO_2, the Kriging specifications using log pollution values are superior to the alternatives. In order to choose between potential specifications, the top general two best IDW and Kriging interpolations are estimated, including both the Universal and Ordinary Kriging with one temporal lag and the IDW using an optimally determined weight parameter with three and all nearest neighbours. Across the entire time series, a tabulation of which model has the lowest RMSE at every month out of the potential top candidates shows that for NO and NO_2 the Ordinary Kriging model with a temporal lag has the best prediction 119 times out of 198 and 137 times out of 213 respectively.

No consistent model had any significant improvement in the prediction power of CO pollution relative to using the study-wide average value. The different specifications explain little of the variation in RMSE with an R^2 of 43.01%. This suggests that the interpolation may not accurately be capturing the true spatial dynamic of this series and more complex underlying features may be influencing local CO values.

Even though the average spatial interpolation does not generate robust predictions, this does not mean that the empirical strategy will not be able to capture the temporal dynamics before and after a transit intervention. The Kriging specification of models remain relatively those with the smallest RMSE. This allows us to have some spatial variability compared to the city-wide average yet provides some level of predictive power over the alternatives. The Ordinary Kriging with one temporal lag is the best prediction 119 times out 191 and is chosen as the preferred specification for CO. Care is taken with these results however, and limited inferences are made with discussion focused on general trends and patterns.

Not all pollutants are well predicted by the Kriging models, and results indicate NO_X is best predicted using models with weights of zero. This would seem to suggest that this pollutant is also better predicted by localized averages without any weighting or Gaussian-based interpolation, and thus has less decay over space and inherently less spatial variability.

To balance the need for concentration heterogeneity with choosing the model minimizing the RMSE, the three nearest neighbours are chosen with a weight of zero. This uses the local average of the closest monitoring stations without any spatial decay.

Figure A3 – 1 of the appendix shows selected pollution-month interpolated spatial grid and corresponding aggregated neighbourhood observation units for the respectively chosen preferred interpolation model.

3 – 5. Empirical Spatial-Temporal Impact Estimates of Urban Transit Initiatives

The mean aggregation of the preferred interpolated data to the neighbourhood level observation units across monthly intervals yields a high-dimensional spatial-temporal longitudinal panel database with the same neighbourhood's pollution concentration repeated at every month. The empirical methodology employed makes use of this structure comparing different neighbourhoods across space before and after the introduction of the transit initiatives.

Under a panel data difference-in-difference econometric strategy the long-run average reduction impacts for various pollutants can be estimated. This is further complemented by a non-parametric bootstrapping procedure to estimate an equivalent short-run difference-in-difference average treatment effect for each impact. This bootstrapping overcomes the limited short-run sampling allowing us to further estimate temporal decays. These estimated spatial-temporal impacts are based on taking the temporal introduction of the transit initiatives and comparing the differences before and after while accounting for the varying spatial orientation of neighbourhoods, namely by considering those that are closer to key areas of the city where we would expect pollution levels to be altered following an intervention .

3 – 5.1. Spatial-Temporal Difference-in-Difference Specification

A variety of models are estimated for each of the transit interventions of interest to evaluate the local impact of pollutants, and any spatial decaying effects, from key locations

around the city related to metro stations and the LEZ. Here, the same econometric model is applied to all the pollutants to observe how equivalent interventions impacted the various pollutants differently. The empirical specifications in equations Eq. 3 – 3 and Eq. 3 – 4 control for the heterogeneity of neighbourhoods while estimating the treatment impact of the respective transit intervention by difference-in-difference across space and time.

$$ln(\mathbb{P}_{it}) = \beta + \alpha_i + \beta_1 \text{Location}_i + \beta_2 \text{Metro}_t + \beta_3[\text{Location}_i \times \text{Metro}_t] + \beta_4 \text{Year} + \varepsilon_{it} \quad \text{Eq. 3 – 3}$$

$$ln(\mathbb{P}_{it}) = \beta + \alpha_i + \beta_1 \text{Location}_i + \beta_2 \text{LEZ}_t + \beta_3[\text{Location}_i \times \text{LEZ}_t] + \beta_4 \text{Year} + \varepsilon_{it} \quad \text{Eq. 3 – 4}$$

Here, $ln(\mathbb{P}_{it})$ represents the log value of the various pollution measures for the $i = 1, \dots, 3{,}623$ census tract neighbourhoods for each monthly interval from $t =$ January 2000 to December 2016. An overall average effect is captured in the intercept, β, while neighbourhood level heterogeneous effects are captured in the fixed effects, the individually varying parameters, α_i. The ε_{it} are classical Gaussian error terms. Year controls are included to capture broad annual trends in concentration trajectories.

The Location_i variable indicates the $\{0, 1\}$ assignment to one of the chosen spatial treatment areas, indicating whether a neighbourhood is in, for example, *Baixa* or near a metro station recently opened. The Metro_t and LEZ_t represent the temporal treatment, assigning pre and post-transit intervention classification.

The specifications above estimate the direct impacts of both the space (location) treatment and the temporal (transit) treatment, as well as the interaction between these two which represents the average treatment effect, β_3, via a difference-in-difference estimation. In this scenario the first difference represents the change in pollution captured before and after a transit initiative, while the second difference captures whether a neighbourhood is proximate or not to five chosen locations across the city where we would expect pollution to be impacted – near the newly opened metro, along the riverfront, in the central business district, near the LEZ or along busy thoroughfares.

If we expect that there are important omitted neighbourhood variables which are not measured or cannot be observed, then a fixed effect estimation can address this omitted heterogeneity. One of the key assumptions in the use of these neighbourhood level fixed effect controls however is that any unobserved and omitted heterogeneous effects among them is time invariant. That is, across the timespan no neighbourhood experienced structural change in their trajectory or dynamics over and above the changes other neighbourhoods experienced.

The fixed effect, or within group, uses the individual group mean to identify the impact of intervention over space and time. This means that the parameter values are estimated using a standard OLS applied to the within-group demeaned values. This estimation is numerically equivalent to including a dummy variable for each census tract neighbourhoods. As the estimation is done on demeaned variables no general intercept value is obtained from the estimation.

Choice of Temporal Treatment

This work looks specifically at two groups of transit related interventions introduced over time. Exploiting the timing of these different initiatives is used to estimate and compare their mitigating impacts on localized airborne pollution across the city. While the mechanisms of these interventions are very different, the environmental aims of reducing air pollution are the same. The goal of this chapter is not to compare and evaluate these two transit interventions against each other, but rather to explore the patterns observed in the generated longitudinal neighbourhood pollution data against known transportation related changes with spatial and temporal dynamics.

The two temporal treatment variables are introduced in a cumulative way. For estimating the impact from opening metro stations, the variable represents the cumulative number of stations opening over the course of the study period. This controls directly for the existing density of the transportation network in place at any given time. The estimate therefore

represents the marginal change of an additional metro station to the existing network of stations already opened.

In terms of the LEZ, the treatment variable is cumulative in the sense of considering the relative intensity of when the policy entered into effect. Different levels of strictness accompanied the introduction of the different phases of the LEZ, and the cumulative variable captures this increasing intensity and restrictiveness as it was phased in over time.

Choice of Spatial Treatment

In identifying the spatial impacts of interest, it is necessary to determine the extent to where we would expect inter-urban air pollution to be most affected by each transit intervention. This could be highly dependent on the study region of question and conditional on local topographic constraints and the built structure of the city. It is not always clear exactly where and how local transit patterns may respond to various changes to transportation infrastructure, and so spill-over impacts from the creation of a metro station in the urban periphery, for example, could be felt in *Baixa* if periphery residents change their behaviour or commuting patterns into the city centre.

Five different key areas around the city are considered: proximity to the metro station which has opened, proximity to the LEZ boundary, distance to the city centre (*Baixa*), distance to the Tagus riverfront, and distance to the nearest freeway. As metro stations open and public transit becomes more accessible, or with the introduction of a traffic limiting LEZ in the city centre, the alterations to commuting and traffic patterns could impact pollution concentration in any combinations of these areas.

One of the key assumptions when estimating impacts is the stable unit treatment value assumption. This states that the treatment assignment of one observation (in a neighbourhood 'nearby' one of the key areas) does not affect the potential outcome of others (pollution levels in the 'non-nearby' neighbourhoods). This could particularly be the case with a spatial

treatment assignments such as proximity to key areas in the city. Given the continuity of pollution over space, the cut-off treatment threshold between what defines ‘nearby’ to the area of interest or not can be subjective. Proximate neighbourhoods are necessarily influenced by adjacent pollution, and so the distinction between these spatially treated units and controls can be fuzzy and should be addressed.

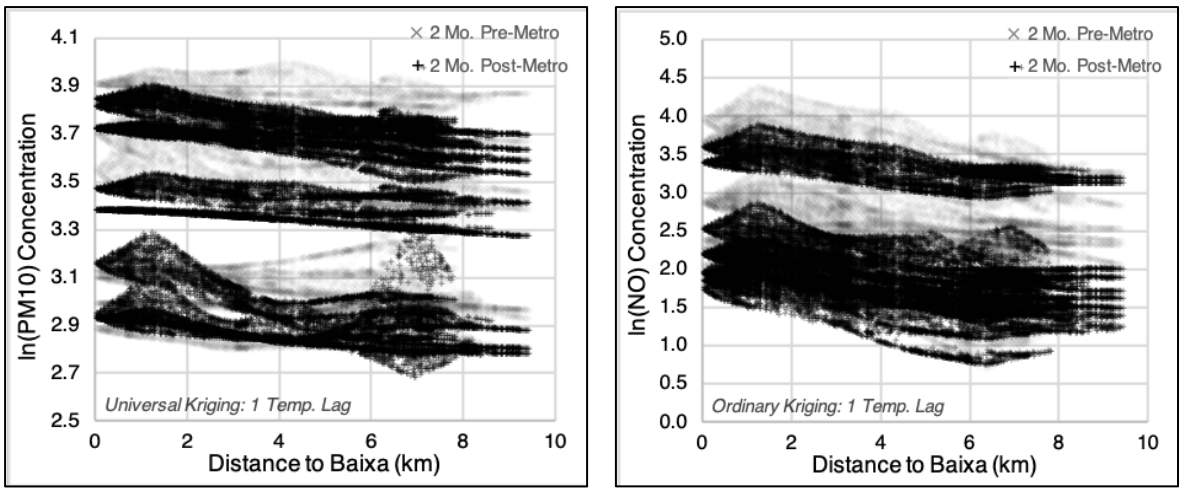
The identification of this cut-off threshold defining the $\{0, 1\}$ spatial treatment needs to represent a positive distance at which point neighbourhoods are considered in proximity to a given area of the city where pollution is likely impacted. These spatially treated units are then compared to those far enough away from these areas so as not to be influenced by the intervention effect (spatial controls). Depending on the context and underlying understanding of the study region, this could range from, for example, considering nearby neighbours as those very local units within 100 meters of downtown versus those within 1,000 meters if we would expect that the transit intervention has a larger impact over space.

To address these possible concerns, a sequence of potential spatial treatment assignments is estimated for all combinations of pollution and interventions. This takes, for example, the distance to the LEZ boundary, and assigns the spatial treatment classification sequentially moving marginally away from the location. This enables a plot of the decaying fuzzy effect that various treatments may have on pollution over space comparing neighbourhoods first at 100 meters away and then at sequentially larger distances.

Each fixed effect model is at every potential spatial treatment assignment in 50 meters intervals between 100 and 500 meters. At larger buffer distances, the impacts are estimated at every 250 meters from 750 meters up to 3 kilometres. The rationale behind estimating the spatial impacts at such a large distance can be seen by looking at an example plot of how PM_{10} or NO concentration decays over space for neighbourhoods pre and post-transit intervention.

Figure 11 shows that decreases in pollution concentration following the introduction of a LEZ was felt up to around 1.25 kilometres away from *Baixa* by comparing the spatial decay two months before and after treatment. Prior to the treatment, pollution effects increased around the city centre locally for both PM_{10} and NO after which point levels begin dropping. After the treatment however, we see consistent spatial decay in levels. This effect is highly conditional on the spatial concept considered, and for example, the spatial decay located around freeways is much more localized and the estimated impacts therefore are only estimated much more locally. Figure A3 –2 of the appendix shows the similar plot for the impact of proximity to freeways.

Figure 11. Spatial Decay of PM_{10} and NO Pre and Post Treatment in Baixa



3 – 5.2. Bootstrapped Short-Term Pollution Reduction Impacts

The estimation of the fixed effect parameters in the difference-in-difference specification makes use of the entire time series of data spanning from the start of monitoring to the end of 2016. The estimated impacts are therefore representative of the average effects spread out over this relatively long span. With limited data in the periods directly before and after a given transit intervention, it is difficult to use a panel data structure to estimate the short-term immediate effects occurring directly following a treatment.

To overcome this limitation, a non-parametric bootstrapping process is used to estimate the immediate temporal effects in the subsequent months following an intervention, and the respective decay of these effects both over space and over time. The bootstrapped average treatment effect is based on systematically increasing the estimation by including new observations in one-month intervals before and after the introduction of the treatment.

In considering only observations directly around any of the respective interventions, the sample size of the data decreases, and so too does the statistical power and variation captured by the original model. The non-parametric aspect means that the estimate returns the bootstrapped value of the mean differences across space and time treatments as opposed to being estimated through a deterministic linear model as is done when a longitudinal database format is available. The bootstrap statistic, β_{BS}, which is evaluated for the subsample of temporally proximate observations for all intervals from one month to a year is:

$$\beta_{BS} = Avg.\left[\overline{\mathbb{P}}^{Sp.Treat=1}_{Tm.Treat=1} - \overline{\mathbb{P}}^{Sp.Treat=0}_{Tm.Treat=1}\right] - Avg.\left[\overline{\mathbb{P}}^{Sp.Treat=1}_{Tm.Treat=0} - \overline{\mathbb{P}}^{Sp.Treat=0}_{Tm.Treat=0}\right] \qquad \text{Eq. 3 – 5}$$

The same set of spatial treatment effects as in the parametric model are used, and so a bootstrap estimate is obtained for increasing distances from *Baixa*, the Tagus riverfront, opening metro stations, the LEZ boundary and freeways. The temporal treatments, however, are converted to dummy variables representing the month of the intervention. This provides an indicator to identify the specific base reference month when a metro opened, or the LEZ was introduced. From this reference month, estimates for the immediate temporal effects can be calculated by systematically estimating the average impact in the first month post-treatment, the second month post-treatment, and continuing sequentially for up to one year following the transit intervention.[10]

[10] In the case of multiple cumulative treatments, such as opening new metro stations, the mutually exclusive before and after observations are taken. Any observations which happen to fall in the same number of months prior to the opening of a station and equally within the months following the opening of another station, are removed.

For each combination of treatment effect of interest, and now further for each month following the intervention, a bootstrapped resampling is used to estimate the average treatment effect. The idea behind this is to use the limited temporal subsample of observations directly immediate to the transit intervention and resample these values with replacement. The average treatment effect is then estimated for the resampled levels of pollution. A total of 500 iterations for each model are run with estimated impact coming from average effect. Bootstrapped standard errors are also calculated and allow us to evaluate the significance of any estimated impacts.

3 – 6. Impact Estimations

This section outlines the results of the estimation for the impact evaluation of the two transit policies of interest: the opening of new metro stations and the introduction of a LEZ. The empirical strategy first looks at the long-run effect of either transit initiative on pollution levels in key areas of the city. This is estimated via difference-in-difference to get the average impact on different neighbourhood locations pre and post intervention. The pollution variables used correspond to the best interpolated model for each respective pollutant as described in section 4.

The difference-in-difference strategy uses the entire time series of data starting after 2000. For every pollutant, this includes up to 216 months and 3,623 neighbourhoods for over 700,000 total observations per model. The average effect across this time span therefore represents a longer-term impact averaged out over almost two decades. Estimates presented are converted directly into the percentage change caused by the interventions.

This model estimation is complemented with the bootstrapped short-run effects which show how pollution levels changed in the months directly after the opening of a new metro station or the timing of the LEZ. The bootstrapped estimates draw from the pollution in neighbourhoods directly surrounding the interventions to compare the pre and post-treatment

effect. It is comparable to the difference-in-difference estimator, calculating the average treatment effect for the subsample of values occurring sequentially in every month before and after the respective treatments up to one year.

Both long-run difference-in-difference estimates and short-run bootstrap estimates for the impact of metros and the LEZ are obtained for every pollutant across the five key areas of the city. The spatial treatment effect estimated varies in distance relative to each of these locations ranging from 100 meters to 3 kilometres away. In order to present all results in a systematic manner, the estimated coefficients from each model are all plotted together to highlight the decaying impact of pollution across space and time.[11]

3 – 6.1. Average Short and Long-Run Impacts of Metro Accessibility on Pollution

Particulate Matter

Overall, we observe positive impacts coming from the opening of new metro stations related to the reduction of particulate pollution in the city. Looking first at the direct impact across the entire study region, the average effect of the metro expansion has a median estimated long-run impact of -2.15% across the varying specifications, corresponding to β_2 from equation Eq. 3 – 3.

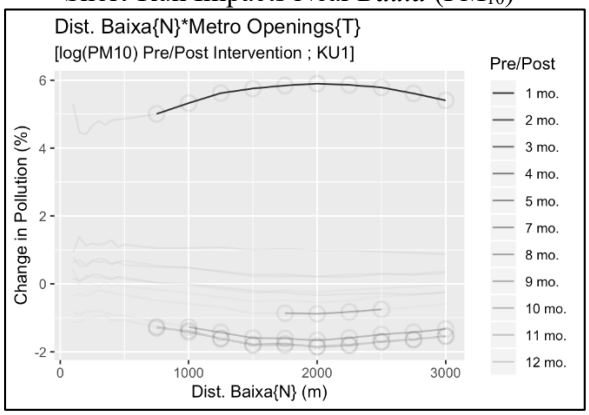
Focusing on the spatial and temporal decay of effects, estimates show a reduction in PM_{10} following the opening of new metro stations extending to some distance away from the city centre. Figure 12 shows the estimated significant impact of metro openings on sequentially increasing spatial treatments, increasing from locally within 100 meters of the city centre and opening metro stations up to 3 kilometres away. The average value of the R^2 is presented across each of the difference-in-difference models. Consistent with expectations from figure 11, the

[11] For brevity, only the most salient features and results are presented. Estimated values, and respective significance levels, are plotted sequentially in order to highlight spatial trends and robustness in considering marginal increases in the estimated values over time and space.

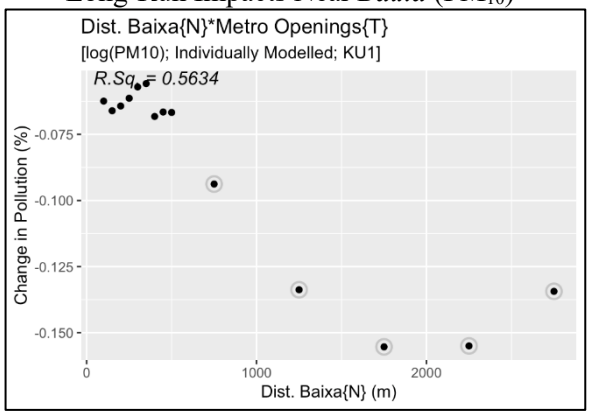
significant impacts are strongest at the point where post-treatment pollution levels appear to decay away from the city centre.

Figure 12. Short and Long-Run Metro Opening Impacts on PM_{10}

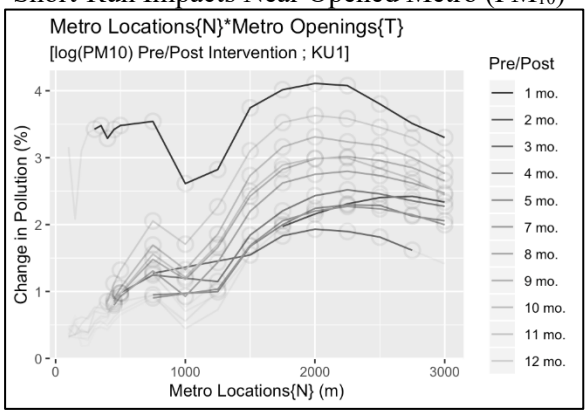
Short-Run Impacts Near *Baixa* (PM_{10})



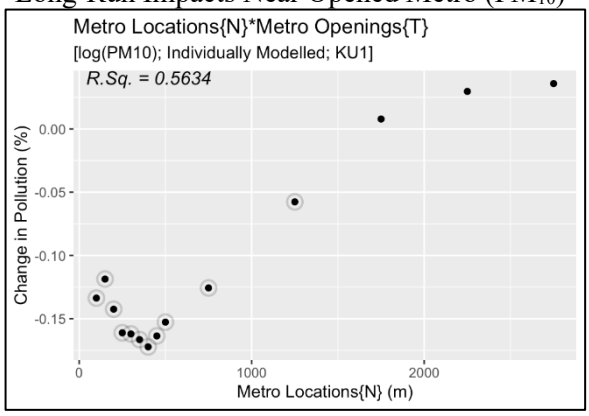
Long-Run Impacts Near *Baixa* (PM_{10})



Short-Run Impacts Near Opened Metro (PM_{10})



Long-Run Impacts Near Opened Metro (PM_{10})

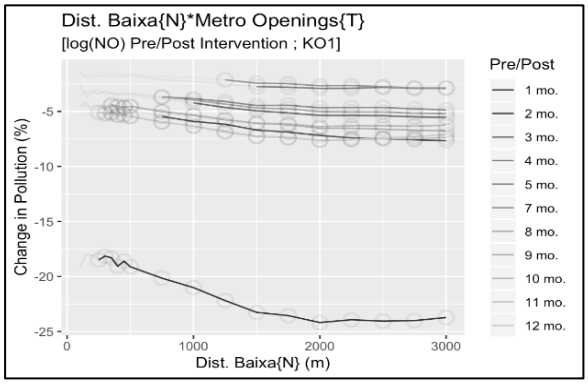


Results indicate that up to approximately 2 kilometres away from the city centre new metro stations across the city has yielded a long-term reduction in local particulate matter around 0.15%. In the short term, the bootstrap plots in the months directly before and after an opening indicate some immediate increase in PM_{10} around *Baixa*, a pattern further observed around the location of the metro station itself. With construction and preparations ongoing up to and including opening day, the immediate jump in pollution could be measuring this effect, compounded if the station opened towards the end of the month. This effect reduces in the subsequent months and immediate short-run local pollution levels reduce by almost 2% up to two and three months following the opening of a new station.

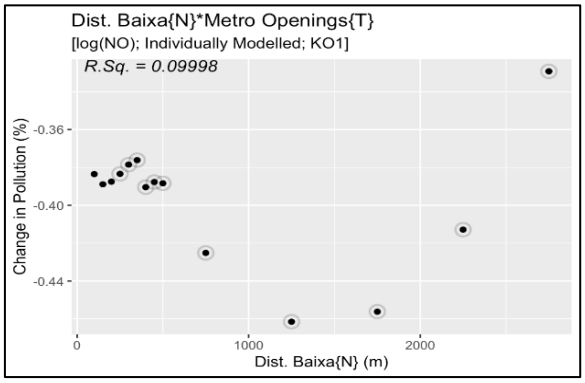
Around the metros, PM_{10} has a long term drop up to 0.175% in the immediate vicinity of the station up to 500 meters away. Weaker yet still negative reductions around 0.05% continue to be experienced in neighbourhoods up to around 1.5 kilometres away from the metro station. Although there appear to be some short-term increase in pollution in the months following an opening, this effect again dissipates over time.

Figure 13a. Metro Opening Impacts on Combustion-Based Emissions Near *Baixa*

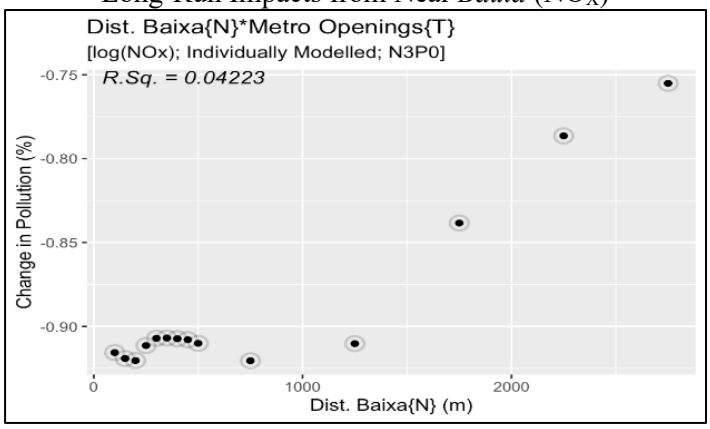
Short-Run Impacts Near *Baixa* (NO)



Long-Run Impacts Near *Baixa* (NO)



Long-Run Impacts from Near *Baixa* (NO_X)



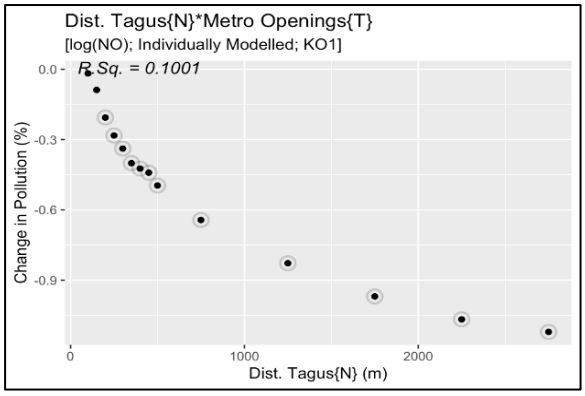
Nitrogen-Based Combustion Emissions

Metro openings decreased NO in *Baixa* by around 0.45% up to around 1.5 kilometres in the long run as seen in figure 13a. Short term effects indicate the largest decrease in the first month, around 20%, and in subsequent months this effect dies off. Up to almost a year later, however, impacts of 2.5% could be felt at large 2 kilometer distances from *Baixa*. Similarly, very localized spatial effects around *Baixa* show decreases in NO_X around 0.9%.

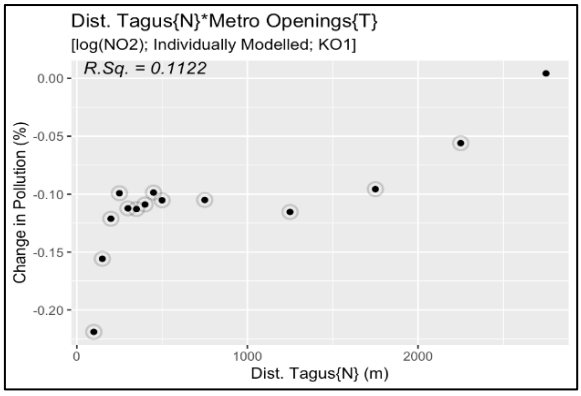
Combustion emissions have the greatest reduction around the riverfront with NO, NO_2 and NO_X all decreasing in this area following the introduction of new metro stations, as seen in figure 13b. The reduction of NO and NO_X is broad, extending away from the Tagus and reaching up to 0.9% and 0.2% up to 3 kilometres away. Around metro stations themselves, there are local reductions of NO_X around 0.3%. This impact dies off over space but decreases NO_X in neighbourhoods up to 2 kilometres away from newly opened stations.

Figure 13b. Metro Opening Impacts on Combustion-Based Emissions Near the Tagus

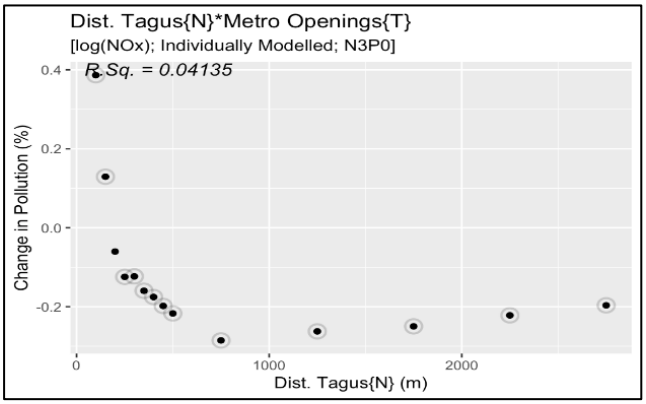
Short-Run Impacts Near the Tagus (NO)



Long-Run Impacts Near the Tagus (NO_2)



Long-Run Impacts from the Tagus (NO_X)



Carbon Monoxide

Although the average predictive power of the interpolated data at each time interval for CO has little improvement, it is still possible that a difference-in-difference specification can capture the temporal impacts from the observed point values. Broadly, the direct estimated impacts caused by the cumulative opening of metro stations, corresponding to β_2, has a median

average reduction of 2.18%. This captures the general temporal impact of these metro treatments on neighbourhood CO concentration.

As one of the primary sources of CO is vehicle emission, we would expect that the opening of new metro stations may reduce emissions from private drivers. Estimates indicate localized reductions near metro stations and broader impacts in *Baixa*, 0.25% up to 750 meters and 0.29% up to 3 kilometres respectively (figure A3 – 3 of the appendix).

3 – 6.2. Average Short and Long-Run Impacts of Low Emission Zones on Pollution

Particulate Matter

The before and after difference is generally larger when comparing the interventions surrounding the LEZ and the introduction of new metro stations. This however only evaluates pre and post intervention averages and the complex mechanisms of pollution abatement cannot cleanly be attributed to any specific source. The larger decrease following the LEZ is not surprising given the zones explicit aim of reducing pollution, however should be interpreted with caution given the number of transit initiatives introduced in conjunction with the LEZ, and thus could be capturing the combined effect of this temporal dynamic.

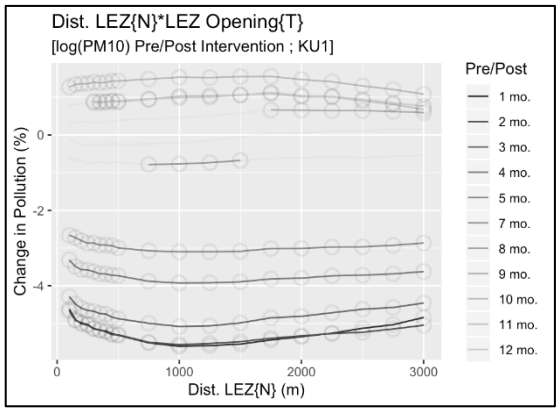
The difference-in-difference estimate can identify the local change in the area directly around the LEZ itself where pollution reduction is more targeted. A long-term reduction up to 0.30% is estimated in the immediate area around the LEZ, with broader impacts extending up to 3 kilometres away from the boundary with a broad area reduction up to 0.43%. Similar effects are confirmed when considering the distance to *Baixa,* where the LEZ is located, with broad effects up to 0.25% felt up to 3 kilometres. The effect of the LEZ has thus had a wide impact on pollution in the city centre over the long term since its introduction.

Large local decreases of PM_{10} are observed in the months directly following the intervention as expected. In the first 5 months after the introduction, the reduction ranged

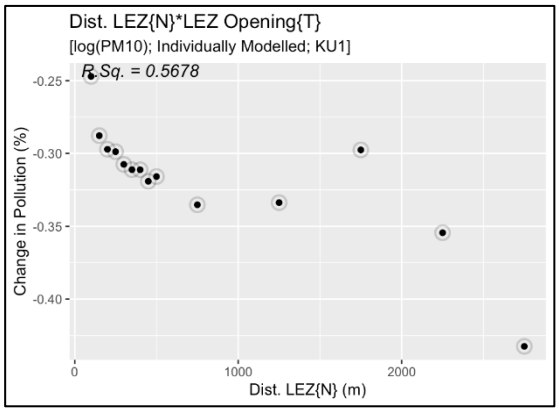
between around 2% to 4%. This effect appears to die off over time and revert, and also observed in the literature as drivers or enforcers may become complacent and regulations laxed.

Figure 14. Short and Long-Run LEZ Impacts on PM_{10} and NO_X

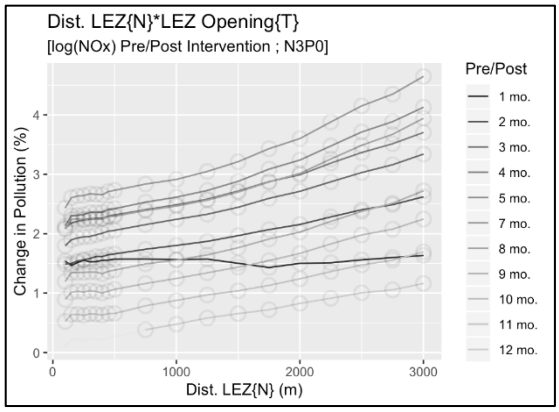
Short-Run Impacts Near LEZ (PM_{10})



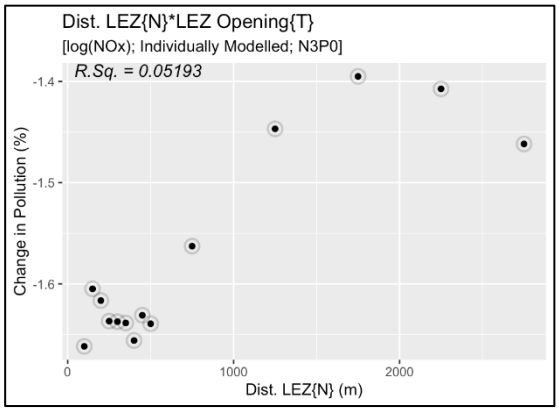
Long-Run Impacts Near LEZ (PM_{10})



Short-Run Impacts Near LEZ (NO_X)



Long-Run Impacts Near LEZ (NO_X)



Interestingly, we see some increases in pollution along the Tagus river over the long run, however small. The estimated impact represents an increase in pollution around 0.2% near the waterfront. While the zone itself has improved in air quality, there is some suggestion that the introduction of the LEZ may have led to shifts in driving patterns where now commuters or heavy vehicles entre into the city centre via alternative routes – many of which are along the river.

Nitrogen-Based Combustion Emissions

In terms of NO_X there are long-run reductions in pollution due to the introduction of the LEZ very localized of around 1.65%. This reduction decays closer to zero over space indicating the strongest reductions of NO_X are closer to the LEZ boundary after its introduction. The short run dynamics reveal some increases in localized pollution. In the month directly following the introduction of the LEZ pollution increase was constant across space, however, in the subsequent months NO_X levels further away from the LEZ began increasing at relatively faster rates.

As LEZ's primarily target traffic patterns, these short-term increases could be a result of peripheral areas surrounding the LEZ having increased spill-over traffic to avoid entering the zone after it was introduced. This can further be seen with slight increases in freeway pollution following the introduction of the LEZ in the magnitude of 0.17% within 200 meters. If traffic patterns are indeed shifted, then we would expect that drivers may take alternative routes to by-pass the LEZ and enter the city centre.

Reductions of NO_2 were broader and included decreases around the Tagus riverfront of around 1.5%. The reductions are very localized and significant for NO_2, and short-term impacts indicate strong decreases very nearby in the months immediately following the introduction which decay both over space and time. Thus, while the LEZ was efficient in reducing the concentration of NO_2 it is a very localized effect.

Similar effects are seen with NO, however, are much more extensive. The pollution reduction of around 3.5% caused by the LEZ and stemming from the riverfront was experienced by neighbourhoods up to 3 kilometres away. Short-run impacts mirror this trend with strong decreases in NO concentration which remain large as we consider a larger zone around the Tagus.

Sulfur Dioxide

The impacts on SO_2 from models looking at metro openings have a poor fit and results that are not meaningful, however estimates are much more robust and intuitive when considering the impacts from the LEZ. This shows the importance of considering the full resulting outcomes of each treatment as they related to different urban transports and ultimately different pollutants. While metro policies are targeted towards light vehicle drivers, incentivizing them to use public transport, LEZ's are targeted towards more heavy polluting vehicles. So, while a metro treatment is not likely to impact SO_2 emissions, given the low proportion coming from light vehicles, a LEZ policy targeting these heavy vehicles directly should.

Similar localized reductions in SO2 are seen around *Baixa* and the LEZ itself from around 5.8% in the immediate proximity with impacts extending out to 2 kilometres from *Baixa* and 1 kilometre from the LEZ. Along the Tagus riverfront, the introduction of the LEZ yielded reductions of over 5% in neighbourhoods within up to 3 kilometres. As we would expect transit patterns to shift, further reductions along the freeways within the city are estimated to reach around 1.5% up to 750 meters away. Plots for auxiliary impacts from sulfur dioxide available in figure A3 – 4 of the appendix.

3 – 7. Conclusions

Although this work does not aim to evaluate the full costs and benefits associated to metro expansions and LEZ's, it does provide context to the relative impact of each initiative in terms of pollution changes and spatial-temporal patterns. The results indicate that within-urban spatial and temporal decay patterns are significant and do not impact all neighbourhoods equally. Transit interventions therefore have potentially very localized effects which may cause varying impacts conditional on how commuting patterns respond. As pollution concentrations

tend to be spatially concentrated in urban areas, it is important to take these high-resolution dynamics into account when evaluating different transit interventions.

Understanding the local and neighbourhood level impacts of pollution can be used to better target hotspots where residents are at greater risks of negative impacts. Further, non-transit related pollution mitigation strategies, such as the planting of trees or developing green infrastructures (walls, roofs, parks), can be better targeted knowing where pollution is most prevalent. While long run pollution reduction around newly opened metro stations are observed, the largest effect appears to come generally from the reduction in pollution in the city centre or near the riverfront. This is similarly seen with the resulting impacts from the introduction of the LEZ, located in the centre.

Of particular interest from these results, in comparison to general average city-level impacts, there is some evidence to support the idea that the introduction of the LEZ yields alterations to transit patterns and ultimately pollution. Results suggest a post intervention increase in pollution at the external boundaries of the LEZ and along freeways. Further, results highlight that certain interventions are only appropriate for certain goals, for example, the goal of reducing SO_2 emissions should be concentrated on LEZ enforcement of heavy-duty vehicles rather than incentivizing light vehicle drivers to switch to public transit.

The results from the applied study highlight the benefit to be gained in terms of increased spatial and temporal complexity and understanding of the impacts. While in the relative context Lisbon is not a heavily polluted city, there are still clear environmental improvements to be made and best practices for other municipalities with similar infrastructure and conditions. The broader impact of metro stations and LEZ fall in line with estimated impacts from other large cities, particularly when looking at the larger short run effects estimated compared with more traditional point differences in pollution levels at different stations inside and outside of LEZs.

More important considerations should further be given to the relative costs of each transit intervention. LEZ's are very low-cost relative to the expansion of a new metro, which include capital heavy construction projects rather than simple enforcement. From a purely pollution reduction point of view therefore, the results from the LEZ suggest that interventions surrounding this goal had relatively large long run reductions in local pollutants.

However, this does not negate or diminish the benefit of metro expansions on pollution reduction. Since the primary goal of metro expansions is not the reduction of pollution, any observed decrease in pollutants constitute a positive spill-over. This value of pollution reduction is a non-market benefit that is often not taken into account when estimating the costs and benefits of any intervention. Therefore, the value of pollution reduction via metro expansions is crucial for evaluating the true benefits, based on current dynamics.

This chapter looks at the role of geostatistical methods to generate longitudinal data for detailed urban environmental statistical analysis. The application of these methods highlight the increasing detail with which urban analytics can be used to better understand dynamic processes at a highly refined spatial and temporal detail, often lacking in studies of the urban environment. This highlights the advantage of leveraging currently available geospatial data to estimate and value urban processes and impacts. From a practical point of view these estimates are crucial for better understanding urban dynamics in different contexts and better valuing location spillovers.

These procedures and estimates highlight the quality of open source data for the generation of high-dimensional longitudinal neighbourhood level databases, enabling the study of spatial and temporal dynamics of urban pollution. As data and computational power allows for higher resolution data at a higher frequency, new and relevant urban-scale intervention analyses can be conducted to better guide any discussions and best practices related to transit or other policy areas influencing the urban environment.

3 – References

Adler, Martin W., and Jos N. van Ommeren. 2016. "Does public transit reduce car travel externalities? Quasi-natural experiments' evidence from transit strikes." *Journal of Urban Economics* 92: 106-119.

Anderson, Michael L. 2014. "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion." *American Economic Review* 104 (9): 2763-2796.

Anselin, Luc, and Nancy Lozano-Gracia. 2008. "Errors in variables and spatial effects in hedonic house price models of ambient air quality." *Empirical Economics* 34 (1): 5-34.

Anselin, Luc, and Julie Le Gallo. 2006. "Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial Aspects." *Spatial Economic Analysis* 1 (1): 31-52.

Baum-Snow, Nathaniel, and Matthew E. Kahn. 2000. "The effects of new public projects to expand urban rail transit." *Journal of Public Economics* 77: 241-263,

Baum-Snow, Nathaniel, and Matthew E. Kahn. 2005. "Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970-2000." *Brookings-Wharton Papers on Urban Affairs* 147-206.

Bertazzon, Stefania, Markey Johnson, Kristin Eccles, and Gilaad G. Kaplan. 2015. "Accounting for spatial effects in land use regression for urban air pollution modelling." *Spatial and Spatio-temporal Epidemiology* 14-55: 9-21.

Boogaard. Hanna, Nicole A.H. Janssen, Paul H. Fischer, Gerard P.A. Kos, Ernie P. Weijers,Flemming R. Cassee, Saskia C. van der Zee, Jeroen J. de Hartog, Kees Meliefste, Meng Wang, Bert Brunekreef, and Gerard Hoek. 2012. "Impact of low emission zones and local traffic policies on ambient airpollution concentrations." *Science of the Total Environment* 435-436: 132-140.

Bowes, David R., and Keith R. Ihlanfeldt. 2001. "Identifying the Impacts of Rail Transit Stations on Residential Property Values." *Journal of Urban Economics* 50: 1-25.

Câmara Municipal de Lisboa (CML). 2018a. "Lisbon: The Economy in Figures."

Câmara Municipal de Lisboa (CML). 2018b. *European Green Capital 2020 – Application*.

Cervero, Robert, and Change Deok Kang. 2011. "Bus rapid transit impacts on land uses and land values in Seoul, Korea." *Transport Policy* 18 (1): 102-116.

Chen, Yihsu, and Alexander Whalley. 2012. "Green Infrastructure: The Effects of Urban Rail Transit on Air Quality." *American Economic Journal: Economic Policy* 4 (1): 58-97.

Ellison, Richard B., Stephen P. Greaves, and David A. Hensher. 2013. "Five years of London's low emission zone: Effects on vehicle fleet composition and air quality." *Transportation Research Part D: Transport and Environment* 23: 25-33.

Ferreira, Francisco, Pedro Gomes, Ana Cristina Carvalho, and Hugo Tente. 2012. "Evaluation of the Implementation of a Low Emission Zone in Lisbon." *Journal of Environmental Protection* 3 (9): 1188-1205.

Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A. Turner. 2018. "Subways and Urban Air Pollution." *National Bureau of Economic Research* Working Paper 24183.

Goetzke, Frank. 2008. "Network Effects in Public Transit Use: Evidence from a Spatially Autoregressive Mode Choice Model for New York." *Urban Studies* 45 (2): 407-417.

Gonçalves, Rui N. 2014. The Lisbon Low Emission Zone Enforcement Methods and Procedures. Portuguese Environment Agency. Presentation to then Workshop on Low Emission Zones (European Comission TAIEX).

Gonzalez-Navarro, Marco, and Matthew A. Turner. 2018. "Subways and urban growth: Evidence from earth." *Journal of Urban Economics* 108: 85-106.

Grupo Metropolitano de Lisboa (ML). 2019. "Relatórios e Contas Consolidado 2018"

Holman, Claire, Roy Harrison, and Xavier Querol. 2015. "Review of the efficacy of low emission zones to improve urban air quality in European cities." *Atmospheric Environment* 111: 161-169.

Instituto Nacional de Estatistica (INE). 2011. Mode of transport used on commuting (No.), Population and housing census - 2011.

Instituto Nacional de Estatistica (INE). 2018. Veículos novos vendidos por 1000 habitantes (N.º) por Local de residência (NUTS - 2013) e Tipo de veículo; Anual - Conservatórias do Registo Automóvel.

Kawabata, Mizuki, and Qing Shen. 2007. "Commuting Inequality between Cars and Public Transit: The Case of the San Francisco Bay Area, 1990-2000." *Urban Studies* 44 (9): 1759-1780.

Li, Zheng. 2018. "The impact of metro accessibility on residential property values: An empirical analysis." *Research in Transportation Economics* 70: 52-56.

Martinez, L. Miguel, and Jose Manuel Viegas. 2009. "Effects of Transportation Accessibility on Residential Property Values: Hedonic Price Model in the Lisbon, Portugal, Metropolitan Area." *Transportation Research Record: Journal of the Transportation Research Board* No. 2115.

Meinardi, Simone, Paul Nissenson, Barbara Barletta, Donald Dadbub. 2008. "Influence of the public transportation system on the air quality of a major urban centre. A cast study: Milan, Italy." *Atmospheric Environment* 42: 7915-7923.

Mohammad, Sara I., Daniel J. Graham, Patricia C Melo, Richard Anderson. 2013. "A meta-analysis of the impact of rail projects on land and property values." *Transportation Research A* 50: 158-170.

Mulley, Corinne, Chi-Hong (Patrick) Tsai. 2016. “When and how much does new transport infrastructure add to property values? Evidence from the bus rapid transit system in Sydney, Australia.” *Transport Policy* 51: 15-23.

Mulley, Corinne, Chi-Hong (Patrick) Tsai, and Liang Ma. 2018. “Does residential property price benefit from light rail in Sydney?” *Research in Transportation Economics* 67: 3-10.

Nunes da Silva, Fernando, Renata A. Lajas Custodio, and Helena Martins. 2014. “Low Emission Zone: Lisbon’s Experience.” *Journal of Traffic and Logistic Engineering* 2 (2): 133-139.

Pereira, Boscolli Barbosa, Edimar Olegario de Campos Jr., and Euclides Antonio Pereira de Lima. 2013. “Biomonitoring air quality during and after a public transportation strike in the centre of Uberlandia, Minas, Gerais, Brazil by Tradescantia micronucleus bioassay.” *Environmental Science and Pollution Research* 21 (5): 3680-3685.

Roukouni, A, Basbas S, and Kokkalis A. 2012. “Impacts of metro station to the land use and transport system: Thessaloniki Metro case.” *Transport Research Arena – Europe 2012* 48: 1155-1163.

Sims R., R. Schaeffer, F. Creutzig, X. Cruz-Núñez, M. D’Agosto, D. Dimitriu, M. J. Figueroa Meza, L. Fulton, S. Kobayashi, O. Lah, A. McKinnon, P. Newman, M. Ouyang, J. J. Schauer, D. Sperling, and G. Tiwari, 2014: Transport. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Sanchez, Thomas W. 1999. “The Connection Between Public Transit and Employment.” *Journal of the American Planning Association* 65(3): 284-296.

Santos, Francisca M., Alvaro Gomez-Losada, and Jose C. M. Pires. 2019. “Impact of the implementation of Lisbon low emission zone on air quality.” *Journal of Hazardous Materials* 365 (5): 632-641.

Titos, G., H. Lyamani, L. Drinovec, F.J. Olmo, G. Mocnik, and L. Alados-Arboledas. 2015. “Evaluation of the impact of transportation changes on air quality.” *Atmospheric Environment* 114: 19-31.

Viard, V. Brian, and Shihe Fu. 2015. “The effect of Beijing’s driving restrictions on pollution and economic activity.” *Journal of Public Economics* 125: 98-115.

Wolff, Hendrik. 2014. “Keep Your Clunker in the Suburb: Low Emission Zones and Adoption of Green Vehicles.” 2014. *The Economic Journal* 124 (578): F481-F512.

Wolff, Hendrik, and Lisa Perry. 2010. “Trends in Clean Air Legislation in Europe: Particulate Matter and Low Emission Zones.” *Review of Environmental Economics and Policy* 4 (2): 293-308.

Xu, Yangfei, Qinghua Zhang, and Siqi Zheng. 2015. “The rising demand for subway after private driving restriction: Evidence from Beijing’s housing market.” *Regional Science and Urban Economics* 54: 28-37.

Zhang, Wei, C-Y Cynthia Lin Lawell, and Victoria I. Umanskaya. 2017. “The effects of license plate-based driving restrictions on air quality: Theory and empirical evidence.” *Journal of Environmental Economics and Management* 82: 181-220.

Zheng, Siqi, Xiaonan Zhang, Weizeng Sun, and Jianghao Wang. 2019. “The effect of a new subway line on local air quality: A case study in Changsha.” *Transportation Research Part D* 68: 26-38.

3 – Appendix

Figure A3 – 1. PM_{10} Optimal Interpolation and Aggregation

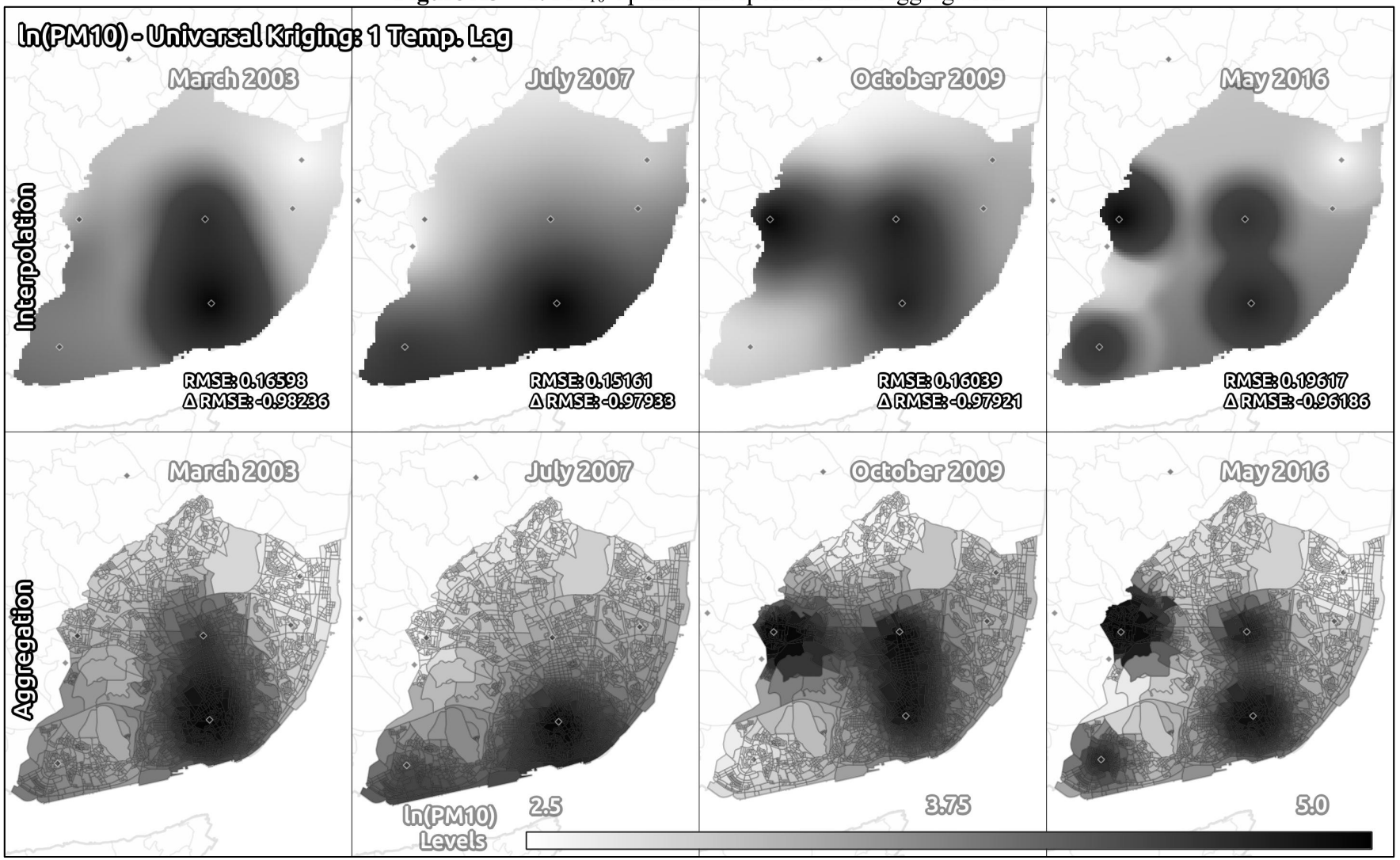


Figure A3 – 2. Spatial Decay of PM_{10} and NO Pre and Post Treatment Near Freeway

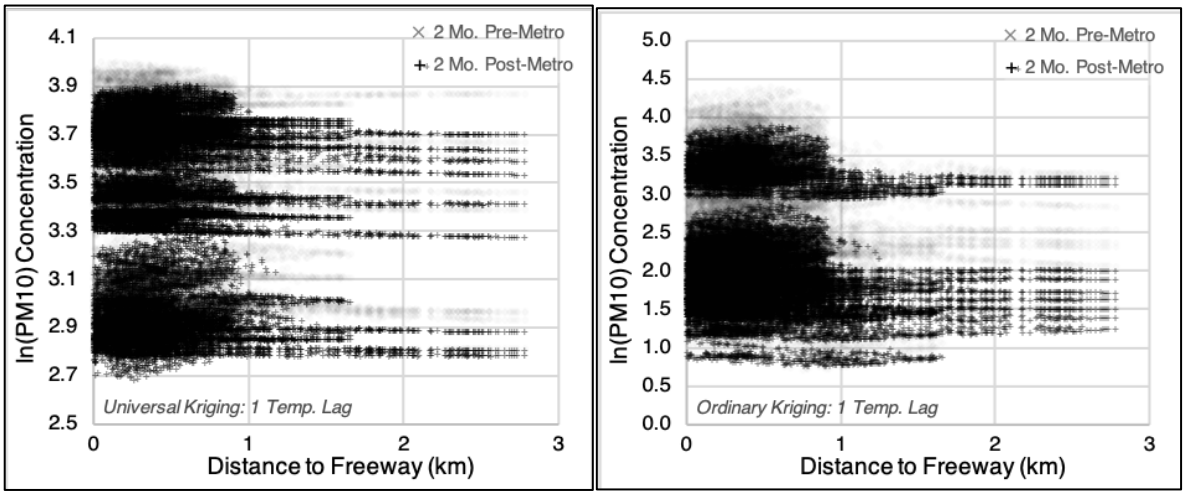


Figure A3 – 3. Carbon Monoxide and Auxiliary Plots

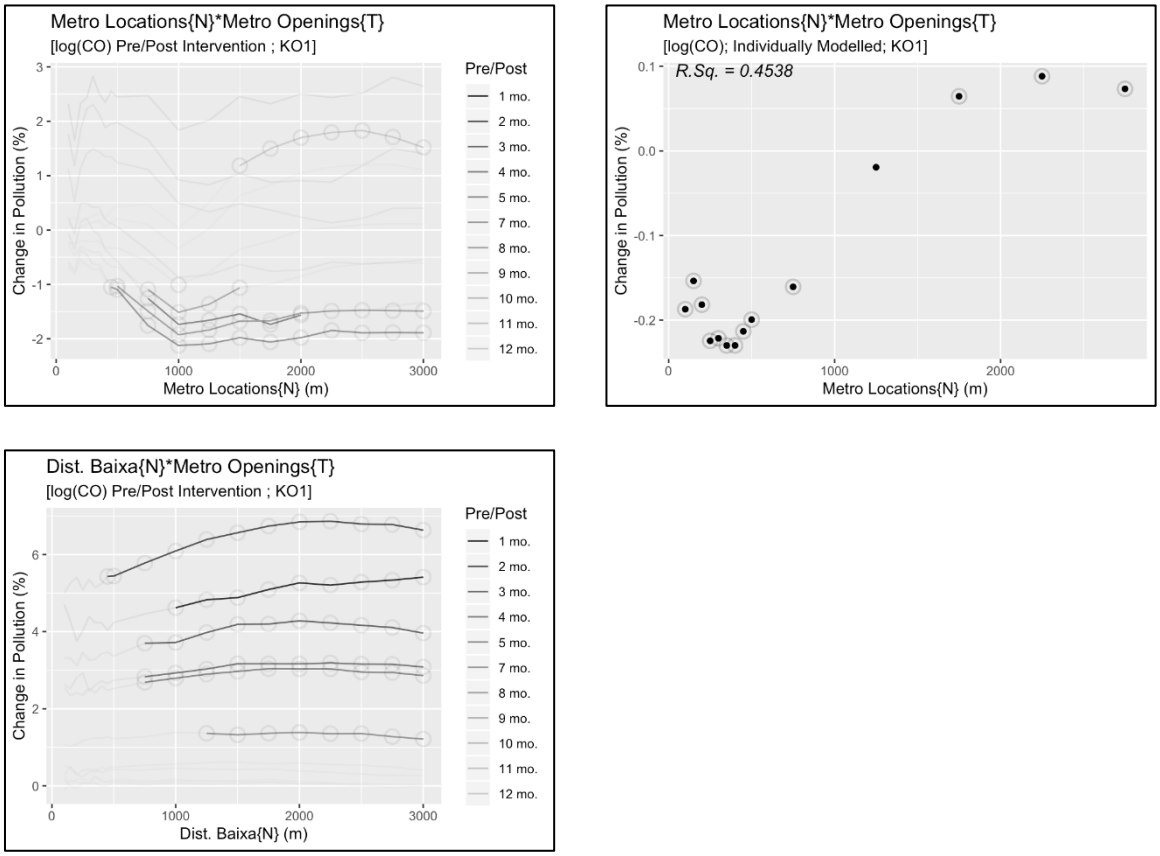


Figure A3 – 4. Auxiliary Plots of LEZ Impacts