

WHERE AND HOW MUCH?

A modelling framework to estimate land value uplifts from transport interventions

Andrés Morales

WHERE AND HOW MUCH?

A modelling framework to estimate land value
uplifts from transport interventions

José Andrés Morales

WHERE AND HOW MUCH?

A modelling framework to estimate land value
uplifts from transport interventions

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
Prof.dr.ir. A. Veldkamp,
on account of the decision of the Doctorate Board,
to be publicly defended
on December 2, 2020 at 14:45 hrs

by

José Andrés Morales

born on April 28, 1986

in Guatemala

This thesis has been approved by:

Prof.mr.dr.ir. Jaap Zevenbergen, supervisor

Dr. Johannes Flacke, co-supervisor

ITC dissertation number 387
ITC, P.O. Box 217, 7500 AE Enschede, The Netherlands

ISBN 90-978-365-5094-9
DOI 10.3990/1.9789036550949

Cover designed by José Andrés Morales
Printed by ITC Printing Department
Copyright © 2020 by José Andrés Morales
All rights reserved. No part of this publication may be reproduced without the prior written permission of the author.



Graduation committee:

Chairman/Secretary

Prof.dr. F.D. van der Meer

University of Twente

Supervisor(s)

Prof.mr.dr.ir. J.A. Zevenbergen

University of Twente

Co-supervisor(s)

Dr. J. Flacke

University of Twente

Members

Prof.dr.ir. A. Stein

University of Twente

Prof.dr.ir. M.F.A.M. van Maarseveen

University of Twente

Prof.dr.ir.ing. K.T. Geurs

University of Twente

Prof.dr. P. Wyatt

University of Reading

Prof.dr. A. van Nes

Høgskulen på Vestlandet &
TU Delft

Acknowledgements

I began this journey outside my home country, Guatemala, pursuing to expand my knowledge and perspective. Little I knew about the immense richness of experiences, beautiful minds and inspiring people that I would come across.

I would like to begin by extending my gratitude to my PhD mentors Prof. Jaap Zevenbergen and Dr Johannes Flacke for giving me the freedom and trust to venture into my research interests. Their guidance by means of insightful discussions and reflections helped me to develop myself as a researcher, to organize my thoughts in numerous occasions and to be very critical about the work I was doing. Foremost, I would like to thank them for their great support and their constant motivation, specially towards the end of this journey.

I would like to recognize and thank for the support received by many persons in ITC and particularly PGM department. A word of appreciation to Lyande Elderink and Javier Morales for inviting me to join the NICHE project which allowed me to conduct a doctoral research. Specially I would like to thank Llyande who was always supportive and interested in my progress. I would like to thank to the many inspiring minds from PGM department that were always there to share a word of knowledge but also with a very friendly touch: Richard, Dimo, Cheryl, Martin, Rohan, Divyani, Mafalda and Ana Grigolón. Special thanks to Petra, Loes and Thereza who always promptly helped me with the many student administrative diligences. A word of gratitude goes also to Carla and Marga, from the library area, for their kind support in accessing existing knowledge.

Conducting this research would have not been possible without the support of relevant organizations and helpful persons outside the ITC Faculty. I would like to thank NUFFIC as funder of the NICHE project and grantor of the PhD scholarship. Many thanks to Juan Carlos Salazar and Humberto Olavarria from "Inspecciones Globales" who opened the gates of their offices and trusted me with their data and local insights. These were essential during the conduction of my research. This gratitude is extended to their team that also received me as one of their colleagues for 6 months during my fieldwork. My appreciation also goes to Erick Uribio and Mario Catalan from ANADIE for their support by sharing relevant documentation and knowledge about the MetroRiel project. I am also thankful for the motivation, insightful comments and ideas from Eva Campos and Silvia García from URBANISTICA, and to Henry Castañeda from the Urban Mobility Bureau at Mixco City Hall. Many thanks to David Rosales from Guatemala City Hall for his

unconditional support, accompanied with invaluable friendship, during the different research stages. My appreciation to Jesús Ronquillo and Luis Castillo for their support during the research in Quetzaltenango. Special thanks to Oliver Hartleben for inviting me to collaborate with the IBI team in Vancouver and California and apply the knowledge built during this research. This experience contributed with valuable ideas and reflections to culminate my research.

During the last two years of my, quite extended, time as a PhD candidate I joined the SIMCI start-up team in The Hague. What a journey! I would like to thank each and every member of this crazy, talented and unique group people. These two years have been valuable steps in my professional experience where I did not only gain some extra technical and broad skills but mainly invaluable friendships. Special thanks to Wessel and Martijn for their support and motivation for me to finalize my PhD. Thanks to Fabio, Yolanda, Marina, Pelle, Riemer, Eelco, Baris, Bart, Mathijs and Gijs with whom I have shared many great times. Thank you for the immeasurable energy, creativity, visionary environment, the "yes we can!" attitude, the unbreakable team spirit and of course your motivation to finish this PhD.

I would definitively run short in words expressing how much I thank to my extended family. Definitively I would run out of space naming all and everyone who made my life so joyful and filled my memory and hearth with truly indelible moments during my life in Enschede. Special thanks go to Parya, Abhishek, Tatjana, Gustavo, Irene, Ana Duarte, Andre, Carissa, Homero, Kostas, Rodrigo, Emma, Vero, Flavia, Arun, Eduardo, Sumit and Aravind for all those family moments and for all the personal growth throughout the happy and sometimes not so happy moments. A special mention to my dear friends in Guatemala with whom, although the distance, the friendship just ages stronger: Josue, Raquel, Alfredo, Diego Castillo, Marielos, Hans, Luis Adolfo, Gaby, Beverly, Diego Thomae, Andreita and Mynor.

Finally, but most important, I would like to thank my family. My loving mother Miriam without whom I would have not come this far and whose example, advice, love and support has been essential my whole life. My grandmother Marta and grandfather Arturo, whose example of hard worker gentleman, humble and loving person are rooted in my hearth. My dear cousin Patty, Manu and my beloved nephew Haritz for all their love and support. My beloved fiancée and partner in life Anita. She has come to shed immense brightness, love and sense to my life. Her support was paramount during the "last push" to complete this journey and she is the cornerstone for the beginning of the next one.

Table of Contents

Acknowledgements.....	i
List of figures.....	vi
List of tables.....	viii
Chapter 1 Introduction.....	1
1.1 Background.....	2
1.2 Transport investments and their effects on land values	4
1.3 Research gap	8
1.4 Research objectives and research questions.....	9
1.5 Conceptual framework.....	10
1.6 Methodology.....	13
1.6.1 Research approach.....	13
1.6.2 Case study area: Guatemala City	13
1.7 Thesis outline	14
Chapter 2 Mapping urban accessibility in data scarce contexts using Space Syntax and location-based methods	17
Abstract	18
2.1 Introduction	19
2.2 Location-based methods and Space Syntax	21
2.2.1 Location-based methods.....	21
2.2.2 Space Syntax (SSx)	22
2.3 Methodology.....	24
2.3.1 Case study areas	24
2.3.2 Data collection, VGI data and pre-processing.....	25
2.3.3 Implementing accessibility analyses	27
2.4 Results and discussion	29
2.4.1 Geographic accessibility	29
2.4.2 Geometric accessibility	34
2.4.3 Associations between the geographical and the geometrical approach	36
2.5 Conclusions.....	41
Chapter 3 Modelling residential land values using geographic and geometric accessibility in Guatemala City	43
Abstract	44
3.1 Introduction	45
3.2 Materials and methods.....	48
3.2.1 Case study.....	48
3.2.2 Data pre-processing, variables and descriptive statistics.....	48
3.2.3 Methods	54
3.3 Results and discussion	57
3.3.1 Accessibility and land-values	57
3.3.2 Model performance and diagnostics	63
3.4 Conclusions.....	65

Chapter 4 Predictive land value modelling using Space Syntax and a geostatistical approach	69
Abstract	70
4.1 Introduction	71
4.2 Study area, data set and methods	73
4.2.1 Guatemala City.....	73
4.2.2 Data set	73
4.2.3 Predictive modelling	77
4.2.3.1 Multivariate regression (MR).....	77
4.2.3.2 Non-spatial variable selection	78
4.2.3.3 Regression-kriging (RK) as an extension of Multivariate regression	78
4.2.3.4 RK for spatialized variable selection.....	80
4.2.3.5 Model assessment and cross validation	81
4.3 Results and discussion	81
4.3.1 The MR and the MR_K	81
4.3.2 RK for spatialized variable selection.....	83
4.3.3 Land value map.....	86
4.4 Conclusions.....	89
Chapter 5 Where and how much? Predicting the impacts of a Light Rail Transit system intervention on the residential land values of Guatemala City	93
Abstract	94
5.1 Introduction	95
5.2 Case study and modelling framework.....	98
5.2.1 Guatemala City and its LRT project	98
5.2.2 Modelling framework	99
5.2.2.1 Data requirements	101
5.2.3 Modelling MetroRiel	103
5.2.3.1 Analysing new accessibility distributions.....	103
5.2.3.2 Using an MRK model to construct an intervened land value map	104
5.2.3.3 Impact analyses	104
5.3 Results	105
5.3.1 Modelled effects on accessibility improvements	105
5.3.2 Residential land value uplifts after MetroRiel	108
5.3.3 Comparison of the total land stock value uplift and the required investment for MetroRiel	112
5.4 Discussion.....	113
5.5 Conclusions.....	115
Chapter 6 Summary of main findings and reflections	117
6.1 Introduction	118
6.2 Summary of main findings.....	118

6.2.1	Chapter 2 - objective 1: To compare location-based methods and Space Syntax for mapping urban accessibility in two cities in Guatemala.	118
6.2.2	Chapter 3 - objective 2: To bridge concepts and definitions to comprehensively address accessibility to uncover its relations with residential land-values in Guatemala City.....	120
6.2.3	Chapter 4 - objective 3: To construct a land value map by means of a geostatistical approach using Space Syntax and a spatialized variable selection.	122
6.2.4	Chapter 5 - objective 4: To propose and operationalize a modelling framework to estimate the residential land value uplifts if introducing a Light Rail Transit system in Guatemala City.....	125
6.2.5	Overarching research goal	127
6.3	Reflections	128
6.3.1	Contributions to scientific research	128
6.3.2	Contributions to the case study area	130
6.3.3	Contributions to planning practice and policy	131
6.3.4	Prospects for future research.....	133
	Bibliography	137
	Summary	155
	Samenvatting	159
	About the author	165

List of figures

Figure 1.1: Conceptual Framework	12
Figure 1.2: Research approach and methods.....	13
Figure 2.1: Case study areas showing: administrative boundaries, road and public transport network	25
Figure 2.2: Methodological framework.....	28
Figure 2.3: Accessibility per variable per mode of transport in Guatemala City	30
Figure 2.4: Accessibility per variable per mode of transport in Quetzaltenango.....	31
Figure 2.5: Integrated macro and micro-location accessibility for Guatemala City (same legend as in figure 2.3 and 2.4).....	33
Figure 2.6: Integrated macro and micro-location accessibility for Quetzaltenango (same legend as in figure 2.3 and 2.4).	33
Figure 2.7: SSx results for Guatemala City and Quetzaltenango.....	35
Figure 2.8: Guatemala City. Insignificant correlations ($p < 0.01$) are cross marked.....	37
Figure 2.9: Quetzaltenango. Insignificant correlations ($p < 0.01$) are cross marked	38
Figure 3.1: Appraisals location (top), and land-value frequency distributions before and after transformation (bottom).	50
Figure 3.2: Scatterplots of x and y co-ordinates against the nl_land -values (left side); and scatterplots of estimated f_x and f_y plotted against nl_land -values (right side).....	53
Figure 3.3: Geometric via geographic-accessibility from low (red) to high (green).	55
Figure 3.4: Pearson correlations between accessibility metrics and log of land value.	58
Figure 3.5: Pearson correlations between geometric accessibility and log of land value.	58
Figure 3.6: Diagnostics of model residuals.	65
Figure 4.1: Spatial distribution of training observations and the hexagonal tessellation, adapted from (Morales et al., 2019b).	74
Figure 4.2: MLE fitted exponential semivariogram functions over the model residuals: MR_K on the left side and RK on the right side. ...	82
Figure 4.3: Graphic summary of the RK-based variable selection process.....	85
Figure 4.4: Mapping of the residuals (left), green and red for positive and negative residuals correspondingly. Frequency distribution of residuals (top-right) and predictions (bottom-right).	86
Figure 4.5: Constructed land value map, road structure is overlaid for reference.	87
Figure 4.6: Prediction error variance.	89

Figure 5.1: Guatemala City, administrative division, TransMetro, MetroRiel and overlay of modelled land values for the year 2014. Adapted from Morales et al. (2020).	99
Figure 5.2: Modelling framework, pointed frame outlines methodology from previous literature.	100
Figure 5.3: Proportional accessibility improvement for selected variables.	106
Figure 5.4: Proportional geometric accessibility improvements at selected radii. Road interventions in red.	107
Figure 5.5: Land value map after modelling MetroRiel intervention	108
Figure 5.6: Index of residential land value increments.	110
Figure 5.7: Comparison of average proportional land value uplift per station and average total against Euclidian distance ranges from MetroRiel stations.	111
Figure 5.8: Comparison of total land stock value uplift before and after intervention.	112

List of tables

Table 2.1: Data collection and pre-processing	27
Table 3.1: List of variables and descriptive statistics.	51
Table 3.2: Regression coefficients and normalized coefficients. Grey colour bars on the normalized coefficients indicate relative importance.	61
Table 3.3: Model performance in contrast with alternative models..	64
Table 4.1: Descriptive statistics accompanied by short descriptions	76
Table 4.2: Reports of the models' coefficients and assessment statistics	83
Table 5.1: Summary of data requirements.....	102

Chapter 1

Introduction

1.1 Background

Cities are the core arenas of social, economic and cultural exchanges and technological advances (Jenkins, 2007; Milder, 2012; Pacione, 2005). Together, cities account for between 70% and 90% of the world's GDP (Birch & Wachter, 2011; Seitzinger et al., 2012). Given the steadily increasing pressure of urbanization, especially in the Global South, cities are key arenas in which critical challenges in human development should be addressed over the twenty-first century (Birch et al., 2011). There is a need to address a myriad of challenges (e.g. congestion, house affordability, environmental deprivation) posed by rapid urban expansion by planning for sustainable development (Drakakis-Smit, 2000; Steinebach, 2009). Improvements to planning and management systems, decision-making processes, the use of data and to the use of multidisciplinary knowledge are some of the preconditions to achieving sustainable processes to drive urban development (Hall & Pfeiffer, 2000). Policies such as Smart Growth and the vision of compact urban forms are commonly referred to as sustainable planning strategies (Colonna, Berloco & Circella, 2012; Goodchild, 1994; Jenks & Jones, 2010; Milder, 2012). This commonly translates into a combination of practices, such as promoting mixed-used densities, that aim to reduce the pressure on urban horizontal expansion in parallel to sustainable mobility.

Lacking a global consensus on a definition of sustainable mobility, the concept commonly refers to the following three premises reflecting three broad interests, namely social, economic and environmental sustainability (Banister, 2007; Kennedy, Miller, Shalaby, Maclean & Coleman, 2005; Litman & Burwell, 2006; Meyer, 2000): (1) the accessibility and development needs of society are met safely, while promoting equity among current and future generations; (2) leveraging local and regional economic development under conditions of efficient, fair and affordable operation; and (3) minimizes environmental pollution, land and energy consumption while predominantly using renewable resources. The planning of sustainable mobility attempts to address the interdependencies and materialization of such premises. It relies on understanding land use and transport interactions (LUTI) as the components from which data about traffic flows emerge while gathering, producing and providing technical information to policy and decision makers (Black, 2018; Colonna et al., 2012; Kii, Moeckel & Thill, 2019; Litman, 2007b).

The broad nature of indicators used to evaluate transport plans refers to the three fundamental dimensions of sustainability (Litman, 2007a). Correspondingly, commonly used evaluation frameworks are available to assess planning projects such as Social Impact Assessment (SIA),

Cost Benefit Analysis (CBA) and Environmental Impact Assessment (EIA) (Ross, Orenstein & Botchwey, 2014). CBA is widely considered to be the most popular framework to deal with the economic aspects of large infrastructure investments (Vickerman, 2007). From an economic perspective, the value increase of properties or land due to transport interventions has not really been a priority indicator, and this dissertation does not suggest that it should be, but rather something that should be pursued as part of a comprehensive evaluation of transport plans (Vickerman, 2017).

There are two reasons that motivate extending and strengthening transport evaluation frameworks using spatial information of land values and the effects of transport investments. First, it would facilitate a richer understanding of the viability of a territory for sustainable land use transformations (e.g. promoting housing affordability in accessible locations) (Jones, Leishman, MacDonald, Orr & Watkins, 2010). Second, it could strengthen the opportunities to formulate financial mechanisms that facilitate the economic viability of transport investments (Bell, Bowman & German, 2009; D. Knowles & Ferbrache, 2016; Li & Love, 2020; Lungo & Smolka, 2005; Pettit et al., 2020). In this regard, Smolka (2012) argues that the technical difficulty of estimating value uplifts is one of the identified challenges to the adoptions of land value capture in Latin American cities. From the academic and practical perspectives, the interest in analysing the relationships between built sustainable transport modes and a territory's economic structures is growing (Krause & Bitter, 2012). However, researchers and practitioners can currently find few analytical frameworks to understand how transport plans could affect land or property values in advance of investments actually taking place.

The underlying motivation for this dissertation was thus to address this analytical need through an effective reproducible framework. In doing so, the research integrates the state-of-the-art of two domains, namely the modelling of urban accessibility and the modelling of property values. Bringing together recent technical advances from these two research strands aims to facilitate bridging the interests of and collaborations between planning practice and land administration. The development of the proposed methods and their application took place in Guatemala City, Guatemala. Yet, decisions made throughout the research were aimed to deliver a reproducible framework that can be adopted, adapted and further developed by other researchers and practitioners beyond the specifics of the case study area.

1.2 Transport investments and their effects on land values

The relations between land values and spatial planning is dual and mutual. Land is the basis for planning while at the same time planning interventions transform land physically and can alter the various dynamics in a territory. Traditionally, land use and mobility dynamics have been the main focus of planning institutions while land values have been mainly of interest to land administration authorities (Bell et al., 2009; Evans, 1987; Kii et al., 2019; Van der Molen, 2002). Yet there are various motivations to bridge the interests of these institutions by providing spatial information about the economic value of land: financing large infrastructures by means of value capture (Bell et al., 2009; Li et al., 2020; Lungo et al., 2005; Pettit et al., 2020; Smolka, 2012; Viguie & Hallegatte, 2014) and utilization of mathematical models that produce such spatial information for mass valuation applications (Pettit et al., 2020).

The planning of sustainable transport infrastructures - such as Light Rail Transit (LRT) and Bus Rapid Transit (BRT) systems - is of growing interest in many Global South countries as a sustainable strategy to meet the challenges of rapid urbanization (Alade, Edelenbos & Gianoli, 2020; Banister, 2007; Cengiz & Çelik, 2019; Ferbrache & Knowles, 2017; Gleave, 2005; Ingvardson & Nielsen, 2018; Liu & Shen, 2011). Innovations in public transport are essential for driving the economic progress of cities and it is well known that improvements in accessibility tend to increase the economic value of land, reflected in property price uplifts and higher rents (Ahlfeldt & Wendland, 2011; Banister & Thurstain-Goodwin, 2011; Filatova, Parker & van der Veen, 2009; Giuliano, Gordon, Pan & Park, 2010; Mohammad, Graham, Melo & Anderson, 2013). Affordable and safe access to transport has the potential to leverage social transformation processes towards greater equality and to incentivise modal shift changes, specially attracting current private vehicle users to public transport (Alade et al., 2020; Ingvardson et al., 2018; Kennedy et al., 2005). Yet quality accessibility could also trigger other (not necessarily positive) effects such as gentrification and land grabbing processes since value uplifts in the form of higher rents only benefit land and property owners (Borras Jr, Franco, Gómez, Kay & Spoor, 2012; Jones & Lucas, 2012; Lin, 2002). For example, value uplifts in Guatemala City after the first implementation of a BRT system made it more expensive for the municipality itself and private developers to invest in affordable residential projects (Morales, 2013).

Well-accepted neoclassical urban economic theory sheds light on how a reduction in the generalized costs of transport tend to be capitalized

in the economic value of locations relative to a city (Ahlfeldt, 2007; Alonso, 1964; Evans, 1987; Fujita & Krugman, 2004; Webster, 2010). A large and growing body of empirical evidence underpin that theory (Banister et al., 2011; Du & Mulley, 2006; Mohammad et al., 2013; Yan, Delmelle & Duncan, 2012). Land value uplifts, defined as the value increases resulting from accessibility improvements (Higgins & Kanaroglou, 2018; Yen, Mulley, Shearer & Burke, 2018), are widely associated with triggering further effects such as economic growth, inward investment and land use transformations (Ferbrache et al., 2017).

This dissertation focuses on how the economic value of land is uplifted as an important - yet commonly neglected aspect - in the planning of sustainable transport infrastructures. The term value is defined as the estimated price reflecting expectations and perceptions of economic worth derived from the utility of land for a specific use at a given location (Adams, 1994). It is logical to argue that transport interventions must be evaluated in the light of their future effects on land values. However, Banister et al. (2011) explain that this has not usually been done since such effects are rarely accounted for in CBA frameworks. Economic growth, new inward investment and land value uplift fall in the category of "wider", "indirect" or "non-transport" effects in CBA terminology. CBA has traditionally relied on inputs that are readily accounted for and that are of direct benefit to the user, i.e. travel time savings (Vickerman, 2007, 2017). It is argued that one of the major limitations of the CBA framework is its inability to account for spatially distributed effects (Oliveira & Pinho, 2010; Walker, Fay & Mitchell, 2005). Furthermore, under an assumption of individual utility maximization in a state of perfect market equilibrium, direct benefits would be proportionally equivalent to the wider effects such as increases in land value, rent uplifts and economic growth. Hence, including land values uplift as a benefit would represent double counting (Vickerman, 2017).

It is not surprising that there is growing interest in evaluating transport projects in the light of their future effects on the spatial distribution of land values (Banister et al., 2011; Grimes & Liang, 2010; Li et al., 2020; Lin, 2002; Metz, 2017; Rietveld & van Wee, 2008; Vickerman, 2017). Practitioners are increasingly motivated to include these effects as indicators when conducting CBAs (Kennedy et al., 2005; Litman, 2007a; Vickerman, 2007, 2017). The assumption of perfect market equilibrium rarely holds in reality, meaning that direct effects (i.e. time savings) cannot be directly translated into wider economic benefits. It is difficult to capture the added benefit of transport interventions specially when those are extensions to already existing and often

mature transport networks, hence making it difficult to justify the usually heavy amounts of required **investment**. However, a direct translation of the monetized accessibility improvements could shed some light on the objective formulation of mechanisms to finance transport investments (e.g. taxation adjustments, betterment levies, private investment). Furthermore, the ability to systematically produce spatial information about the potential of such monetary effects could better inform transport infrastructure design processes (e.g. optimization of access stations, transport corridor layouts, road connections) and choices between design alternatives. This is suggested in the context of moving towards a comprehensive and balanced selection of indicators that also consider other economic, social and environmental aspects (Litman, 2007a; Vickerman, 2017).

However, the task of predicting land value uplifts resulting from transport investments is neither trivial nor easy. Data availability plays an important role in any type of analysis dealing with accessibility and spatial distribution of urban markets and its use in CBAs reports (Banister et al., 2011; Pettit et al., 2020; Viguie et al., 2014). Analytical limitations due to data scarcity is common in many cases, particularly in Global South countries (Ahlström, Pilesjö & Lindberg, 2011; Viguie et al., 2014; Yeh & Gar-On, 1991; Yeh, 1999). Furthermore, the interactions between transport investments and property markets are highly complex. Value uplifts can be empirically observed in the various stages of a transport investment, namely after project announcement and before construction (ex-ante intervention), during and after construction (Yen et al., 2018). Effects can be heterogeneous along a transport corridor based on the variability in how users value transit-oriented investments as well as their transport modality preferences (Higgins et al., 2018). Sharma and Newman (2018) analysed the value uplift in the emerging city of Bangalore using panel data estimations that explain 74% of the observed data variability (data used to calibrate their model). They identified an uplift of up to 25% within a 500 m catchment area and 4.5% for the entire remaining city. The authors argue for the potential for putting in place value capture mechanisms. Diao, Leonard and Sing (2017) tested various models, including the spatial difference in differences (SDID) approach, explaining up to 90% of the observed data and identified an uplift of up to 7.8% post-intervention in Singapore. Devaux, Dubé and Apparicio (2017) implemented the same methodological approach, explaining up to 94% of the observed data and identified an uplift of 25% in property prices within a catchment area of 400 m of stations along a metro extension in Laval, Canada. Cervero and Duncan (2002) implemented various linear and non-linear regression techniques, explaining up to 60% of the observed data and reported a value uplift

between 23% for a typical commercial land parcel and up to a 120% increase on parcels located close to the central business district (CBD) and within 1 km of a transit station in San Diego, USA.

Overall, the research strand dedicated to analysing value uplifts after transport interventions are made (ex-post) is both fertile and international in nature. Ingvardson et al. (2018), Debrezion, Pels and Rietveld (2007), D. Knowles et al. (2016) and Mohammad et al. (2013) provided more extensive reviews on published empirical research about the wide range of proportional value uplifts that can be associated with transport interventions, particularly LRT systems. Ingvardson et al. (2018) extended their review with a comparison of reported modal shift from car ridership to BRT and LRT systems in cities across the US and Europe. Results are hardly comparable due to the variability of methods, modelling strategies, data used and the specificities of each case study (Ingvardson et al., 2018; Martínez & Viegas, 2009). Therefore, it is impossible to establish some transferable reference on the expected magnitude and spatial distribution of value uplifts. Overall, it has been observed from the literature that effects range from -45% to 100% or more (Cervero et al., 2002; Pan, 2013).

In contrast, literature reports on the prediction of value uplifts before the interventions are made (ex-ante) are scarce. Viguie et al. (2014) approached the problem using an urban economic formulation based on household utility maximization. They implemented the formulation in a land use transport interaction model calibrated for Paris. They reported that accessibility capitalization is particularly sensitive to population expansion. Ahlfeldt (2013) proposed a non-spatial non-linear regression approach to estimate the elasticities of property price as a function of public transport access to selected labour markets in London. That model was first calibrated using property transactions and then used to predict value uplift from expected travel time reductions. Gallo (2018) implemented a non-spatial linear regression approach to estimate elasticities of average property value per ward as a function of counts of public transport stations. The model was calibrated using observations of asking prices aggregated at the district level for Naples. Cengiz et al. (2019) also used a non-spatial linear regression approach to estimate elasticities of property values as a function of Euclidian distance to existing transport stations. The model was calibrated using property values within the commonly used 0.5 km catchment area of an existing transport corridor. The model was then used to predict the uplift from a future intervention. Pettit et al. (2020) presented a planning tool designed in collaboration with estate valuers and planners in Sydney, that allows building "*what if?*"

scenarios of locations of new train stations and rapidly visualizes computed land value uplifts within a 1 km circle around proposed stations. The predictive function relies on a geographic weighted regression that explains up to 85% of the input data. It incorporates accessibility metrics in the form of distances to the CBD and train stations among others. The authors emphasized the potential of the tool for mass appraisals and supporting policy makers in the implementation of land value capture strategies.

1.3 Research gap

Limitations in the existing literature leads to a description of the research gap that is addressed in this dissertation from various angles. First, the implementation of a comprehensive accessibility concept in such modelling strategies is lacking. Geographic accessibility is defined as the opportunity at an origin to reach a destination, or vice-versa, given the impedance between the two locations (Albacete, Olaru, Paül & Biermann, 2015; Batty, 2009; Curl, Nelson & Anable, 2011; Geurs & Van Wee, 2004). Operationalizing public mobility benefits by Euclidian distance to transit stops makes it impossible to associate the value uplift with an interpretable metric of improved geographic access as it accrues to users (i.e. reduced travel times to the CBD or other facilities). Also, it is a constraint as it would not allow the estimation of the value uplift effects of accessibility improvements that are due to new transport technologies relative to the city in question (e.g. the first line of an LRT system). Introduction of new transport technologies, such as LRT systems, are rarely isolated interventions, but rather a composite of urban transformations that commonly include modifications to the existing road network. Such modifications would thus also have implications for accessibility by private transport.

Modifications to an urban layout would likely have effects on its geometric accessibility. Geometric accessibility is defined as a type of resource that is determined by the network centrality and focuses on the topological, metric and geometric properties of urban layouts in a multi-scale approach (Bafna, 2003; Batty, 2004; Hillier, Turner, Yang & Park, 2010; van Nes, 2019; Volchenkov, 2019; Webster, 2010). Such an accessibility resource, as analysed in Space Syntax (SSx), has been frequently reported to be correlated with traffic flows, land use patterns and functional hierarchy (Jiang, Claramunt & Batty, 1999; Kaplan, Burg & Omer, 2020; Karimi, 2012; Li, Zhou & Wen, 2019; Serra & Hillier, 2019). Moreover, recent literature suggests that SSx metrics add relevant spatial information that improves the understanding of the variability of land and property values (Di Pinto & Rinaldi, 2019; Enström & Netzell, 2008; Law, Penn, Karimi & Shen, 2017; Xiao, Orford & Webster, 2016a). Broadening the way in which

accessibility is understood and modelled from a geographic to a geometric perspective is thus relevant. This is particularly true if we consider the trade-off between a less data-intensive approach (i.e. only road network representation is required) and the reported co-variability of SSx metrics with property markets. Such trade-offs become especially relevant especially when addressing data scarce contexts. Yet SSx applicability should be tested in urban setups that are dissimilar than those previously investigated (e.g. more heterogeneous urban developments).

Second, markets are far from being in perfect equilibrium (Vickerman, 2017). In areas where market imperfections are aggravated or unknown, estimating the elasticities of land values as a function of accessibility would be preferable to an urban economic formulation. A model calibrated incorporating such elasticities could not only provide insights into the relations between accessibility and land values (i.e. inferential modelling) but extend its application for predictive purposes.

Third, the spatial scope of analysis cannot be restricted to pre-assumed catchment areas (i.e. the 0.5-1km buffer around stations) that is typically applied. Instead, a city scale approach would allow the calibration of elasticities based on richer datasets whilst gaining broader spatial insights on the potential land value effects of a proposed transport investment. When applying this consideration, it becomes increasingly relevant to utilize statistical approaches to address spatial dependence. This is a common problem in property value studies where patterns are observable in the spatial distribution of model residuals (Bourassa, Cantoni & Hoesli, 2010; Gallo, 2018; Krause et al., 2012).

1.4 Research objectives and research questions

This research will propose and implement a modelling framework to estimate the spatially distributed land value uplifts of proposed transport infrastructure by means of operationalizing a comprehensive accessibility definition (i.e. incorporating Space Syntax metrics) into a predictive model. To achieve this goal, we formulated the four research objectives described below.

Objective 1: To compare location-based methods and Space Syntax for mapping urban accessibility in two cities in Guatemala.

- 1 How to measure accessibility at a city and neighbourhood scales while accounting for data scarcity?

- 2 What are the relations between Space Syntax and urban access to various destinations as a first step to evaluating its applicability to explain variations in land values?

Objective 2: To bridge concepts and definitions to comprehensively operationalize accessibility indicators and uncover their relations with residential land-values in Guatemala City.

- 1 How to combine Space Syntax and location-based methods to explain the variability of land values in Guatemala City?
- 2 What are the elasticities of land values as a function of urban access in Guatemala City?

Objective 3: To construct a land value map by means of a geostatistical approach using Space Syntax and a spatialized variable selection.

- 1 How do Space Syntax-based metrics add relevant spatial information to the modelling of land values after accounting for spatial dependence?
- 2 What is the spatial distribution of residential land values in Guatemala City?

Objective 4: To propose and operationalize a modelling framework to estimate the residential land value uplifts if a Light Rail Transit system were to be built in Guatemala City.

- 1 How to model the spatially distributed effects of a Light Rail Transit system on residential land values?
- 2 Where and what is the potential residential land value uplift that could be expected following the introduction of a Light Rail Transit system?

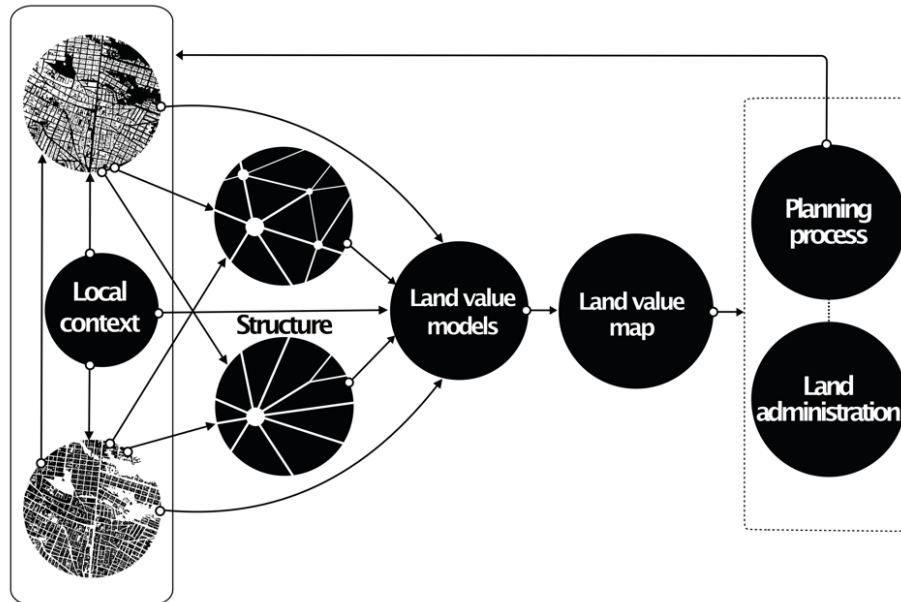
1.5 Conceptual framework

Figure 1.1 shows the conceptual framework that guided this research. That framework emerged from the position that spatial information of land value (i.e. land value maps) is becoming increasingly important in planning, particularly of transport infrastructures. There are three specific instances of the planning process, as defined in UN-HABITAT (2005) and Sharifi and Zucca (2009), in which land value maps are argued to be particularly relevant: intelligence, design and assessment (or choice). These steps are defined respectively as: (1) the understanding of a base-line situation where requirements are formulated to achieve a vision; (2) the formulation and drafting of plans/projects to satisfy requirements established in the intelligence

phase that could redefine the distribution of urban access and impact therefore land values distribution; and (3) the evaluation of proposals in the light of selected indicators using evaluation frameworks such as the CBA. While planning processes have been traditionally concerned with the interactions between transport and land uses, incorporating spatial information about land values as a relevant input in such processes could motivate the bringing together of common interests and collaborations between planners and land administration authorities. Land value maps are central to the functions of such authorities and consequently also the potential effects of transport investments on those.

Our framework suggests that in order to construct land value maps - and in a manner that the same approach can be used to analyse the potential land value uplifts- it is vital to rely on a land value predictive model. Such a model must meet certain requirements in order to help address the research gap. First, it should operationalize a comprehensive definition of urban access. It is hypothesised that a land value model would benefit from incorporating a robust operationalization of geographic access. Such access emerges from the combination of the distribution of land uses that are relevant to the phenomena (i.e. distribution of land values) and the mobility infrastructure availability to reach those. Second, the model should increase its ability to explain land value variability by the complementarity of spatial information added by geometric access metrics. It is hypothesised that distribution of geometric access (i.e. topology-based) will be comparable to the distribution of geographic access (i.e. time-based) as a first step to test its complementarity in explaining land values. Furthermore, a systematic mapping of geographic and geometric access would allow to visualization of certain city structures of centrality or poly-centrality, relevant to the understanding of spatial distribution of land values.

Geographic access



Geometric access

Figure 1.1: Conceptual Framework

Third, the model should rely on a city-wide spatial scope. This would allow the model to benefit from “learning” the relationships between access and land values from a richer dataset compared to only focusing on an assumed catchment area around transport investments. This means that it would be possible to understand the potential effects of transport investments from a broader perspective -compared to focusing only in assumed catchment areas - and in the context of existing land value structures (e.g. monocentric or poly-centric). Fourth, the model would not only provide insights about the current relationships between access and land values but it should be possible to extend its applicability for predictive purposes together with the incorporation of additional variables relating to the local neighbourhood context (e.g. social aspects, sub-markets) to tackle potential problems of spatial dependence. The third and fourth requirements pose particularly new approaches to the research strand that incorporates geometric access to model land values.

1.6 Methodology

1.6.1 Research approach

To achieve the objectives and provide answers to the questions formulated, a quantitative correlational research approach was taken. A correlational research design is non-experimental and it focuses on finding relationships between variables without explicitly addressing causality as opposed to a comparative-experimental research (Curtis, Comiskey & Dempsey, 2016; Johnson, 2001). Relationships between variables can also be described as the data structure for which various statistical techniques can be applied to numerically define such relationships. For our purpose, by taking a correlational approach the research unveils statistical relationships between a comprehensive definition of urban accessibility (i.e. independent variables) and the determination of land value (i.e. dependent variable) to then extend its applicability into two predictive analyses: the construction of a baseline land value map and the predictive analysis of the effects of future access improvements on such land values. During the process, the research adaptively integrates state-of-the-art quantitative methods from two broad research domains: urban accessibility and property/land value modelling. Figure 1.2 shows a synthesized architecture of the methods utilized in the research in the context of the correlational design and the research objectives. Thin-lined boxes indicate domain-specific quantitative methods and thick-lined boxes indicate methods that are part of the correlational research design.

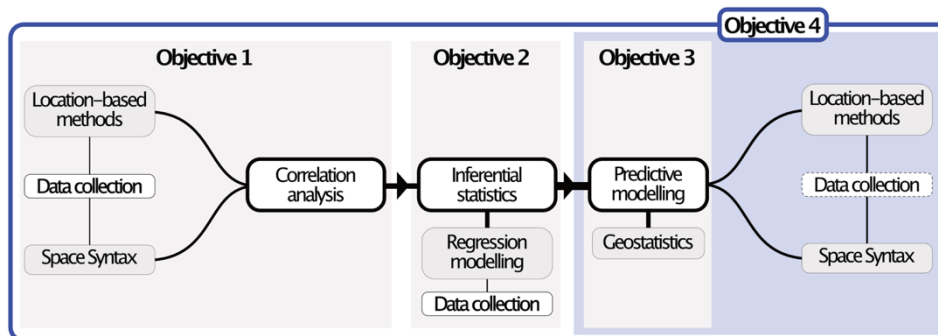


Figure 1.2: Research approach and methods.

1.6.2 Case study area: Guatemala City

The selected case study area is Guatemala City, which is located in Guatemala, Central America. As in other countries in Latin America, the country has a colonial heritage in its planning tradition (Ford, 1996; Griffin & Ford, 1980). This is reflected in historic gridiron networks and common Global South problems (Glebbeck & Koonings, 2015; Pacione,

2005, pp. 447-602) such as: (1) heterogeneous and fragmented urban development; (2) the presence of informal settlements in central areas; (3) deteriorated historic cores; (4) top-down, but weak planning practice; and (5) congestion-related problems due to the uneven and unplanned horizontal expansion and centralized economic land uses.

Guatemala City is the country's capital and is located in the country's central region. It accommodates around 26% of the country's population. It extends over 996 km² within its municipal administrative boundaries, excluding the conurbation areas in contiguous municipalities. The city has expanded from its historic core, starting with planned expansions, and then moving towards unplanned peripheral developments alongside the main infrastructure developments. The first planned expansions are associated with current location of the CBD. Horizontal expansion has been mainly shaped by topographic conditions. Currently expansion mostly occurs in the South-Eastern, South-Western and Western areas, outside the city's administrative boundary.

From the most recent census, in 2018, it is known that 70% of the city's population lives in the peripheral areas. However, the major concentration of economic activities (i.e. jobs location) remains centralized in the city core. This leads to significant needs for mobility by citizens which results in ever-increasing congestion problems. As a response, the local municipality and government have put in place efforts to improve public transport mobility. TransMetro is a BRT system that already has various lines running across the city. MetroRiel is a proposal to implement an LRT system by means of restoring the old railways that ran across the city from South-West to North-East.

The investment required to build MetroRiel is expected to be approximately US \$700 M (IDOM, 2016). Although the project has already been announced and it was expected to start its construction phase soon after the feasibility report was completed (in 2016), it experienced some delays due to bureaucratic delays and budget constraints. Given its local relevance, intervention spatial-scale and being representative of the surging interest in LRT systems internationally, the proposal is used as an application case study for the operationalization of the modelling framework presented in this dissertation.

1.7 Thesis outline

This dissertation comprises six chapters. Chapter one is this introduction. Chapters two to five describe the research findings

related to the four research objectives. Chapter six presents the synthesis of the research.

Chapter One introduces the motivation and background that lead to describing the research gap addressed in this dissertation. Research objectives are listed with their corresponding research questions. The chapter also presents the conceptual framework and the methodology utilized in the research.

Chapter Two addresses the first research objective¹. It presents a comprehensive review of concepts and methods to quantitatively analyse urban access, both geographically and geometrically. The methodologies currently available to analyse urban access from these two perspectives are presented, location-based and Space Syntax respectively. The chapter includes a discussion of the selection of Space Syntax, including the views of critics and comparable approaches. Accessibility is then analysed for Guatemala City and another city in the same country. The results underpin the discussion of the relationships between the distribution of geographic and geometric accessibility as well as the applicability of Space Syntax in the case study area.

Chapter Three addresses the second objective². This chapter proposes the formulation of one metric that analyses potential accessibility as the ease of reaching geometric access as the resource, namely geometric via geographic access. Through a set of modelling experiments, it sets the strategy for an operationalization of a comprehensive definition of urban accessibility into a land value model. The chapter attempts to deal with the problem of spatial dependence as detected in the land value data structure, by incorporating additional variables reflecting sub-market and neighbourhood characteristics. The results provide the basis for a discussion on the elasticities between land values and access metrics, as well as the confirmation of a monocentric structure in Guatemala City.

¹ Chapter two is based on: J. Morales, J. Flacke, J. Morales and J. Zevenbergen. Mapping Urban Accessibility in Data Scarce Contexts Using Space Syntax and Location-Based Methods. *Applied Spatial Analysis and Policy*, 12(2), 205-228, 2019.

² Chapter three is based on: J. Morales, J. Flacke and J. Zevenbergen. Modelling residential land values using geographic and geometric accessibility in Guatemala City. *Environment and Planning B: Urban Analytics and City Science*, 46(4), 751-776, 2019.

Chapter Four addresses the third objective³. This chapter addresses the spatial dependence problem and provides an improved modelling strategy to refine the selection of variables in the model presented in Chapter three. Light is shed on new findings about how Space Syntax metrics do add relevant modelling information to explain variability of land values under spatialized modelling conditions. Among these findings, it was revealed that the access metric formulated in Chapter three turns out to add more information to the model compared to the access to the CBD. The chapter presents a newly constructed land value map of Guatemala City for the year 2014.

Chapter Five addresses the fourth objective⁴. This chapter introduces a modelling framework to estimate land value uplifts as a function of the future accessibility improvement that can arise from transport investments. The framework comprises a structure of the data and the methodologies developed in the previous three chapters. The proposal for a Light Rail Transit system is used to empirically operationalize the framework and visualize its potential future effects on Guatemala City's land value structure.

Chapter Six presents a summary of the main findings in the context of the overarching objective, sub-objective, research questions and limitations. The chapter finishes with reflections on the main contributions and makes recommendations for future research.

³ Chapter four is based on: J. Morales, A. Stein, J. Flacke and J. Zevenbergen. Predictive land value modelling in Guatemala City using a geostatistical approach and Space Syntax. *International Journal of Applied Earth Observation and Geoinformation*, 1-33, 2020.

⁴ Chapter five is based on: J. Morales, J. Flacke and J. Zevenbergen. Where and how much? Predicting the impacts of a Light Rail Transit system intervention on the residential land value in Guatemala City. *Computers Environment and Urban Systems* (Submitted), 2020.

Chapter 2

Mapping urban accessibility in data scarce contexts using Space Syntax and location-based methods *

* This chapter is based on: J. Morales, J. Flacke, J. Morales and J. Zevenbergen. Mapping Urban Accessibility in Data Scarce Contexts Using Space Syntax and Location-Based Methods. *Applied Spatial Analysis and Policy*, 12(2), 205-228, 2019.

Abstract

Data scarcity is still a common barrier to adequately understanding urban access in Global South countries. Widely used location-based methods address the traditional definition of accessibility as the easiness to reach land-uses by means of available mobility modes. Space Syntax instead analyses accessibility as network centrality focusing only on the topological and geometric properties of urban layouts, making it comparatively less data intense. However, the interpretation of its outputs is limited to its own theory. Knowledge is missing on how such metrics are comparable to the metrics produced by location-based methods. The objective of the research was to compare both approaches for mapping urban accessibility in two cities in Guatemala. Our hypothesis tested the assumption that Space Syntax metrics could consistently reflect accessibility conditions that so far have only been measured by location-based methods. We proposed an approach using volunteered geo-information and produced accessibility maps following both approaches that were then compared using Pearson correlations. Space Syntax metrics at low and high radii are consistently correlated with location-based access to land uses that reflect location quality at neighbourhood and city-wide scale correspondingly. Space Syntax metrics at lower radii reflect time-based access restrictions either posed in the location-based analyses or by reduced accessibility by public transport. The hypothesis acceptance, $p < 0.01$, expands the scope of accessibility knowledge derivable from limited data availability using Space Syntax, which is relevant for its applicability in data-scarce contexts by planners and researchers in the Global South. Rather than replacing location-based methods Space Syntax offers an important complementary measure to geographical accessibility. This having been said, Space Syntax could contribute to early-stage planning by gaining overall insights into patterns of urban access.

2.1 Introduction

Understanding urban accessibility is fundamental for land use and transport planning (Curl et al., 2011; Curtis & Scheurer, 2010; Geurs et al., 2004), as it is one of the key aspects for agglomeration economies, economic growth, and quality of life (Ahlström et al., 2011; Kourtiti, Nijkamp & Partridge, 2015; Rietveld, 2015). Two conceptions of urban accessibility can be distinguished. *Geographic accessibility* is the most common one and is defined as the opportunity at origin to reach a destination, or vice-versa, given the impedance between both locations (Albacete et al., 2015; Batty, 2009; Curl et al., 2011; Geurs et al., 2004; Handy & Niemeier, 1997; Ingram, 1971). The combined effect of land use distribution and infrastructure components at a given location determines geographic accessibility (Geurs & van Eck, 2001). *Geometric or general accessibility*, on the other hand, is concerned with network centrality and focuses on the topological, metric and geometric properties of urban layouts (Bafna, 2003; Batty, 2004; Hillier et al., 2010; Webster, 2010).

Two methodological approaches correspond to the two concepts of access. Location-based measurements have been the preferred methods to analyse geographic accessibility (Curl et al., 2011; Geurs et al., 2001). In turn, Space Syntax (SSx) is a set of theories and methods with long-standing development whose purpose is to analyse geometric accessibility (Hillier, Leaman, Stansall & Bedford, 1976; Karimi, 2012; Webster, 2010). The availability of geographic data (e.g. land use, road and public transport networks), the easiness of interpretation and applicability of geographic information systems (GIS) have facilitated implementing location-based methods for transport planning purposes. However, the scarcity of official data and capacities for processing the same is still an important barrier in Global South countries (Ahlström et al., 2011; Yeh et al., 1991; Yeh, 1999) such as Guatemala. Common problems are incomplete or outdated data sets as resources might not be available for periodic collection and maintenance.

Alternative sources of information such as volunteered geographical information (VGI) might be potentially useful when dealing with scarcity of official data (Arsanjani, Zipf, Mooney & Helbich, 2015), jointly with considering a geometric accessibility concept. The SSx method is less data-intensive than traditional location-based methods. Only a representation of a road network is needed for the analysis. Previous work has already reported associations between SSx metrics with relevant urban phenomena: flows of people (Hajrasouliha & Yin, 2015), land use and construction density (Hillier, Greene & Desyllas, 2000; Hillier et al., 2010; Kim & Sohn, 2002; van Nes, Berghauser-

Pont & Mashhoodi, 2011) and real estate values (Matthews & Turnbull, 2007; Netzell, 2012).

SSx has also been debated regarding its dual analytical approach (Hillier & Penn, 2004; Porta, Crucitti & Latora, 2006; Ratti, 2004). Batty (2013) emphasized the problem of mathematically relating the topological-based measurements with the intuitive geographic ones (e.g. distance or time) and proposed an analytical framework to reconcile SSx with metric information. The SSx approach has attempted to prove itself a complementary tool to aid planners and researchers in accessibility studies, particularly in data-scarce contexts. However, the interpretation of its outputs remains limited to its own theory and knowledge is missing on how such metrics are comparable to the measurements produced by location-based methods. These observations restrict its applicability as an analytical approach when data availability is limited.

The objective of this chapter is to compare a geographical and a geometrical approach for mapping urban accessibility. Our hypothesis tested the assumption that Space Syntax metrics could consistently reflect urban access conditions that so far have only been measured by location-based methods. By testing this hypothesis, we attempted to contribute in empirically bridging both approaches and expanding the scope of knowledge derivable from SSx. This is relevant for planning practice as regards the applicability of available methods to address accessibility-related planning tasks in the context of Global South cities with data-challenging environments. Two cities in Guatemala were studied in order to examine the applicability of both approaches in different heterogeneous and fragmented contexts. We developed a methodological framework for analysing accessibility using SSx and location-based methods. This included a tailored based access per mode of transport to key land uses that are relevant in planning practice and are commonly associated with urban-economic dynamics. We further derived two SSx metrics at the road-level at various spatial scales. Finally, the results from both approaches were compared using Pearson correlation. The strength and significance ($p < 0.0q$) were evaluated. We elaborated on how geometric accessibility measurements provided information that was comparable to geographic access to various land-uses per mode of transport, its limitations and its applicability in practice.

The remainder of this chapter is organised as follows: section 2.2 introduces the location-based and SSx-based accessibility measurements used in this research. Section 2.3 describes the methodological framework and introduces the case study areas.

Section 2.4 presents the results and discussion. Finally, section 2.5 addresses the conclusions of this chapter.

2.2 Location-based methods and Space Syntax

2.2.1 Location-based methods

Location-based methods are widely used in research and practice (Albacete et al., 2015; Geurs et al., 2004; Handy et al., 1997; Koenig, 1980; Wegener & Fürst, 2004). They aim to analyse accessibility considering four components (Geurs et al., 2001, p. 35): (1) mobility infrastructure (i.e. roads, public space, public transport), (2) land-use location, (3) temporal conditions of the previous two, such as variability of travel-time and available land uses during the course of the day or week and (4) personal-level characteristics and restrictions. A plausible accessibility model would attempt to address these aspects as far as possible in accordance with its purpose. However, it will be limited by the availability of geographic data.

Three commonly used location-based measurements are: (1) impedance to closest facility, (2) cumulative opportunity, and (3) potential accessibility. The first analyses proximity following the criteria of shortest trip where impedance is commonly defined by travel time (per mobility mode), distance or cost. Cumulative opportunity measures the number of reachable attractions within a given impedance threshold and takes the form of equation 2.1.

$$A_i = \begin{cases} \sum M_j, & \text{if } dij \leq R \\ 0, & \text{if } dij > R \end{cases} \quad (2.1)$$

Where A is the access at origin i ; M is the size of the attraction at destination j ; d is the impedance between i and j ; and R is the radius restriction. The potential accessibility can be traced back to Stewart and Warntz (1958) and Hansen (1959). It accounts for the size of attraction (e.g. number of jobs) and the effect of distance on the interaction probability between origin and destination. Such effect is commonly named distance decay. The measurements take the form of equation 2.2.

$$A_i = \sum M_j \alpha \exp(-\beta * dij) \quad (2.2)$$

Where M is equal to the size of the attraction at j ; and α and β are constant parameters that determine the distance decay. These three measurements are simple and less data-intense compared to other location-based measurements such as those based on balancing

factors and derived from time-space geography (Curl et al., 2011; Geurs et al., 2001). The components of equations 2.1 and 2.2 can be adapted to data availability. For example, impedance can be measured in planar or network distance, time, or cost. Although planar or even network distance could be used if data is scarce, real mobility conditions are represented better when using travel time or cost per mode of transport. The size of attraction ' M ' in both equations could simply represent the number of facilities available (e.g. number of public spaces). Even though a more realistic representation could be for instance to include floor area.

Limitations of these measurements have been described by Geurs et al. (2001). A cumulative opportunity does not distinguish impedance or attraction size differences between the various destinations reached within the fixed threshold. These limitations are overcome by the potential accessibility measurement. However, decay parameters should be calibrated per mobility mode and trip purpose, which is more data demanding. Results are less intuitive to interpret, although acceptable to non-specialists. Some drawbacks of the potential accessibility measurement are: influence of self-potential, attraction within origin zone; no distinction between matching types of attraction and individual preferences; only addressing the spatial distribution of attraction supply, not the demand of these. Extensions of the basic gravity model have addressed these drawbacks at a cost of more data needs and interpretability.

2.2.2 Space Syntax (SSx)

SSx is a network analytical formalism to analyse a type of accessibility that also has an economic significance (Webster, 2010). Hillier, Penn, Hanson, Grajewski and Xu (1993) describe this access type as the easiness to move *through* and *to* places given the spatial arrangement of urban layouts, which has shown to be correlated with flows and attraction of movement. Urban economies are tightly linked to these dynamics as certain land uses benefit from these flows based on a maximum profitability rationale (Hillier, 2007a). Positive correlations between SSx with real estate values and construction density support such a relation (Kim et al., 2002; Matthews et al., 2007; Netzell, 2012). It follows that we can expect a positive correlation between accessibility as analysed in SSx and location-based accessibility to various land uses that follow an economic rationale, or service type of activities where the purpose is to be reachable.

Sharing similar grounds with SSx is the Multiple Centrality Assessment (MCA) method (Porta, Crucitti & Latora, 2005). The main difference between the two is that SSx analyses are computed using a dual graph,

while MCA is based on primal one. Opposite to the dual graph, in the primal approach intersections are treated as nodes and streets as edges. While SSx is known for pioneering in the studies of network centrality applied to cities, MCA presents itself as an enhanced method with recent evidence of its capacity to correlate with location of economic activities (Porta, Latora & Strano, 2010; Porta et al., 2012). However, in our research we consider it appropriate to implement the SSx approach as it benefits from a notably larger body of literature empirically supporting its applicability in various urban studies, in planning and design processes and with respect to the availability of applications for direct implementation within GIS. Following Law (2017), metrics analysed at the street level via a dual approach would be adequate to compare with accessibility metrics that are derived from travel times and location of places along the street, not at street-junctions.

SSx analyses over road centre-lines are done using a segment angular analysis (SAA) technique (Hillier & Iida, 2005; Turner, 2007). It is a geometric weighting method that works as an impedance parameter based on the idea that persons seek to minimize their angular deviation when choosing trip routes (Dalton, 2003). Implicitly SSA accounts for the continuity of road segments, but without incurring in an explicit generalization process (network simplification) such as the “street-name approach” or the “continuity negotiation algorithm” (Jiang & Claramunt, 2002; Porta et al., 2006).

Two main variables are analysed: *integration* and *choice*. Integration is equivalent to network closeness, and choice to network betweenness (Freeman, 1977; Porta et al., 2005). *Integration* measures how close each segment is to any other segment in the network. *Choice* measures the cumulative number of times that each segment is used in shortest trips from every segment towards every other segment. Impedance in SAA is based on angular deviation between segments, unlike the measurements of time or distance in geographic access. Thus, angular integration at any given x segment takes the form of equation (2.3)

$$Int_{\theta}(x) = (\sum_{i=1}^n D_{\theta}(x, i))^{-1} \quad (2.3)$$

where n is equal to the number of segments in the system, and $D_{\theta}(x, i)$ is the angular *depth* between x segment and any other segment in the network, i . Depth indicates the cumulative angular deviation.

Angular choice is expressed in equation (2.4), where $\sigma(i, x, j) = '1'$, once x is used to go from i to j , else = $'0'$ and being $i \neq x \neq j$. Hillier, Yang and Turner (2012) suggest a normalization procedure for integration (NAIN) and choice (NACH) to a scale ranging from -3 to 3. While

normalizing choice is highly recommended, normalising both values allows comparing the results between segments within a city, and with other cities. Yet, Hillier et al. (2012) report some inconsistencies about the use of NAIN.

$$Ch_{\theta}(x) = \frac{\sum_{i=1}^n \sum_{j=1}^n \sigma(i,x,j)}{(n-1)(n-2)/2} \quad (2.4)$$

In SSx terminology global integration and choice measurements are carried out at city-wide spatial scales. Local integration and choice values are analysed by introducing metricized restriction radii (Hillier et al., 2010). High local integration values are associated with walkable areas that have dense and consolidated networks. High local choice values are associated with streets that serve to connect the neighbourhood-level areas to higher-hierarchy roads. Various integration values at increasing radii are argued to be empirically correlated with various types of movement patterns (Hillier, 2007b, 2009; Hillier et al., 1993; Penn, 2003).

2.3 Methodology

2.3.1 Case study areas

Our case study cities are in Guatemala, Central America: Guatemala City (GC) and Quetzaltenango (QT). As in other countries in Latin America, the country has a colonial heritage in planning tradition (Ford, 1996; Griffin et al., 1980). This is reflected in historic gridiron networks and common Global South problems (Glebbeek et al., 2015; Pacione, 2005) such as: heterogeneous and fragmented urban development, deteriorated historic cores, top-down, but weak planning practice and congestion-related problems due to the uneven and unplanned horizontal expansion and centralized economic land uses. Both cities have expanded from an historic core, starting with planned expansions, and then moved towards unplanned peripheral developments following the main infrastructure (see figure 2.1). The first planned expansions are associated with current location of the core-business district (CBD). However, they differ significantly in size and stage of urban development, reflected in different streets configurations and ongoing economic dynamics. These differences made these cities adequate to test the applicability of our approach in different urban setups.

GC is the country's capital located in the central region. It accommodates around 26% of the country's population. It extends over 996 km² within the municipal administrative boundary, excluding the conurbation areas in contiguous municipalities. Horizontal expansion is mainly shaped by topographic conditions. A segment of a non-finished peripheral ring connects the foundational core with the

west and south-west areas. Current expansion mostly occurs in the south-eastern, south-western and western areas, outside the administrative boundary. These areas are nurtured by the main infrastructure and the intra-regional CA-1 road.

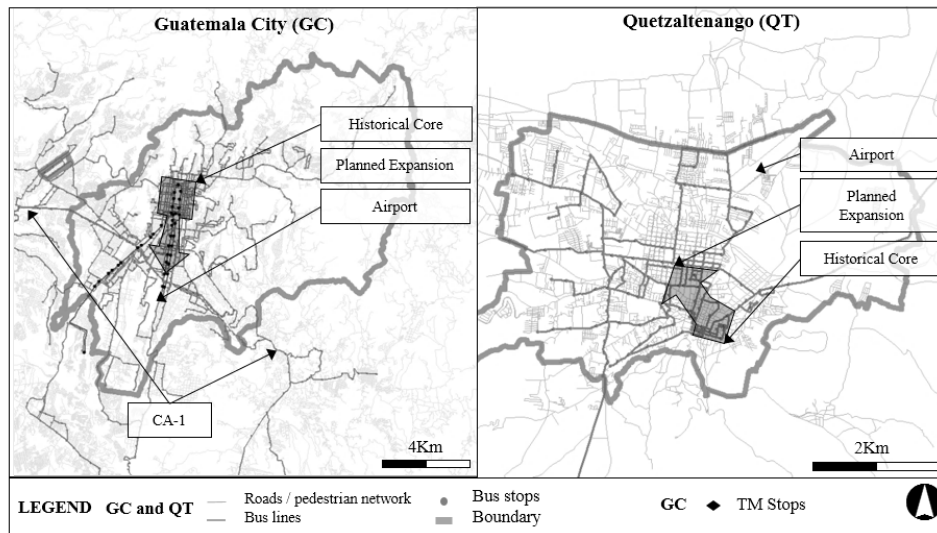


Figure 2.1: Case study areas showing: administrative boundaries, road and public transport network

QT is the second most important city, located in the west of the country. It accommodates around 5% of the population in a minor extension of 120 km². Location of important infrastructure such as the airport, a peripheral road segment combined with topographical conditions influence current expansions mostly towards the northern and north-western areas. Still, further than the foundational core and the first expansion, a slow infill process is observable in the rest of the urban area.

2.3.2 Data collection, VGI data and pre-processing

Table 1 describes the data per accessibility approach that were used for the analyses. Official data were collected during fieldwork in the period of August 2014 to March 2015. The main problems with the data obtained from official institutions were lack of up-to-dateness and incomplete geographical coverage. Data sets in GC only cover the administrative boundary, even though the functional city extends beyond those. Data from the contiguous municipalities do not exist. Therefore, we implemented a tailored approach that included extracting and pre-processing various sources of VGI.

Topological inconsistencies were the main problem when using OSM network data (Cooper, 2014; Cooper & Chiaradia, 2015; Gil, 2015). Pre-processing involved the following steps: filtering out all the roads where vehicles are not allowed following OSM tags convention, planarizing the networks, except at overpass locations, removing duplicated features and detecting and correcting unconnected road segments. We simplified the networks using the Douglas-Peucker algorithm with an offset tolerance of 3 m for the SSx analyses. Then the simplified networks were fragmented at each vertex.

We built time-based network models for private-vehicle (PRIV) and public-transport mobility (PUB) for each city. First, we examined for the correct categorical road classification based on OSM tags convention (e.g. motorway, primary, secondary, residential). A few important roads were found misclassified and were corrected. Speed limits were consulted with the respective authorities and added to the road segments according to their classification. Then we used Waze (Waze, 2015) in GC and field observations in QT to calibrate final travel-times. The PUB models in both cities incorporate the roads as pedestrian networks. Travel-time calibration was carried out in consultation with local planners and experts.

In GC, origin-destination (OD) matrices per mode of transport (PRIV and PUB) during peak hours (6:00-9:00 am) were available for the year 2005. Each matrix contains 173 rows and columns, corresponding to 173 Traffic Analysis Zones (TAZ). Even though trip volumes in these data are outdated, we assumed they reflect overall mobility patterns that have not changed significantly. It means that the major attractors of morning trips, central locations of jobs and commerce, are the same up to date. In QT, OD data are more recent (2014) but drastically less detailed as each TAZ represents one of the 12 postal zones. Only the total number of trips attracted and generated during peak hours is available. Overall, the use of OD data was an alternative solution to overcome unavailable data on job locations. The total number of trips attracted was normalised per TAZ area and used as a proxy variable to indicate the density of job opportunities.

Table 2.1 shows that VGI was the main data source to compensate incomplete land-use data, extracted as points-of-interest (POIs). Still, exhaustive pre-processing was needed to use these data. In OSM most of the land uses were found as POIs, but in some cases they are digitised as polygons, or both. POIs of restaurants or banks within a large-scale mall (as POI or polygon) were simultaneously digitized, so they had to be detected and removed to avoid double counting. OSM conventions are not used consistently and the completeness of the data

sets relies on the active contributors. Thus, other sources (GoogleMaps and Wikimapia) were used to cross-check and complete each land-use. Official land-use data was used mostly as a secondary reference. The final compiled datasets with POI locations were discussed with local planners to validate their use.

Table 2.1: Data collection and pre-processing

	Data	Official data	VGI sources	Observations	Data used and pre-processing
Accessibility approach Geometric / Geographical	Road network	Municipality in GC and public university in QT	OSM, Waze	Official data incomplete in coverage and with no attributes to model travel-time	OSM roads systematically cleaned from topological errors
	Public transport network		Not available	Location of bus stops not available. Inconsistent digitalizations of bus lines in GC	Simplified digitalization of lines, and generation of bus stops at no less than 300mt between each other
	Origin - destination matrices	Municipality in GC and private consultant in QT	Not available	Used as a proxy for job location	Estimation of density of trips attracted during morning peak hours
	Jobs location	Not available	Not available		
	Large scale malls	Municipality in GC and public university in QT. Available with data on floor area	OSM, Google maps, Wikimapia. Mostly only location available without floor area data.	Official data set in GC from 2004. Data incomplete in both cities. Spatial database in QT (from 2013) with inconsistencies in its land-use attributes	POIs extracted and compiled from VGI sources. Where possible official data sets used as secondary reference to validate VGI locations. Feedback from local planners on the location patterns of these uses, specially where official data were not available.
	XL grocery shops				
	universities				
	Culture				
	XL sport facilities				
	Local grocery shops				
	Restaurants / banks				
	Parks				
	Municipal markets				
	Hospitals	Geoportal of the National Planning Secretary		Only available as points	Use of official data
	Schools				
	Clinics				

2.3.3 Implementing accessibility analyses

Figure 2.2 shows our methodological framework to map urban accessibility using a geographical and a geometrical approach. At a first glance, differences in data requirements are noticeable in each approach. In the geographical approach we started with the analysis of access to different land uses, by grouping them based on two sets of variables: macro- and micro-location. Macro-location addresses all those variables where, due to their characteristics (relevance and scale), people are more willing to overcome impedance. Micro-location addresses all locations with characteristics related to neighbourhood scale; thus, people are relatively less willing to overcome impedance. During the fieldwork we validated the relevance of these variables by means of a workshop with local experts in planning and real estate markets. Participants also ranked the variables on the basis of the relative importance of each variable to the local land market.

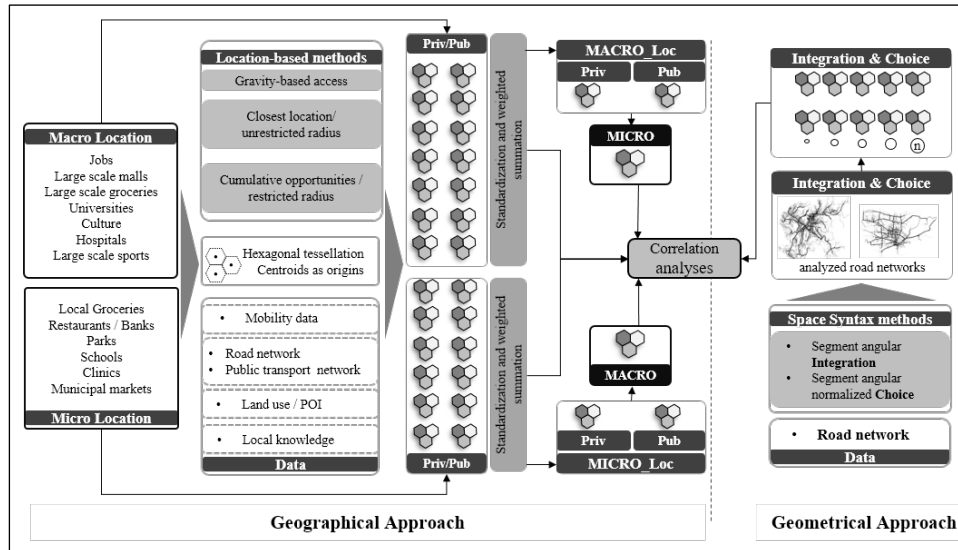


Figure 2.2: Methodological framework

We generated a hexagonal tessellation for each study area. A hexagonal shape was chosen for the following advantages: (1) spatial sampling performing slightly better when using geostatistical techniques compared to a grid sampling (Birch, Oom & Beecham, 2007; Condat, Van De Ville & Blu, 2005), if the analyses were to be used to investigate their relation with other socio-economic dynamics; (2) more symmetric nearest neighbourhoods, avoiding the ambiguities of a rectangular grid (Birch et al., 2007); (3) better visualization (Birch et al., 2007; Burdziej, 2012). The size of the cells has a circumscribed diameter of 300 m, equivalent to a distance of three blocks in the study areas. The size provides a reasonable resolution at a moderate computational demand. Cells that were not overlaying any roads were removed. The remaining hexagonal centroids were used as origins for the analyses.

We used the location-based methods described in section 2.2.1. The GC OD matrices were used to estimate the decay parameters α and β , following a probabilistic approach (Ingram, 1971). Population-weighted centroids were used during this process to avoid aggregation biases (Hewko, Smoyer-Tomic & Hodgson, 2002). Using ordinary least squares, decay functions per mode of transport were fixed as: $\alpha=1.2$ and $\beta=0.052$ for PRIV; $\alpha=1.3$ and $\beta=0.002$ for PUB. These values were used with equation 2.2 and the density of trips attracted as “M” to map access to jobs in both cities. The rest of the macro-variables were analysed using the shortest travel-time to reach a facility. All the

micro-location variables were analysed using a cumulative-opportunity method, restricted to 10 minutes for both mobility modes.

Then we produced integrated maps per mobility mode (PRIV and PUB), at macro and micro level. First, we standardized the results to a '0' - '1' scale using the non-linear standardization methods (Nyerges & Jankowski, 2009). Then the results were combined per mode of transport using weighted summation. The ranking by experts was then used to estimate the normalized weights using the rank sum method (Malczewski, 1999). Finally, the PRIV and PUB results were also combined using the same procedure. The weights were equal to the percentage of trips generated per mode of transport from each TAZ. This step made the whole analysis sensitive to types of users and the accessibility per modality that benefits each location the most.

In the geometrical approach, we used SAA to analyse integration, choice, and their normalised value at different spatial radii. We started with a minimum radius equal to the average neighbourhood size (0.8 km). Then we expanded it to 1.5 Km and 2.5 km. From there we produced analyses by increasing the radius by 2.5 km up to the longest radius that would still produce information visually different than a global metric (7.5 km in GC and 5 km in QT). We aggregated the results, originally stored per road segment, to the hexagonal cells. Thus, each cell contains the average integration and the maximum choice values. Finally, we produced correlation matrices comparing the results from both approaches at $p < 0.01$.

2.4 Results and discussion

2.4.1 Geographic accessibility

Figures 2.3 and 2.4 show accessibility results per variable and per mode of transport for both cities. In GC, job access highly benefits the core area and rapidly decreases outwards. Access to large scale (XL) malls and grocery shops mostly benefit a corridor area from north-west to south-east. Access to universities and hospitals tend to benefit the most central east-western areas.

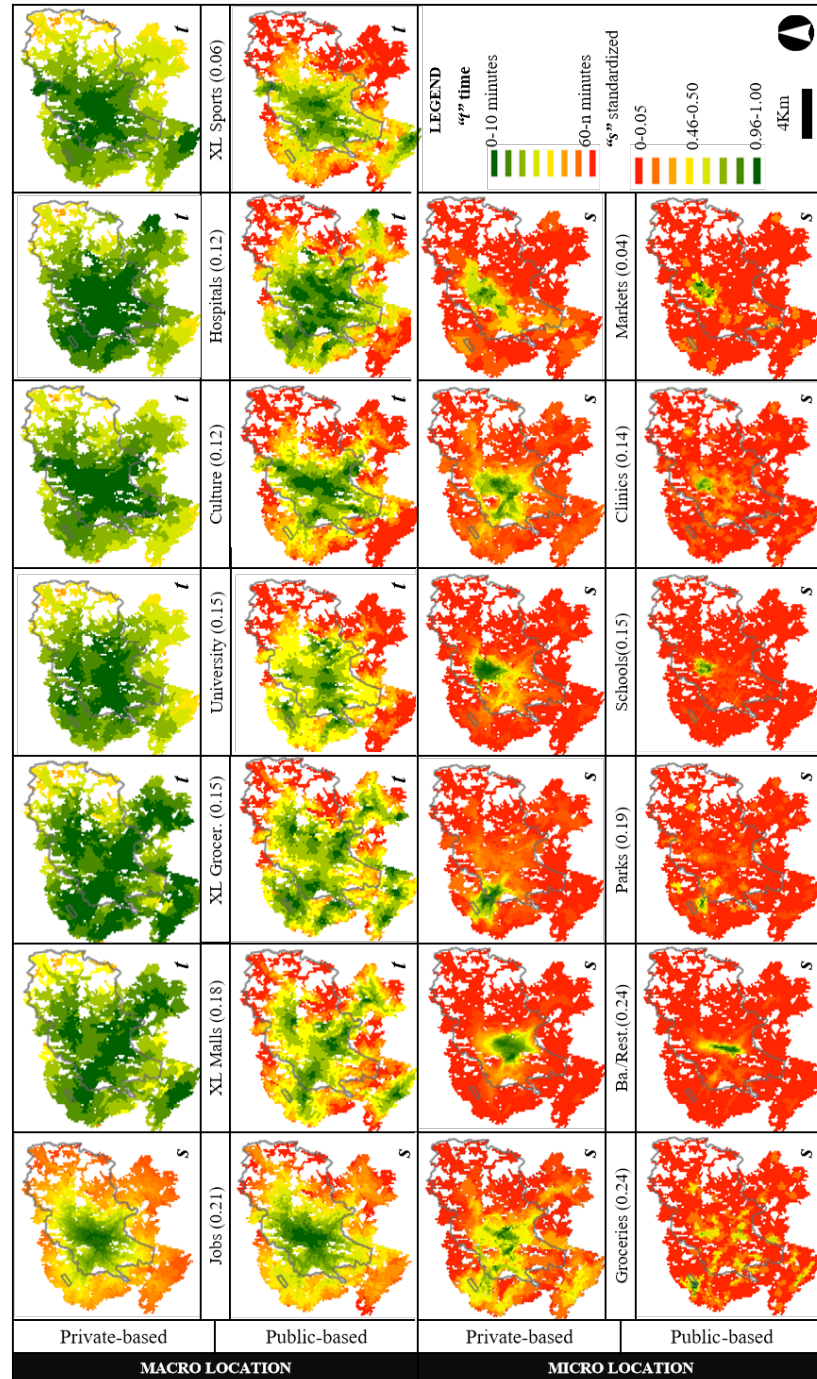


Figure 2.3: Accessibility per variable per mode of transport in Guatemala City

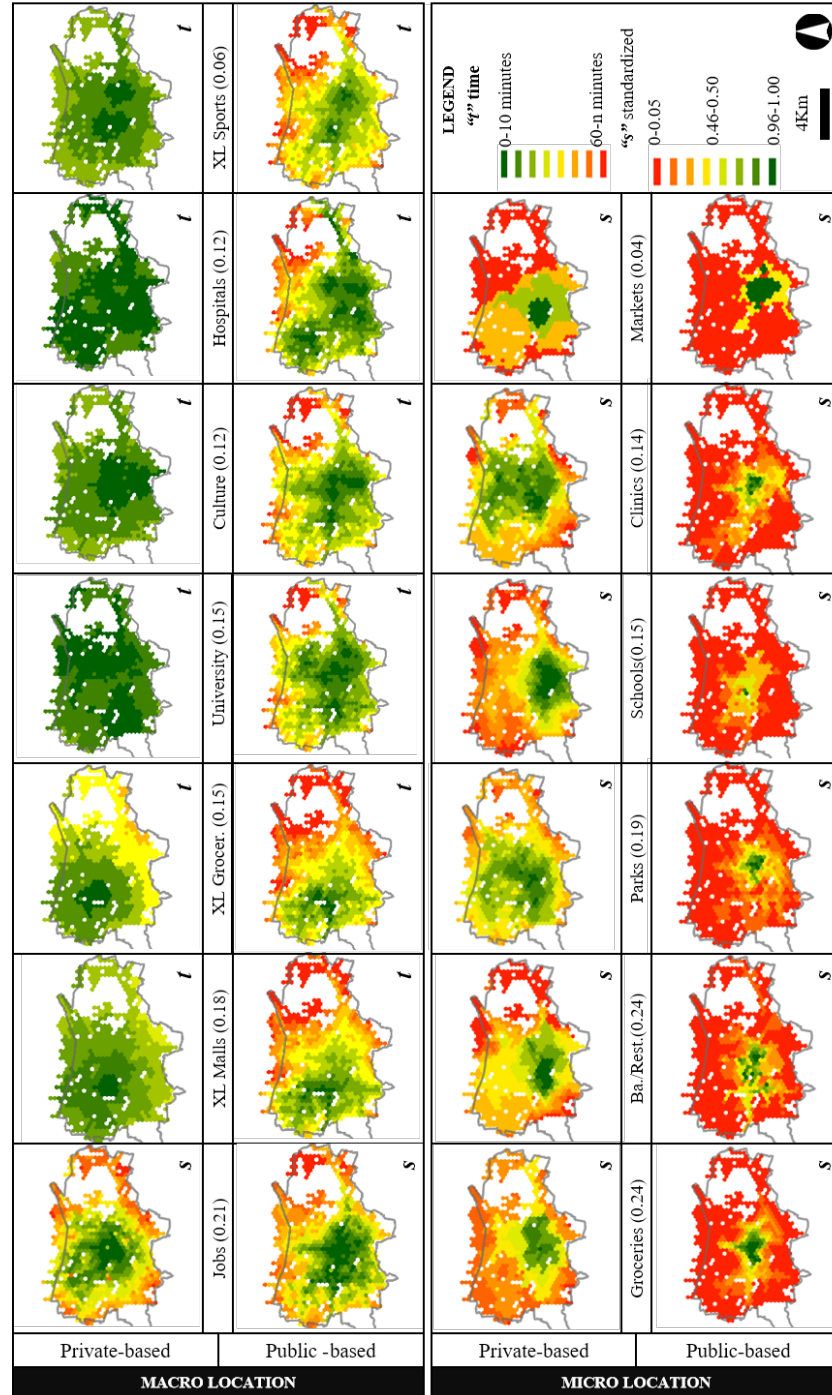


Figure 2.4: Accessibility per variable per mode of transport in Quetzaltenango

Access to culture and XL sport-related facilities (e.g. stadiums) outlines a north-south corridor. At the micro-location, low and medium-low access predominates in the maps. Still, the core corridor simultaneously benefits from access to multiple locations of grocery shops, banks and restaurants, schools, clinics and municipal markets. Access to parks points out an urban area in the north-west. Here, various small parks are found within residential neighbourhoods.

The core area shows a relatively low-medium parks access. However, important public and larger open spaces are located in this corridor, especially in the historic CBD. This outlines the limitation of not including the size of each facility in the analysis. Overall, public transport availability and travel-times strongly constrain accessibility at macro and micro-location scale. Repeatedly, the north-eastern areas lack good accessibility.

In QT, private access to jobs, cultural facilities and XL sports favours the most central and historic areas. High access to XL malls and XL grocery shops reflects an emerging commercial pole located in the north-eastern area. Highest access to universities and hospitals are less associated with high access to other variables. Overall, public transport restricts high accessibility to different variables, mostly to core areas. Public transport availability and quality highly restricts accessibility. Mostly, central and historic areas benefit from high access to various land uses. The central area still benefits from medium high access to XL malls and grocery shops. Also, it benefits the most from high access to all the micro-location variables using both mobility modes.

Figures 2.5 and 2.6 show integrated macro- and micro-location accessibility maps for PRIV (a), PUB (b) and both modes combined (c) for each city. Standardized "s" colour ranges from figures 2.3 and 2.4 are applicable to these figures. Accessibility scores range from 0 (red) to 100 (green). Road segments scoring high SSx NACH values, classified in two, are overlaid on the access maps showing the combined scores (c). Discussion of this layer is provided in next sections.

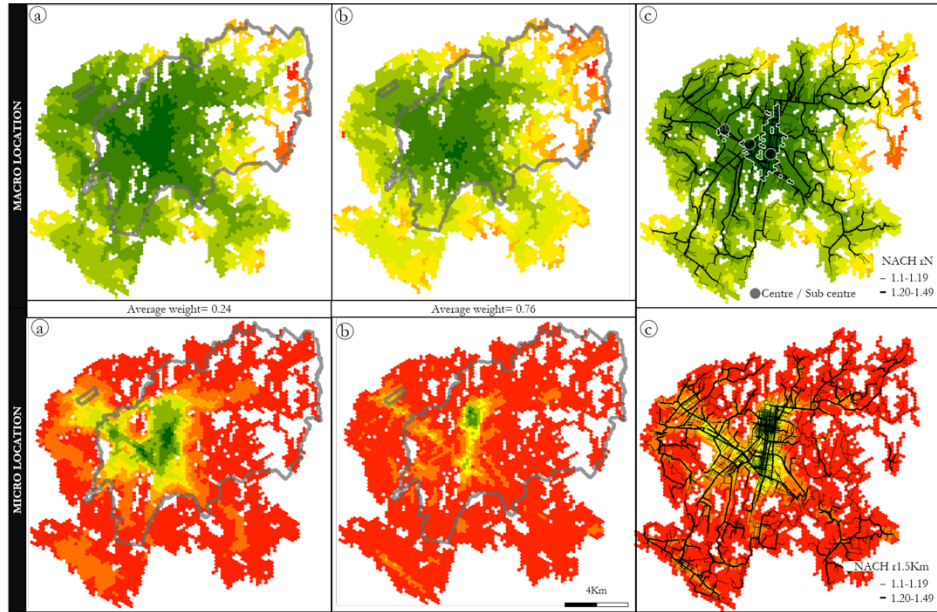


Figure 2.5: Integrated macro and micro-location accessibility for Guatemala City (same legend as in figure 2.3 and 2.4).

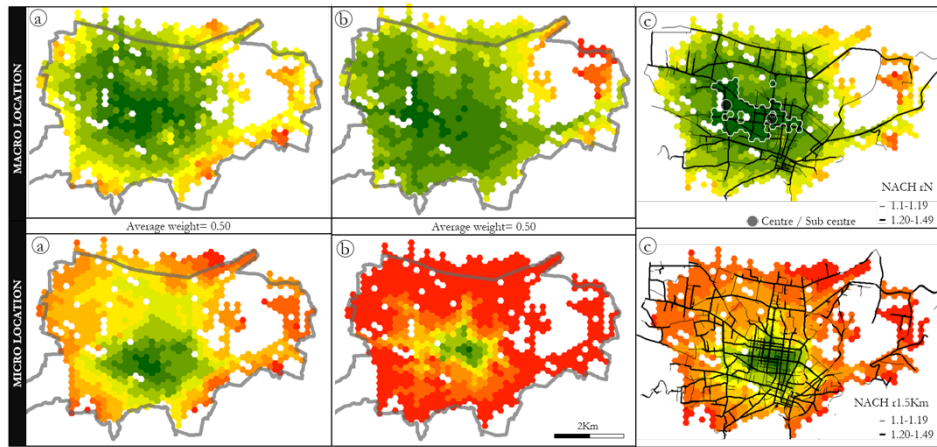


Figure 2.6: Integrated macro and micro-location accessibility for Quetzaltenango (same legend as in figure 2.3 and 2.4).

In GC the core corridor benefits the most from PRIV and PUB access given the current mobility infrastructure to access macro-location variables (maps "a" and "b"). The combined effect of macro and micro-location accessibility, maps "c", is mostly influenced by PUB accessibility. That is because the average percentage of public transport users per TAZ is higher (0.76) than private vehicle ones

(0.24). Highest macro accessibility (scores 0.91-1) outlines three important areas in map (c): the current CBD and two inter-connected sub-centres aligned along the CA-1. These cores match the locations perceived by local experts as important economic centres and important job sources. Contrastingly, current policies are mostly focused on centrality at the north-south core corridor. However, an opportunity to strengthen an existing polycentric structure extends towards the north-west. Micro-location emphasises the historic centre, core corridor, and quickly decreases towards the other sub-centres with medium scores. Medium-low scores (0.3-0.4) point out areas that benefit from minor concentrations to a combination of various micro variables.

In QT the modal split is equal for the whole area (50%-50%). The highest combined macro accessibility benefits central areas and extends towards the emerging commercial pole. Accessibility slowly decreases towards the periphery. The east periphery has the lowest macro accessibility. Micro-location benefits core areas the most, including the historic centre. Accessibility quickly decreases towards the periphery. Unlike GT, the emerging pole benefits the most only from high macro accessibility.

2.4.2 Geometric accessibility

Figure 2.7 shows the SSx results of integration and NACH at selected radii. In GC, the core areas benefit the most from geometric accessibility at various spatial scales. Neighbourhood scale integration (r0.8 km) outlines urban areas with compact grids (small blocks). These areas correspond to some of the old planned neighbourhoods at the core areas and the core settlements of the peripheral municipalities. These areas have the highest potential for pedestrian movements. High integration (1700~) at r5 km highlights the core corridor and extends towards the southern part following important roads. Eastern, southern and western peripheral areas account for medium integration (600-100), while some areas in between these and the core account for lower integration values. Overall, these patterns already provide visual insights on the important association between integration and the agglomeration of economic and service activities in the core areas versus the peripheral areas, as outlined by geographic access. This predominance is reflected in the disposition of city structuring roads when observing NACH analyses (r5 km and rN). The core corridor is the only area framed and traversed by these important roads, while towards the periphery the urban patches are connected to this framed core by tree-type configurations. The peripheral roads play an important role structuring the western area, producing an important

intersection where one of the sub-centres is outlined via macro accessibility.

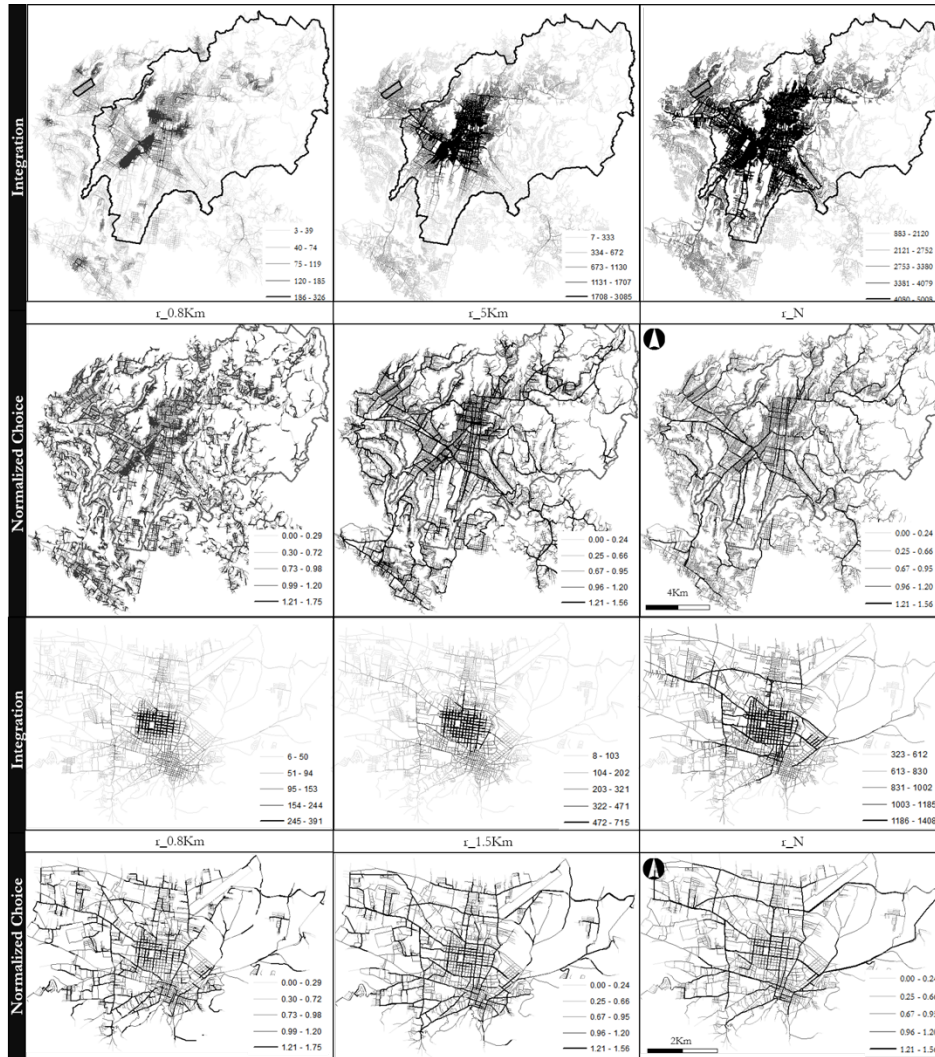


Figure 2.7: SSx results for Guatemala City and Quetzaltenango

In QT the central area is highly integrated at different spatial scales, more than the historic centre, which is associated with the high centrality of such an area as visualized via geographic access. High integration at r_N outlines areas towards the north-west, matching the location of the emerging commercial pole. Contrastingly, other peripheral areas are poorly integrated at different spatial scales, denoting less consolidated urbanization. High NACH values at the lowest radius outline service roads at neighbourhood scale. Similarly

to GC, both the central area and the historic centre seem to be well connected by means of city structuring roads outlined by NACH analyses (r 1.5 km and rN). Those connect to other important roads: a periphery east-west road on the north (intersecting with the emergent pole) and a major south-north road on the west side.

During the analyses, integration and NAIN were both explored. However, we observed inconsistencies of the normalized measurement at low radii, as reported in Hillier et al. (2012). In GC, analyses at r0.8 km and 1.5 km were highlighting isolated segments in peripheral areas. In QT, the same inconsistencies were observed with analyses at r0.8 km, 1.5 km, 2.5 km and even 5 km. This was expected as the urban layout in QT is less consolidated compared to GT.

2.4.3 Associations between the geographical and the geometrical approach

Figures 2.8 and 2.9 show the results of Pearson correlation matrices ($p < 0.01$) between geographical (location-based) and geometrical (SSx) accessibility, per mobility mode for each city. Square size and colour range, from small to large and from white to black, indicate correlation strength. The matrices reveal the associations of the various accessibility measures within and between approaches. We confirm a positive correlation between geometric and geographic accessibility to various facilities and aggregated macro- and micro-location accessibility. The results do not indicate a directional causality. However, high geographic accessibility is the result of location and concentration of various land uses. Various land uses are more prone to change or relocate over time, compared to changes in the urban layout. Thus, we could think of the geographic accessibility as a result of a cumulative process of land-use location influenced by seeking optimal geometric location. We first discuss the correlations between the various location-based analyses and then the correlations between location-based measurements and SSx.

Figure 2.8 shows that the distribution of PRIV accessibility in GC to various macro-variables is slightly less similar compared to PUB. While PRIV mobility infrastructure is more evenly distributed, varying locations of macro-variables produce more irregular patterns. It reflects that some areas do not benefit simultaneously of the same access levels to all macro-variables. In turn, the PUB infrastructure restricts homogeneously higher access to macro-variables only to the core areas. The remaining areas have simultaneously poorer access to multiple macro-variables. Higher and significant correlations between macro-variables and micro-variables in the upper matrix outline that

only PRIV mobility offers simultaneous access to macro- and micro-variables, unlike PUB.

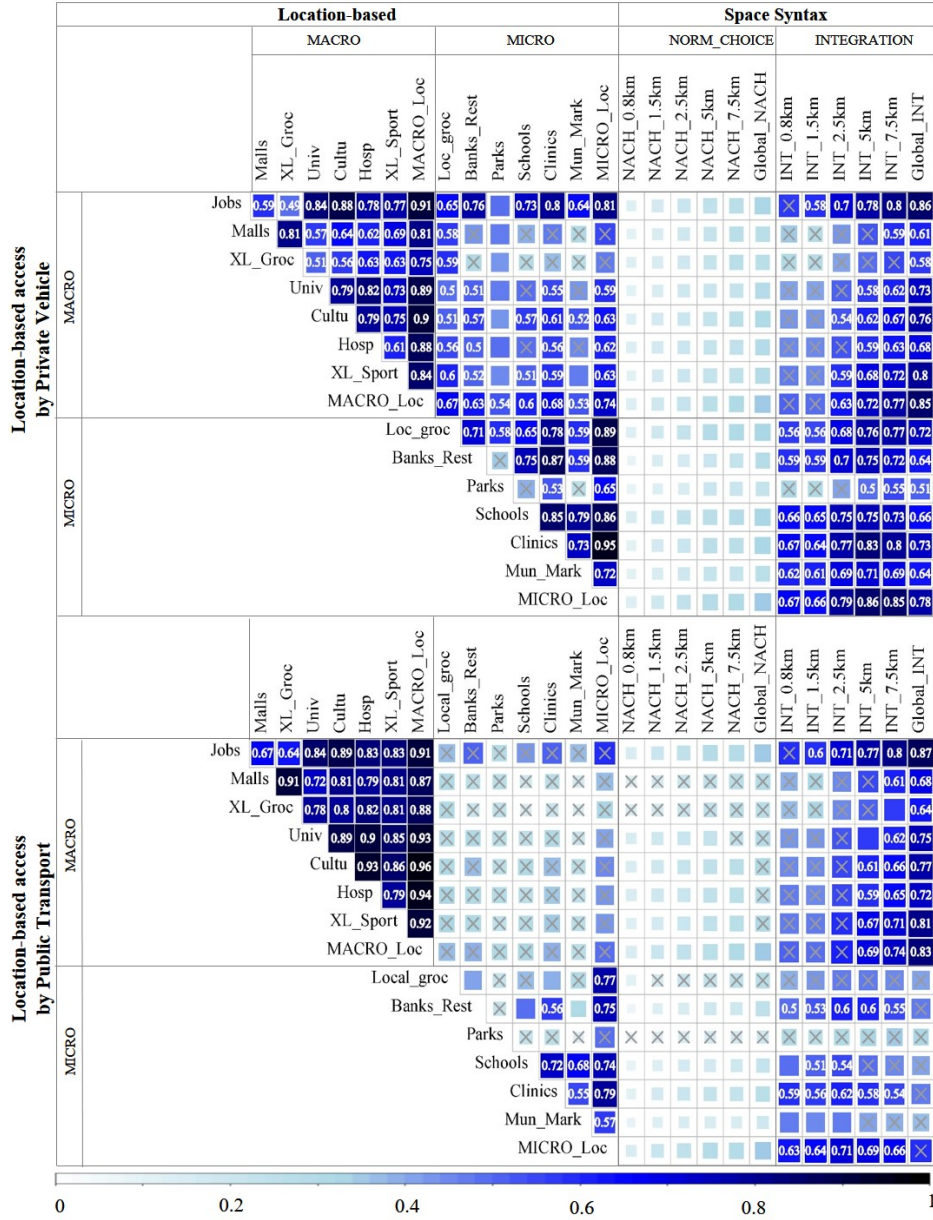


Figure 2.8: Guatemala City. Insignificant correlations ($p < 0.01$) are cross marked.

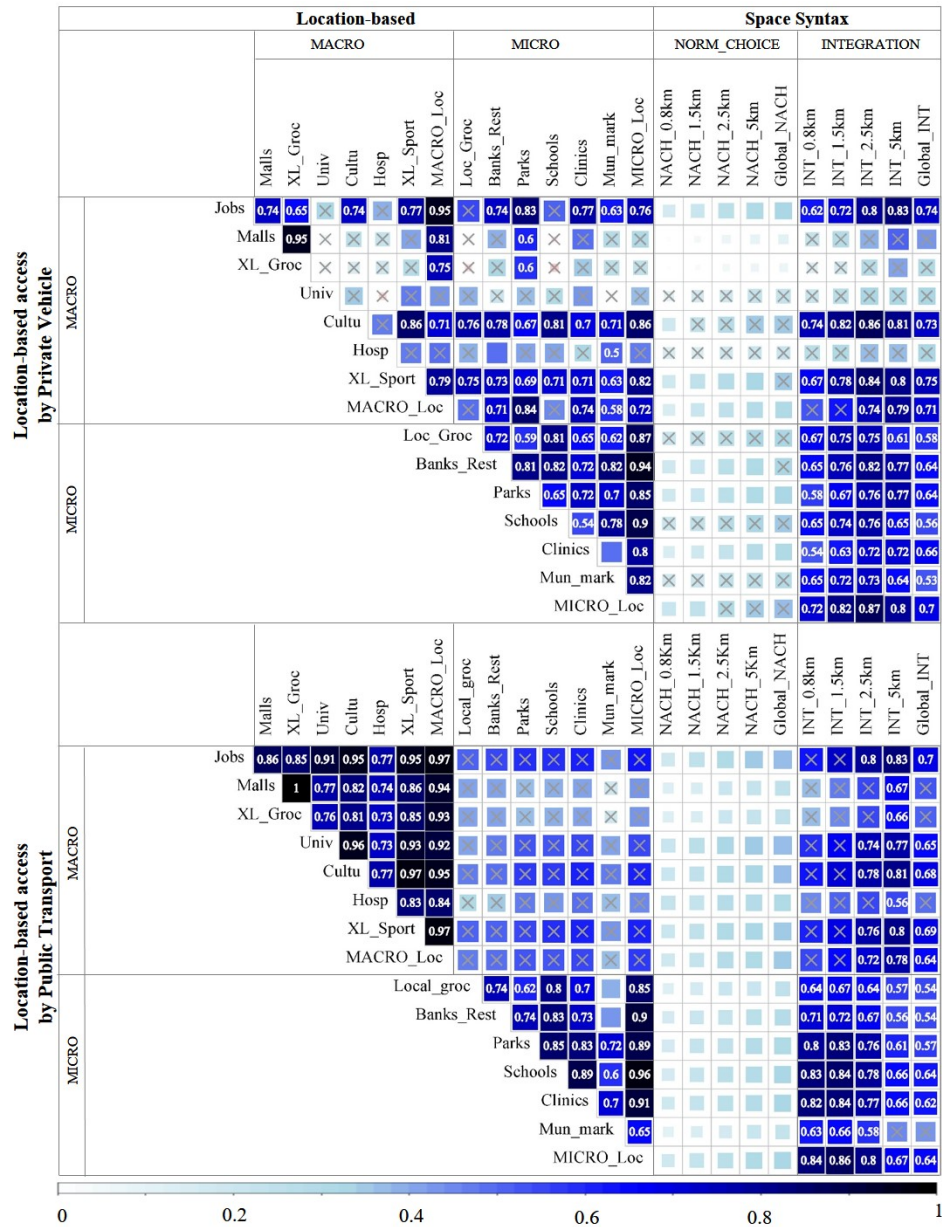


Figure 2.9: Quetzaltenango. Insignificant correlations ($p < 0.01$) are cross marked

Figure 2.9 shows that in QT PRIV accessibility patterns to malls, XL-grocery shops, hospitals and universities differ highly from accessibility to various other land uses. Partially, this is explained as the first two are located in the emerging commercial pole. In turn, the PUB

infrastructure simultaneously benefits with high accessibility to the same areas, a reduced portion of the city. The rest simultaneously benefits from medium or poor access to various macro-variables. Access distributions to micro-variables are significantly similar both for PRIV and PUB. Unlike the location of macro-variables, these are concentrated in the same areas. Similar to GC, significant and insignificant correlations between the macro with the micro-variables show that PRIV mobility means higher access to both groups of variables than PUB mobility.

Potential PRIV access to jobs strongly correlates with various location-based metrics in both cities. The cumulative access to areas that attract more trips during morning peak hours works as a latent variable showing areas favoured by good access to various commercial and service type of land uses (job sources). As was expected, it emphasizes the important role of such a metric to understand the distribution of economic opportunities, as well as its potential to visualize aggregated geographic access via a robust location-based metric using only trip data.

The highest correlations between location-based measurements (PRIV, PUB) and NACH were found at rN in both cities. Although not all of them are significant at $p < 0.01$, they are at $p < 0.05$. In QT correlations between PRIV location-based accessibility to malls, XL-grocery shops and NACHrN are insignificant in both cases. Even though correlation levels are low, these modestly provide evidence of location preference of profitable or service land uses. It shows an association between route choices as a function of travel time with a cognitive criterion of least angular deviation. By looking at maps c in figure 5 and 6, we could suggest that NACH metrics provide relevant information that is complementary to location-based metrics by outlining those structuring roads from where geographic access distributes across a city. We observe that the distribution of location-based accessibility closely follows the arrangement of the important roads. Furthermore, this confirms the preference of the various land uses for roads where higher flows of people are expected.

Macro variables consistently show the highest correlation with high radii integration, rN in GC and mostly with r5 km in QT. This difference shows the effect of the different territorial extensions and stages of urban development. In QT, high rN integration is associated with a higher concentration of economic activities, or at least where there is potential for those. The emergence of the west economic pole responds to this potential. However, location-based access to most of macro-variables tends to outline more the central areas. Therefore, high SSx

integration at r5km captures better the current location of most of these uses. Furthermore, this also explains the disassociation between PRIV location-based accessibility to malls and XL-grocery shops, and integration values.

In both cities micro-location accessibility shows the highest correlation with integration at lower radii. We found that lower radii reflect restricted time-based mobility in two ways: the 10-minute threshold imposed in the analysis and mobility restrictions imposed by the current public transport infrastructure. In GC PRIV micro-access shows the highest correlation with integration-r5 km and PUB with r2.5 km. In QT PRIV micro-access shows the highest correlation with integration-r2.5 km and PUB at r1.5 km. These differences also show that micro-location land-uses in QT are spatially distributed and more accessible to pedestrian movements, compared to GC.

The correlations provide insights about how integration at different spatial scales (radii) is associated with access to various land uses. SSx analyses the geometric access as an aggregated resource, assumed to be preferred by simultaneous location and concentration of the various land uses. Aggregated location-based variables (Macro_loc and Micro_loc) attempt to reflect this dynamic. In our case, aggregated Macro_loc and Micro_loc address an expert-based ponderation of each land use regarding the local land market. As expected, both aggregated variables are associated the most with NACH values at rN in both cities, no matter the transport-mode. Correlation between Macro_loc and integration in GC peaks at rN and at r5 km in QT for both transport modes. The relative importance of access to jobs becomes evident in the case of QT when addressing PUB access. In GC Micro_loc is associated most with integration at r5 km-r7.5 km for PRIV access and r2.5 km-r5 km for PUB access. In QT Micro_loc is associated most with integration at r2.5 km and r1.5 km correspondingly. The divergence per mode of transport in both cities emphasizes the mobility restrictions posed by PUB mobility using the same 10-minute travel-time, but more evident for the case of QT.

Location-based methods were adapted to consider the accessibility components as long as data allowed it. The results are rich in information about accessibility to various land uses: indeed, 32 maps per city. In turn, the geometrical approach considered only two measurements (integration and choice) at different analysis radii: 12 maps per city. SSx outputs are intuitive in terms of visualizing accessible locations, where we can assume that profitable uses or service facilities are present. Differences in data requirements and pre-processing give significant advantage to the geometrical approach.

However, previous training on SSx and local knowledge is required for adequate interpretation. Besides, the applicability of the SSx might be limited in cases such as QT, where ongoing development is sparse. Furthermore, VGI was applicable in both approaches in this research. After exhaustive pre-processing these data became indispensable to implement the geographic approach. In turn, data for a geometric approach could be manually or digitally derived from other sources such as satellite imagery.

2.5 Conclusions

Urban accessibility was successfully mapped using a geographical and a geometrical approach and the results were compared using correlations. We found consistent correlations between accessibility measurements from both approaches, which allows us to confirm our initial hypothesis. The results do not explain directionality of the causal relation between geometrical and geographic accessibility, but it is logical to think that in Global South cities with weak land-use policies, geometric accessibility influences a location process where various land uses seek to benefit from reachable locations (integration) and exposure to movement (choice).

Our research does not solve the analytical dual approach problem outlined in Batty (2013). However, the results establish an empirical connection between the morphological-geometric properties of an urban area and the distribution of time-based geographic accessibility. The quantitative relations provide additional knowledge on the interpretability and limitations of the information that is produced in Space Syntax using little data. Our results are placed in the context of studies such as Kim et al. (2002) and van Nes et al. (2011), where positive associations are claimed between SSx metrics with building density and location of various land uses, which are observed in the present work, and is in line with Space Syntax capacity to aid in predicting flows of people (Hajrasouliha et al., 2015).

The consistency of the correlations between geographic and geometrical measurements simultaneously validates the use of VGI, and the applicability of Space Syntax in these two cities. Location-based analyses using VGI data produced plausible results that outlines how access to various land uses is distributed across each city. In turn, Space Syntax analyses applied to different urban contexts produced information that is statistically comparable with the results of a location-based approach. Based on the correlation results we claim that a geometrical approach using SSx delivers plausible information from where inferences could be made about geographic accessibility. Space Syntax turns out to be a more data/time efficient approach.

We found some limitations when applying Space Syntax in less consolidated urban areas, besides the problems with the normalised measurements reported in Hillier et al. (2012). Global integration measurements might not reflect the ongoing land-use processes in less-consolidated areas consistently. We cannot conclude from the results which spatial radii best describe geographic access, as these differed per city. This is one important drawback of SSx analysis, as there is not enough evidence to suggest which spatial radii correlate better with what in different cities. This observation could also apply to the MCA method (Porta et al., 2012). Further research could address these observations by replicating the research in a larger set of case studies and expanding the comparisons by including network centrality metrics from other methods, such as the MCA.

The methodologies and results are important for the planning practice. In Guatemala the results provide important information for transport and land-use planning, which was not available before. Mobility and land-use projects could be assessed using the methodologies presented here. We suggest that our methodological framework, including the use of VGI, is replicable to other cities in the region and Global South, where data are scarce, but information is highly relevant for planning tasks. Space Syntax would be a valid approach for planners and researchers in areas where not even VGI is available, for example, by analysing the attractiveness of places, the city structure, or geometric accessibility impacts of changes in street configuration. Finally, although the outputs of both approaches are statistically comparable, rather than replacing location-based methods, Space Syntax offers an important complementary measure to geographical accessibility. We describe Space Syntax as an accessibility tool able to support early-stage planning processes when limited data are available. Trained interpretations enriched with local knowledge could provide first-level insights on concentrations of profitable and public services land use, and the distribution of access to those. Such interpretations could be further examined in detail for areas of interest using location-based methods, potentially supported by VGI. Accessibility analysis could benefit from addressing time, land-use and geometric aspects to improve the understanding in other fields such as human geography and urban economy in cities in the Global South.

Chapter 3

Modelling residential land values using geographic and geometric accessibility in Guatemala City *

* This chapter is based on: J. Morales, J. Flacke and J. Zevenbergen. Modelling residential land values using geographic and geometric accessibility in Guatemala City. *Environment and Planning B: Urban Analytics and City Science*, 46(4), 751-776, 2019.

Abstract

Location and accessibility are core concepts for land-value research. However, the perspective is still limited in their conceptual and methodological application to cities from the Global South. The objective of this research is to bridge concepts and definitions to comprehensively operationalize accessibility indicators and uncover its relation with residential land-values in Guatemala City. We developed a multivariate regression model using the following access metrics: (1) geographic-access indices that were computed using time-based analyses per transport mode; (2) geometric-access metrics estimated via Space Syntax at various spatial scales; (3) a proposed geometric via geographic-access metric computed as potential access to network centrality. A variable selection process allowed to assess the information contribution of each variable in building a parsimonious model. We assessed the model in the context of model variations that represent common approaches used in existing literature. Geographic access to the central business district has the highest impact on the land-values, followed by proximity to urban areas with high geometric-access, measured as geometric via geographic access. Geometric accessibility at neighbourhood and city-wide scales add spatialized information that contributes to a parsimonious model and reduces spatial dependence. The model yielded the highest goodness of fit and prediction accuracy compared with the model variations. We concluded that Guatemala City land-values follow a predominant monocentric structure. Additionally, potential access to vital urban areas as identified via Space Syntax denotes the presence of economic activities, or potential for such, which were not explicitly addressed through the geographic-access metrics. The results have limitations but pose methodological possibilities relevant for research and practice in similar Latin American cities.

3.1 Introduction

Following Krause et al. (2012), trends in land and property value research can be grouped in three. First, uncovering the relation between the various factors that can influence the value formation is commonly done using hedonic pricing, and implemented through multivariate linear regression techniques (Liu, Zheng, Turkstra & Huang, 2010; Munroe, 2007). However, there is a growing body of literature concerned with the presence of spatial dependence and the statistical methodological expansions to address it, such as spatial econometrics (Bourassa, Cantoni & Hoesli, 2007; LeSage & Pace, 2009). Second, literature is placing attention on studying the value of land after recognizing differences between this value and the value of the improvements (i.e. construction). Land is fixed both in territorial location and its supply (Evans, 1987), which could make it more volatile and sensitive to the effects of demand, location and economic shocks (Krause et al., 2012). Third, interest is taken place on studying the value effects derived from sustainable urban processes such as mixed-uses, mass transport and street connectivity (Anantsuksomsri & Tontisirin, 2015; Matthews et al., 2007). It can be said that this has always been central in this research area, but only implicitly by means of the urban accessibility concept (Adair, McGreal, Smyth, Cooper & Ryley, 2000; Ahlfeldt, 2007; Des Rosiers, Thériault & Villeneuve, 2000; Giuliano et al., 2010; Liu et al., 2010).

Location and accessibility are core concepts in land-value and house-price research (Orford, 2002; Webster, 2010). Important urban economic theories (Alonso, 1964; Evans, 1987; Muth, 1969) and a large body of literature prove that relative location and access externalities are associated with the economic value of various land-uses (Dale-Johnson & Jan Brzeski, 2001; Kivell, 1993; Liu et al., 2010; Peiser, 1987). Basic accessibility concepts and methods have been commonly at the core of previous research on land-value modelling, particularly in urban areas (Iacono & Levinson, 2011; Saeid, 2011). However, cities are no simple phenomena and there is still room for exploring their application, particularly in cities from the Global South.

We address location using two types of accessibility. *Geographic-accessibility* reflects the easiness to reach a location or to be reached given available infrastructure to overcome impedance between origin and destination (Batty, 2009). Distance, travel-time, cumulative opportunities and gravity-type measurements are metrics commonly used in research and practice (Geurs et al., 2001). In turn, *Geometric-accessibility* is defined as a resource given the topologic-geometric characteristics of a city urban layout (Jiang et al., 1999; Webster, 2010). Network-closeness and network-betweenness are common

network centrality metrics that are associated with geometric-accessibility (Freeman, 1977; Porta et al., 2005). Space Syntax (SSx) is a set of theories and methods based on graph theory and dual representations of the urban layouts that is pioneer in network centrality applied to urban areas (Bafna, 2003; Hillier et al., 2012; Webster, 2010). Integration and choice are SSx metrics, adopted in this research, equivalent to closeness and betweenness correspondingly.

The trade-off neoclassic theory (Alonso, 1964; Evans, 1987, pp. 18-21) constitutes a fundamental basis for land-value and house-price literature (Ahlfeldt et al., 2011; Kivell, 1993, pp. 14-36). Accessibility is then mainly addressed as proximity to a central business district (CBD) and other location externalities. Proximity measurements are commonly included as aerial or network-based distances (Bourassa et al., 2007; Dale-Johnson et al., 2001; Heikkila et al., 1989; Liu et al., 2010; Munroe, 2007; Orford, 2002; Waddell, Berry & Hoch, 1993). Metric distances are not realistic measurements as road characteristics and transport modes pose different travel times or costs over the same distance (Mavoa, Witten, McCreanor & O'Sullivan, 2012; Ryan, 1999). Some literature used travel-times instead (Ahlfeldt et al., 2011; Iacono & Levinson, 2015; Ottensmann, Payton & Man, 2008; Pujol, Pérez & Sánchez, 2013). However, the focus has mostly been placed on travel times by private mobility. Relatively few research has tested more complex methods such as the potential measurements to replace the traditional CBD proximity (Adair et al., 2000; Ahlfeldt, 2007; Du & Mulley, 2012; Giuliano et al., 2010; Osland & Thorsen, 2013). The rationale is that a monocentric CBD assumption might not adequately capture access to economic opportunities due to emergence of polycentric structures.

The availability of public transport is mostly addressed as proximity to access points (i.e. stops and stations) to such infrastructure (Ryan, 1999), (e.g. Anantsuksomsri et al., 2015; Iacono et al., 2011; Ibeas, Cordera, dell'Olio, Coppola & Dominguez, 2012; Rodríguez & Mojica, 2009; Waddell et al., 1993). Such approach is commonly used to sort out the effects of public transport investments on house-value properties, again assuming mono-centricity (Ryan, 1999). Only very limited research has used simultaneously travel-times measurements by more than one transport mode (Adair et al., 2000; Des Rosiers et al., 2000; Du et al., 2012). Adair et al. (2000) used travel times per transport mode and estimated mean access values for each neighbourhood in its study area.

Only a small body of literature has addressed geometric accessibility combined with basic geographic access metrics. Desyllas (1997) investigated the relationships between the evolution of land-use and geometric-access with land-values in Berlin. Matthews et al. (2007) investigated the effects of integration and proximities to selected land-uses on house-prices in Washington. Enström et al. (2008) investigated the effects of integration on office-rent variation in downtown Stockholm. Chiaradia, Hillier, Barnes and Schwander (2009) researched on the associations between geometric access and dwellings values in London. Saeid (2011) analysed the relation between integration with land-values in Wroclaw. Xiao et al. (2016a) studied various spatial scales at where geometric-accessibility, together with selected geographic-access metrics, better explains house-price variability in Cardiff.

Overall, the outcomes of these works provide evidence on how geometric access can explain land-values more accurately. SSx metrics at various spatial scales provide additional information about the quality of the urban layout that other access metrics cannot capture. Furthermore, these metrics could have the potential to improve the quality of the model itself in terms of reducing heteroskedasticity and spatial autocorrelation (Xiao et al., 2016a). SSx methodologies have evolved and developed in more sophisticated analyses compared to the early applications (Steadman, 2004; Turner, 2007), but only limited research have implemented along newer techniques in land-value investigations (Chiaradia et al., 2009; Xiao et al., 2016a). Additionally, these works are mostly limited to case studies from developed and planned cities.

We define the research gap as the need of bridging available concepts and methods in the field of accessibility studies with the land-value modelling task. We introduce the following hypotheses. (1) Addressing the disparity of geographic-access opportunities due to available transport modes and the geometric-access at various spatial scales could contribute to an increased capacity to explain land-values variability and prediction accuracy. (2) Geometric-accessibility capitalizes land not only at location, but also as a reachable resource by means of geographic-access. Thus, geometric via geographic accessibility is defined as the easiness to reach geometric access by means of private or public transport-based mobility. Such metrics could reflect access to urban areas with presence of facilities or economic activities (or a potential for such), which are not explicitly addressed in other geographic-access metrics.

The objective of this chapter is to bridge concepts and definitions to comprehensively operationalize accessibility indicators and uncover its relation with residential land-values in Guatemala City. We developed one multivariate regression model that used the following access metrics: (1) geographic access indexes that were computed using time-based analyses per transport mode; (2) geometric access metrics estimated via SSx at various spatial scales; (3) a proposed geometric via geographic access metric computed as a potential access to network centrality as analysed in SSx. A parsimonious model is estimated following a variable selection procedure and then assessed using various performance statistics.

The remainder of this chapter is organized as follows. Section 3.2 introduce the study area, data pre-processing and methods. Then the results are presented and discussed in section 3.3. Here we interpret the impacts of the various access metrics on the value of residential land, as well as the performance of the model. Finally, we address the conclusions in section 3.4 where we reflect on the limitations and implications of the results in further research and practice.

3.2 *Materials and methods*

3.2.1 *Case study*

Our case study city is Guatemala City, Guatemala. Similar to other Latin American cities with colonial origins, Guatemala's historic centre has a gridiron urban structure (Gellert, 1990; Pacione, 2005, pp. 447-602). First expansions by the end of 1800s and up to middle 1900s were carried out by local planners. Later, a combination of socio-political conditions, natural events and a massive migration from rural areas resulted in unplanned expansion towards the periphery. A CBD of white collars emerged towards the south of, and linked to, the historic core following important infrastructure such as the inter-American road (CA-1) and the international airport (figure 3.1). Jobs and important economic activities are highly centralized there, whilst population density tends to show a decentralized pattern (Municipality, 2009). The case study is relevant for the region as various Latin American cities display similar city structure and dynamics (Ford, 1996; Ingram & Carroll, 1981).

3.2.2 *Data pre-processing, variables and descriptive statistics*

Land-value data were collected during fieldwork (August 2014 - April 2015). We built a spatial database indexing 2000 records of real-estate property appraisals (observations) dating from the years 2008-2014

and carried out by a Guatemalan private office (AO). Observations were geo-referenced to the centroid of each property using the WGS84 co-ordinate system (Decker, 1986). According to the AO, the observations report “arm-length values” reflecting optimum transaction values where there is no pressure to sell or buy, and parties have complete information. Land-values are in local currency Quetzal (Q) per square meter of plot surface area. These were deflated to the year 2014 using the Guatemalan consumer price indexes (INE, 2016), and transformed to natural logarithms to deal with a non-normal distribution and potential non-linear relations with the predictors. We only used observations of residential uses (excluding flats) and plots with surface areas between 100 and 1,000 m². Figure 3.1 shows the location of the 1,026 observations used in this research.

Accessibility metrics were mostly produced in in chapter 2 and are aggregated in a hexagonal tessellation (Morales, Flacke, Morales & Zevenbergen, 2019a). The size of the hexagons (300 m inner diameter) provided an adequate resolution at an affordable computational demand. Geographic access metrics were available per transport mode. Additional variables, further described, were also mapped and aggregated to this tessellation (except property-level characteristics). Where one hexagon contained more than one observation, same predictor values were attached to the observations.

Table 3.1 shows descriptive statistics of the land-values and the predictors included in our model. Geographic accessibility metrics addressed location externalities that are assumed and empirically proven to influence residential markets in existing property-value literature (Des Rosiers et al., 2000; Du et al., 2012; Liu et al., 2010). According to local experts, access to these externalities is relevant to define location quality (Morales et al., 2019a). Access to neighbourhood-scale groceries, banks and restaurants, parks, schools, clinics and open markets were assumed to benefit a location within limited time radii and are of an “optional nature”, meaning that users appreciate close proximity and amount of available options. Thus, access to those was measured as a cumulative opportunity within a 10-minute travel time (Morales et al., 2019a). We assumed that a cumulative measurement would capture the benefit from accessible concentrations of such facilities, compared to the shortest travel-time to one of the facilities.

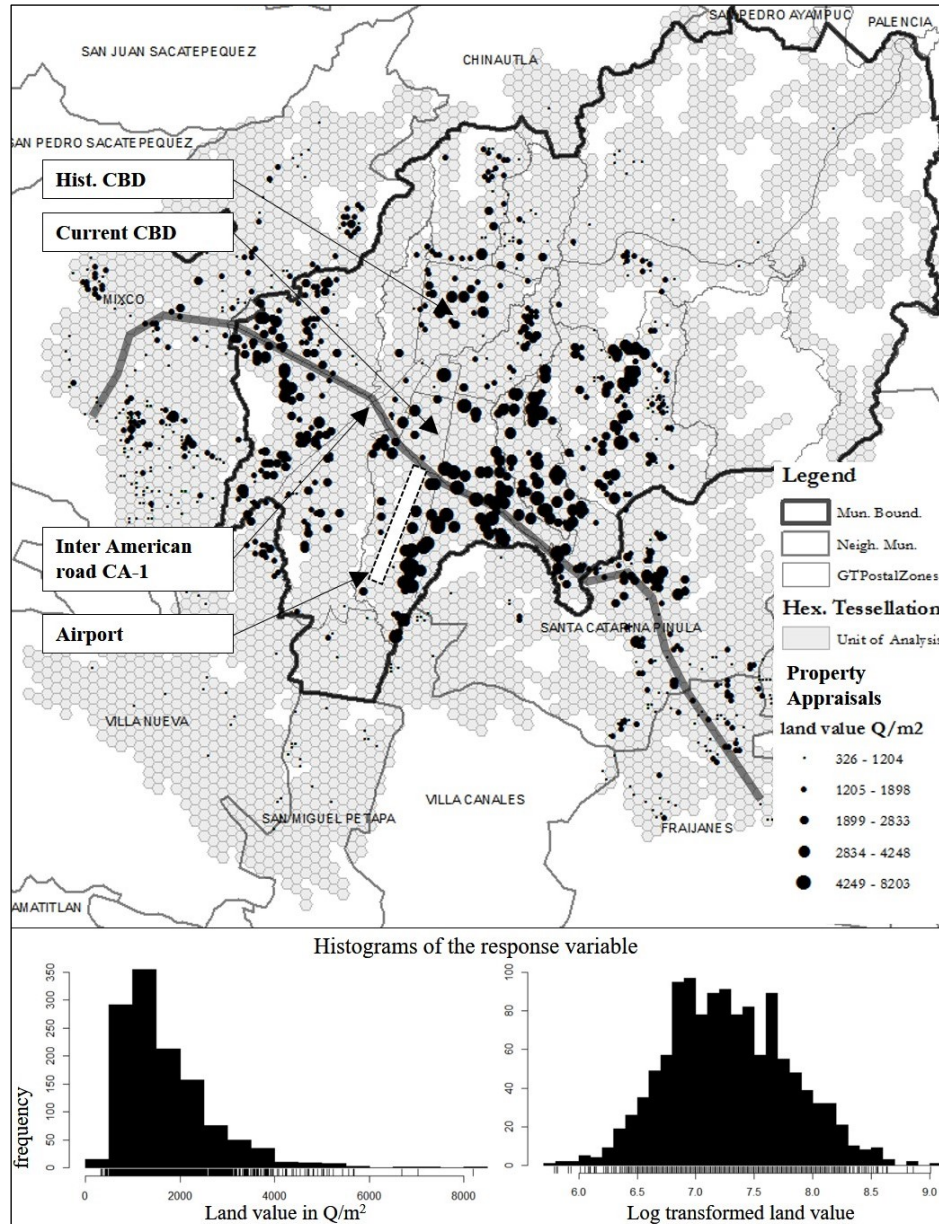


Figure 3.1: Appraisals location (top), and land-value frequency distributions before and after transformation (bottom).

Table 3.1: List of variables and descriptive statistics.

Group	Acronym	Description	Type	Mean	St. Deviation	Min	Max	% coded 0	% coded 1
Response variable	lv	Land value in Q/m ²	Ratio	1678.45	958.41	325.75	8203.43	□	□
Geographic accessibility by public transport	stand_groceries	Standardized access score (from 0-1) based on number of facilities accessible within 10 minutes driving	Ratio	0.08	0.14	0.00	0.80	□	□
	stand_bank_rest			0.04	0.09	0.00	0.94	□	□
	stand_parks			0.05	0.09	0.00	0.76	□	□
	stand_schools			0.03	0.10	0.00	1.00	□	□
	stand_clinics			0.06	0.13	0.00	1.00	□	□
	stand_markets			0.03	0.13	0.00	1.00	□	□
	time_cbd	Travel time to CBD	Ratio	27.21	9.74	2.13	51.59	□	□
	time_jobs	Gravity-based access to job opportunities		116375.94	50122.88	1244.84	240368.04	□	□
	time_xl_mall	Travel time to closest facility using public transport		39.50	18.56	3.00	109.00	□	□
	time_xl_grocery			29.08	15.26	2.00	93.00	□	□
	time_university			39.86	22.89	2.00	130.00	□	□
	time_culture			42.44	21.23	2.00	111.00	□	□
time_hospital	30.90		19.11	0.00	101.00	□	□		
time_xl_sports	50.20	24.12	4.00	136.00	□	□			
Geographic accessibility by private vehicle	s_groceries	Standardized access score (from 0-1) based on number of facilities accessible within 10 minutes driving	Ratio	0.24	0.20	0.00	0.91	□	□
	s_ban_rest			0.10	0.16	0.00	0.99	□	□
	s_parks			0.13	0.19	0.00	1.00	□	□
	s_schools			0.12	0.18	0.00	1.00	□	□
	s_clinics			0.17	0.21	0.00	0.95	□	□
	s_markets			0.10	0.18	0.00	1.00	□	□
	time_cbd	Travel time to CBD using private vehicle	Ratio	59.72	20.92	8.70	127.80	□	□
	time_jobs	Gravity-based access to job opportunities		48502.84	20746.17	9964.52	96362.19	□	□
	time_xl_mall	Driving time to closest facility		14.60	6.42	0.00	33.00	□	□
	time_xl_grocery			11.36	6.47	1.00	33.00	□	□
	time_university			17.02	9.65	1.00	48.00	□	□
	time_culture			19.26	9.69	0.00	42.00	□	□
time_hospital	12.56		8.24	0.00	41.00	□	□		
time_xl_sports	21.48	9.87	1.00	51.00	□	□			
Geometric accessibility (Space Syntax)	int_08	Average integRatio r_0.8Km	Ratio	49.95	35.86	6.37	272.05	□	□
	int_15	Average integRatio r_1.5Km		106.21	90.15	6.33	558.83	□	□
	int_25	Average integRatio r_2.5Km		206.21	191.25	8.80	1060.01	□	□
	int_50	Average integRatio r_5Km		524.75	475.18	58.28	2399.83	□	□
	int_75	Average integRatio r_7.5Km		922.39	754.09	135.55	3244.11	□	□
	int_n (global)	Average integRatio r_N		3063.56	863.52	1396.33	4807.18	□	□
	nach_08	Maximum normalised choice r_0.8Km		1.26	0.18	0.00	1.56	□	□
	nach_15	Maximum normalised choice r_1.5Km		1.23	0.18	0.00	1.53	□	□
	nach_25	Maximum normalised choice r_2.5Km		1.21	0.18	0.00	1.44	□	□
	nach_50	Maximum normalised choice r_5Km		1.16	0.18	0.00	1.43	□	□
	nach_75	Maximum normalised choice r_7.5Km		1.14	0.19	0.00	1.44	□	□
	nach_n	Maximum normalised choice r_N		1.10	0.19	0.00	1.49	□	□
Proximity to infrastructure	dist_mroad	Planar distance to main roads as classified in OSM, expressed in meters	Ratio	333.11	313.19	0.19	1492.12	□	□
	dum_prox_bus	1 if location is within 500mt distance to any bus line	Dummy	□	□	0	1	33%	67%
Sumbarkets	dum_west	1 if area is within the west municipalities	Dummy	□	□	0	1	78%	22%
	dum_east	1 if area is within the east municipalities		□	□	0	1	84%	16%
	condo_seg	Average total selling price of horizontal housing offer classified from 0 to 3 accord.		1.83	1.11	0.00	3.00	□	□
	flat_seg	Average total selling price of vertical housing offer classified from 0 to 3 accord.	Ratio	0.84	1.07	0.00	3.00	□	□
Neighbourhood characteristics	pop_dens	Population density projected to 2015*	Ratio	46.83	21.62	3.38	98.26	□	□
	soc_economic	Predominant socio economic level (from 1-5) per census block		4.28	0.97	1.00	5.00	□	□
	percent_priv	Percentage of private-vehicle-based generated trips per TAZ		0.22	0.18	0.03	0.89	□	□
	dens_flats	Density of recent horizontal housing offer (2013-2014) expressed in units/km ²	Ratio	4785.69	7024.00	0.00	36605.64	□	□
	dens_condos	Density of recent vertical housing offer (2013-2014) expressed in units/km ²		5367.09	4381.67	0.00	15915.49	□	□
Plot characteristics	year	Year when the property was appraised	Ratio	6.46	1.49	4.10	8.98	□	□
	plot_area	Plot surface area in m ² without construction		277.24	182.10	100.29	997.00	□	□
	const_area	Total construction area in m ² without plot surface area		251.85	111.94	100.48	784.00	□	□
	dum_geometry	1 if plot geometry is rectangular, else 0	Dummy	□	□	0	1	26%	74%
	dum_intrablock	1 if plot is located on the corner of the block		□	□	0	1	86%	14%
	pot	Value indicates building potential according to current policy	Ordinal	3.25	1.08	0	5	□	□
**Grev highlighted cell indicates measurements that were transformed into natural logarithms (ln)									

**Grey highlighted cell indicates measurements that were transformed into natural logarithms (nl)

The CBD, jobs location, large-scale malls, large-scale groceries, universities, culture, hospitals and large-scale sport facilities are destinations for which people are more willing to overcome impedance,

and influence location quality at a city scale. Access was measured as shortest travel-times, except for jobs access (Morales et al., 2019a). This was mapped as potential access using the Hansen (1959) formulation. Decay parameters (α, β) were estimated using an origin-destination (OD) matrix containing trip volumes of traffic analysis zones (TAZ) per transport mode during morning peak-hours from year 2005. Travel-times were estimated for each OD pair. Then, trip volumes were ranked by travel-time and a decay function was fitted allowing to identify the optimum parameters for each transport mode. Values of trips attraction were calculated at the hexagon level. These were used as proxy values to indicate accessible jobs at each origin (hexagon centroids).

Geometric access includes SSx metrics, integration and choice, at various spatial scales (radii). High geometric access at lower radii (e.g. r_0.8 km) benefit areas associated with pedestrian flows and important streets at a neighbourhood level. Access at higher or global (r_N) radii analyses benefit areas associated with important economic activity and reveal a roads hierarchy at a city-wide scale. Proximity to mobility infrastructure includes distance to main roads and whether or not a location is within a threshold distance of access points to a bus line. Following Bourassa et al. (2007) we used submarket variables to address auto-correlated errors. Auto-correlated errors indicate the presence of a spatial structure due to nearby observations sharing similar locational and unmeasured externalities (Basu & Thibodeau, 1998; LeSage et al., 2009). Submarket classifications were mapped based on location (east or west periphery) and using expert knowledge from the AO. Submarket predictors classify the urban areas in reference to the selling-price segmentation of available new residential projects. Thus, indicating the non-existence of projects in an area (0), and the incremental selling-price coded from 1 to 3.

Neighbourhood characteristics include the following information: population density projected to 2015 (World_Pop, 2014); classification of socio economic groups per census track (URBANISTICA, 2009); percentage of private mobility users per TAZ (percent_priv); and density of new residential projects. The "percent_priv" was estimated as the proportion of trips generated by private vehicles, from the total of trips generated per TAZ. Densities of new residential projects (dens_flats, dens_condos) were assumed to be able to capture popularity and demand for residential location. Geo-referenced locations of projects dating 2013 and 2014, provided by the AO, were processed using a point-density analysis at a 2 km radius (average neighbourhood size).

The year variable aimed to capture any temporal trend that was independent to monetary inflation. The plot surface area aimed to capture the relation where appraised value per m² decreases with increasing surface area (Lin & Evans, 2000). Construction area was a proxy variable included to isolate the land-value increase due to residence structural characteristics. We included dummy variables indicating whether the plot had a regular geometry or not, and the property intra-block location. These are locally used as appraisal “adjustment factors” (DICABI, 2005). The POT variable, ranging from 1 to 5, attempted to capture the relation between the land-value with potential for future development into non-residential uses and/or larger constructions following current Territorial Ordinance Plan (Municipality, 2009). Although it only applies for properties within Guatemala municipal boundaries, we classified the remaining plots using the same guiding principles.

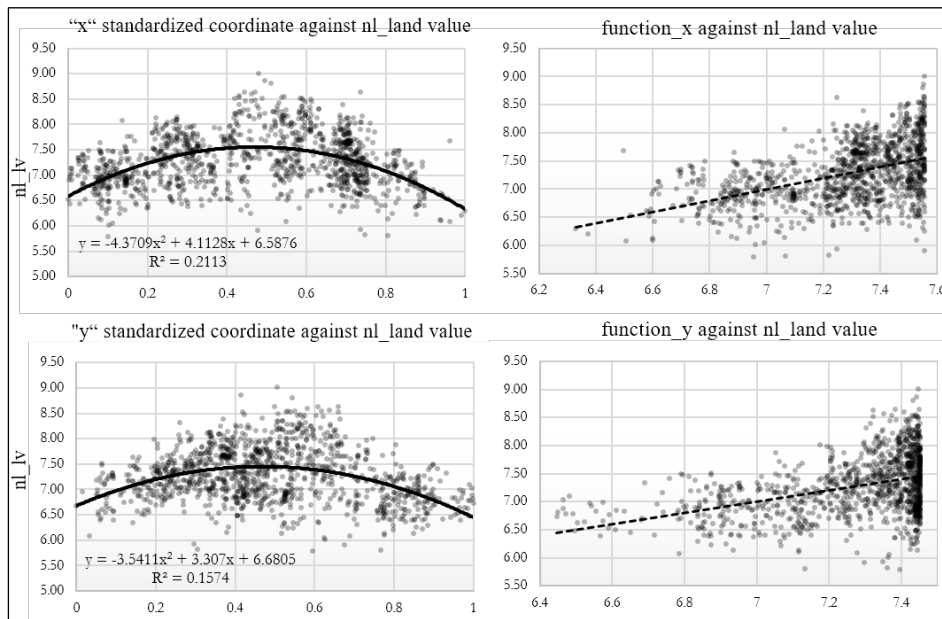


Figure 3.2: Scatterplots of x and y co-ordinates against the nl_land-values (left side); and scatterplots of estimated f_x and f_y plotted against nl_land-values (right side).

Observation co-ordinates were used as predictors, also following Bourassa et al. (2007). Left side of Figure 3.2 shows initial scatterplots of each “x” and “y” co-ordinate at the upper and bottom areas correspondingly. Standardized co-ordinate values on the horizontal axis are plotted against land-values on the vertical axis. By fitting a quadratic function, the concave trends on both co-ordinates (x and y)

reflect that the highest land-values are in the city core (matching CBD location) and these decrease as the co-ordinates move towards east or west (x co-ordinate), and south or north (y co-ordinate). Such trends were expected to be captured by the accessibility variables. Yet, we kept the co-ordinates as predictors to either capture any remaining trends or to reduce auto-correlated errors. Using the quadratic functions, we estimated new values to use as predictors (f_x and f_y) to remove the non-linearity. On the right side of the figure are the corresponding scatterplots.

3.2.3 Methods

We used multivariate regression as a commonly accepted method in this research area, particularly in hedonic pricing studies. The regression model included the following access metrics: (1) geographic access indexes were computed based on travel-times per transport mode to various land uses; (2) geometric access incorporated SSx network metrics at various spatial scales; (3) we proposed a *geometric via geographic-access* metric formulated as the potential access to network integration (as measured in SSx). Additionally, other relevant variables were also included, such as submarket information, neighbourhood and property-level information. Following an initial correlation exploration, the model was first fully specified. Then a variable selection process was implemented to assess the relevance of the information added by each variable and deduce the most parsimonious model. Finally, the model was assessed using performance statistics contrasted with *alternative* model versions representing common approaches in land-value modelling literature.

Geographic access indexes per facility were estimated using equation (3.1). Computations were done at the tessellation level for the entire city area. Access measurements per transport mode were all standardized to ranges from 0 (poor access) to 1 (high access). Then, *combined_access* to facilities at location i is given by the standardized access value by private vehicle times *percent_priv*, plus the equivalent formulation for access by public transport. Unlike (Adair et al., 2000), our index was adjusted to the variation of the transport-mode predominance across the city. This variation is given by the *percent_priv* variable (see previous section) and represents the changing proportions of modal split aggregated to the hexagonal level. We assumed that areas where residents predominantly use public transport, public transport-based access was locally more important compared to private-based access and vice-versa.

$$combined_access_i = (A_{priv_i} * perc_priv_i) + (A_{pub_i} * (100 - perc_priv_i)) \quad (3.1)$$

Geometric via geographic access was estimated as *integration_gravity* per hexagon i . We used equation (3.2), which is a potential access formulation (Hansen, 1959) where the attraction size is given by the average global integration (r_N) at any reachable hexagon, hex_j , where n is equal to the number of segments in the system, and D_θ is the angular *depth* between a x segment and any other segment in the network, i (Hillier et al., 2012).

$$integration_gravity_{hex\ i} = \sum \left(\frac{\sum_{i=1}^n \left(\frac{n-1}{D_\theta(x,i)} \right)_{hex\ j}}{N_{seg\ at\ hex\ j}} \right) \alpha \exp(-\beta * t_{hex\ i-hex\ j}) \quad (3.2)$$

We used global integration as it had the highest positive correlation with land-values. Decay parameters, α and β , exponentially penalize attraction size based on travel-time (t) between origin ($hex\ i$) and reachable destinations ($hex\ j$). The parameters applied in this computation were the same as used in Morales et al. (2019a) to estimate job access: $\alpha = 1.2$ for private and 1.3 for public transport; and $\beta = 0.052$ and 0.02 correspondingly. Although, this leads to potentially high collinearity with job access, it was still expected that *integration_gravity* would add additional information at a higher level of resolution. Figure 3.3 show the results of implementing this metric.

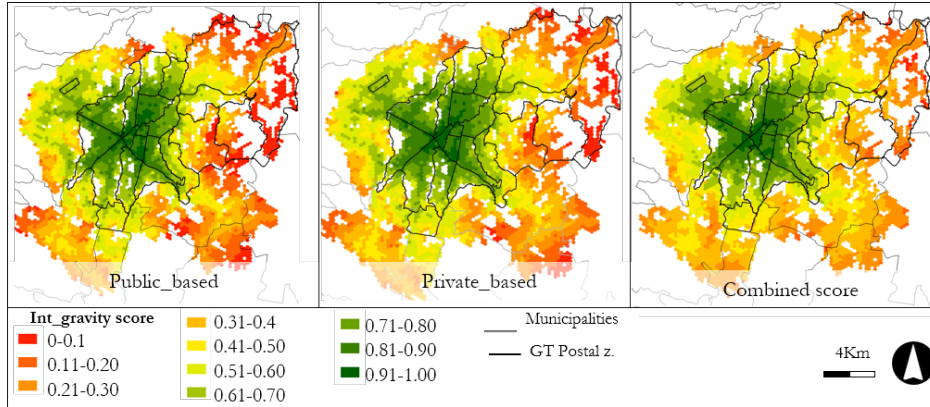


Figure 3.3: Geometric via geographic-accessibility from low (red) to high (green).

The regression model assumes that land-values can be decomposed into numerical contributions from the locational predictor attributes (Liu et al., 2010). A generalized specification follows equation (3.3). Where the land-value (nl_{lv}) of a property is estimated as a function of a constant intercept β_0 , plus the summation of the products between

estimated coefficients (β_i) with the measured attributes X , plus a random error ε .

$$nl_{lv} = \beta_0 + \sum \beta_n * X_p + \varepsilon \quad (3.3)$$

Our model takes the form of equation (3.4). Where the land-value predictors are grouped into geographic access, geometric access, geometric via geographic access, infrastructure proximity, submarket, neighbourhood and plot-level characteristics. The model was fitted via ordinary least squares in R software (Team, 2016) using 879 randomly selected observations (75%) to train the model, and the remaining observations for cross validation.

$$nl_{lv} = \beta_0 + (\sum_{Geograp} \beta_i * X_i) + (\sum_{Geomet} \beta_i * X_i) + (\beta_i * X_{Geomet\ via\ Geograp}) + (\sum_{Infra} \beta_i * X_i) + (\sum_{Sub-Mark} \beta_i * X_i) + (\sum_{Neigh} \beta_i * X_i) + (\sum_{Plot} \beta_i * X_i) + \varepsilon \quad (3.4)$$

It was expected that not every predictor would add meaningful information to the model and that access variables would introduce some multicollinearity, which makes almost impossible to discriminate predictors based on their coefficients and significance. Therefore, we applied an automated bi-directional stepwise regression using the MASS package (Ripley et al., 2016, pp. 143-144), where variables are retained or discarded based on their contribution to the Akaike Information Criterion (AIC) statistic (Bozdogan, 1987). For the final reduced model we used the Breusch-Pagan test and White's correction to correspondingly detect heteroskedasticity and adjust the coefficients and the Variation Inflation Factor (VIF) to detect multicollinearity (Fox & Weisberg, 2016).

The estimated coefficients represent the impact of each variable on the land-value. However, those are not comparable among each other as they are scale-dependant. First, not all the predictors are expressed in the same units (e.g. logarithmic scales versus standardized access-scores). Second, the predictor values come from a systematic mapping for the whole city area. So, the value ranges differ even for predictors expressed in the same units (table 3.1). Therefore, standardized beta-coefficients (Fletcher, 2015) are also reported.

We used the following performance statistics to assess the models: adjusted R^2 , AIC, semivariogram modelling to assess the presence of auto-correlated errors (Pebesma & Graeler, 2016), the root mean squared error over prediction (RMSEP) of test-data and the normalized RMSEP to compare the proportional difference of error variance among

models. The RMSEP used the difference between the maximum and minimum observed values of the test-data (2.82).

Semivariogram models assume that the similarity between any pair of residuals is inversely proportional to the lag (distance) between them. Dissimilarity is estimated as gamma values. Three parameters can be read from the semivariogram: (1) the nugget is the value of gamma at lag=0 and associated to statistical noise; (2) the partial sill is the difference between the nugget and the value of gamma when auto-correlation becomes negligible; (3) the range indicates the distance between observations where auto-correlation becomes negligible and the partial sill is reached. The variograms fitted here are exponential and assume isotropy, meaning that the spatial dependence at a given lag is constant throughout the area and that the resulting range is effective up to three times its distance.

The model was assessed in comparison to six *alternative* models, which were formulated as variations of equation 3.4 and estimated using the same approach. First and second are model variations where corresponding access predictors only address one of the transport modes. The third model is as the original but excludes all of the geometric-access predictors. The fourth considers exclusively a mono-centric assumption and excludes potential jobs access. In turn, the fifth model replaces CBD access with potential job access, relaxing the mono-centric assumption. The sixth model is a variation that excludes all of the submarket predictors and the observations co-ordinates.

3.3 Results and discussion

3.3.1 Accessibility and land-values

Figure 3.4 shows pairwise correlations between the any access metric relying on transport modes (geographic access and geometric via geographic access) with the land-values. Correlations of the private and public transport-based access metrics were included for reference purposes. A label indicates in which definition each metric (travel time, cumulative opportunity or potential access) was originally measured.

The strengths of the correlations for each variable are similar for both transport modes except for the first five land uses. Private mobility strongly increases access opportunities to concentrations of such land uses. This is reflected in higher pairwise correlations with higher land-values in contrast with access by public mobility. This comparison seems applicable to all the variables considered. However, cumulative -opportunity metrics, aiming to analyse quality location at a neighbourhood scale, made the mobility disparity more evident.

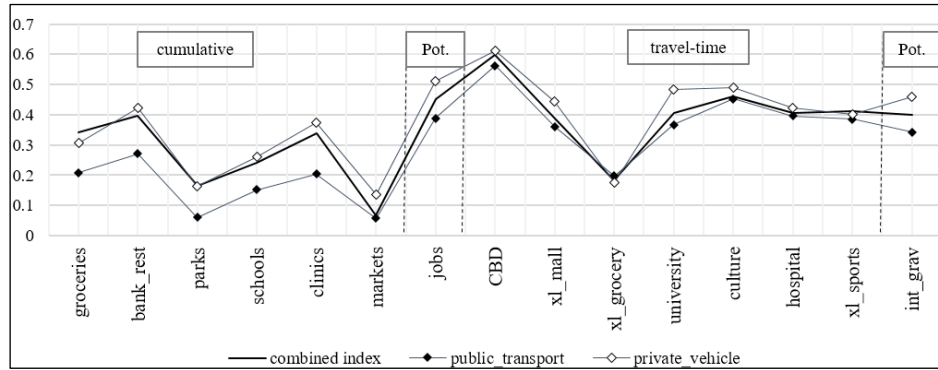


Figure 3.4: Pearson correlations between accessibility metrics and log of land value.

Correlation strengths of the combined indices fall between the correlation strengths by transport mode. This was expected as the indices incorporate and reflect the difference on the strength of the pairwise correlations by transport mode, previously discussed. However, the combined indices are sensitive to spatial variations of the access benefits derived from simultaneous availability of transport modes as well as to the residents' predominant preference for those. This was expected to contribute with additional information to the model, compared to only addressing one transport mode.

Importance of a monocentric land-value structure is preliminarily detected as the strongest correlation is observed with CBD access (~ 0.6). Access to jobs, geometric via geographic access (integration_gravity), universities, cultural institutions and large-scale malls reach correlations up to ~ 0.5 . Then, access to banks and restaurants, clinics, hospitals and large-scale sports infrastructure reach correlations up to 0.4.

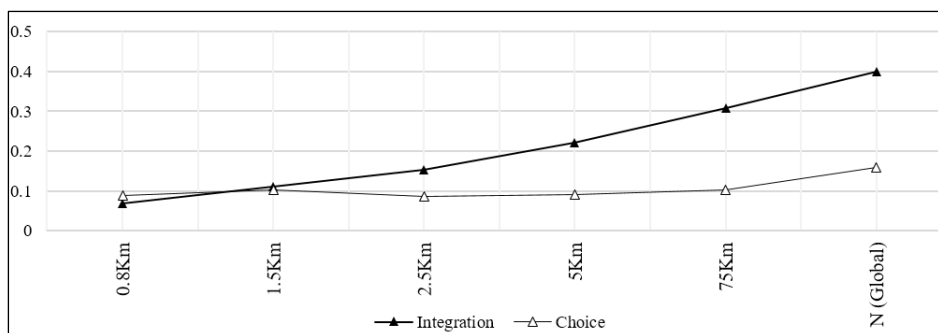


Figure 3.5: Pearson correlations between geometric accessibility and log of land value.

Figure 3.5 shows correlations of the geometric access metrics that were computed via SSx. Geometric access metrics estimated via SSx, all show positive correlations. Those increase when increasing the radii of analysis (0.8 km-Global), from 0.09 to 0.16 for the choice values and from 0.07 to 0.4 for the integration values. This means that, at first glance, geometric-access is a quality-location that tends to be more important when indicating relative location to the city as a whole and lesser at the neighbourhood scale. Also, compact grids with small blocks are less favoured than longer blocks.

Table 3.2 shows model results after the variable selection process. Only relevant variables were kept in the model based on their information contribution, reflected through the AIC statistic. Variables expressed in logarithmic form are interpreted as the percentage change in land-values per 1% change in the predictor value. This is also applicable to coefficients of indices ranging from 0 to 1 (i.e. geographic and geometric via geographic access). Other coefficients are interpreted as the percentage change in land-values given 1-unit change in the predictor.

From the cumulative access metrics, only access to banks and restaurants and schools were retained. Both have a positive impact of 0.80% and 0.42% per 1% increase of access to those facilities. A 1% increment in CBD access has a positive impact of 2.53% on the value of land. Meaning that land-values at the CBD are ~253% more expensive than those at periphery. Job access has an unexpected negative sign. Values decline 1.26% per 1% increase in the access index. We tested the sign reversal by manually removing, one by one, the following predictors that showed multicollinearity and correlation with job access: *integration_gravity*, *global integration*, *CBD* and *culture access*. Nevertheless, we found the sign remained negative. We hypothesise that in our case study when adding job access to the model simultaneously with other predictors, the index might be actually capturing the negative effects of urban centrality (e.g. pollution and congestion). Particularly when considering that this metric uses the number of trips attracted to an area as a proxy to availability of jobs.

Land-value increases by 1.16% with a 1% increase in access to large-scale malls. In turn, a 1% increase in access to large-scale grocery shops means a decline of 1.17%. The comparison between these two coefficients could be revealing that the attractiveness of agglomerated commerce and recreational activities is the one that overweighs the potential negative impacts (noise and congestion) that the vicinity of either facility poses. Furthermore, a 1% increase in the access score of

the following externalities implies an increase of land-values: universities (1.21%) and hospitals (1.21%). Unexpectedly, access to cultural facilities has a negative effect of a 1.87% decrease in land-values. In Guatemala City cultural facilities are mostly located in the core and historic areas of the city. Negative externalities such as pollution, congestion, street robbery, and informal commerce are common in these areas, which could explain the positive association with the land-values in the pairwise correlation, but a negative sign when placed jointly with other predictors.

As in Chiaradia et al. (2009), SSx integration has contrasting effects according to the analysed spatial scale. An increase of 1% in integration at lower radii (0.8 km and 2.5 km) implies a decline of land-values of 0.09% and 0.12% correspondingly. In Guatemala City most of the areas highlighted at low radii integration correspond to deteriorated single family neighbourhoods. These are areas with compact gridiron street patterns built during the first planned expansions. Middle-low and low-income groups predominantly reside in these areas, vulnerable to pollution and street robbery. In turn, the effect of integration is positive at a city-wide scale (global). Land-values increase by 0.28% per 1% increase of global integration. SSx choice at 1.5 km has a positive impact of 0.38%. Neither distance to main roads nor normalized choice at higher radii were selected. From the results we interpret that residential land-values benefit only when located nearby streets that are important at a neighbourhood scale, even though, very often, such streets could also be part of or connected to city-structuring roads outlined by choice analyses at higher radii.

Geometric via geographic access has a positive influence of 0.62% per 1% increase in the access index, implying that quality locations benefit from geometric access not only at location, but also as a reachable resource. This metric might be capturing access to vital urban areas, or with potential for doing so, containing economic activities not addressed explicitly by the other predictors.

The strongest multicollinearity (VIF ~50) was observed with jobs and geometric via geographic access, meaning, as expected, there is a strong correlation between them and implying that their interpretation should not be considered without caution. However, we are able to claim that each predictor still adds complementary information to the model following their impact on the AIC statistic. This is further discussed in the model assessment.

Observations within a 300m distance of the bus lines network showed a decline of 0.11%. This was expected as previous access predictors

address the true benefit of public transport, while the mere physical proximity to the network itself might be associated with negative externalities such as noise and pollution.

Table 3.2: Regression coefficients and normalized coefficients. Grey colour bars on the normalized coefficients indicate relative importance.

Predictors		Coeff.	Std. error	t	Significance	VIF	n_coeff.
Intercept		-0.32	1.09	-0.28	0.77		
Geographic accessibility	access_groceries	0	0	0	0	0	0
	access_ban_res	0.80	0.22	3.62	0.00 ***	3.6	0.15
	access_parks	0	0	0	0	0	0
	access_schools	0.42	0.17	2.38	0.02 *	2.8	0.09
	access_clinics	0	0	0	0	0	0
	access_markets	0	0	0	0	0	0
	access_cbd	2.53	0.45	5.55	0.00 ***	19.0	0.47
	access_jobs	-1.26	0.29	-4.42	0.00 ***	48.7	0.50
	access_xl_mall	1.16	0.24	4.75	0.00 ***	7.2	0.23
	access_xl_grocerie	-1.17	0.20	-5.85	0.00 ***	4.0	0.20
	access_university	1.21	0.22	5.50	0.00 ***	7.7	0.26
	access_culture	-1.87	0.38	-4.87	0.00 ***	21.0	0.40
access_hospital	1.21	0.24	5.01	0.00 ***	7.4	0.24	
access_xl_sport	0	0	0	0	0	0	
Geometric Accessibility	nl_int_08	-0.09	0.03	-2.80	0.01 **	4.1	0.10
	nl_int_15	0	0	0	0	0	0
	nl_int_25	-0.12	0.03	-3.77	0.00 ***	7.2	0.18
	nl_int_50	0	0	0	0	0	0
	nl_int_75	0	0	0	0	0	0
	nl_int_n (global)	0.28	0.11	2.51	0.01 *	11.9	0.16
	nl_nach_08	0	0	0	0	0	0
	nl_nach_15	0.38	0.13	3.00	0.00 **	1.5	0.07
	nl_nach_25	0	0	0	0	0	0
	nl_nach_50	0	0	0	0	0	0
nl_nach_75	0	0	0	0	0	0	
nl_nach_n	0	0	0	0	0	0	
Geomet. via	access_int_grav	0.62	0.30	2.07	0.04 *	47.1	0.26
Proximity to infrastructure	dist_mroad	0	0	0	0	0	0
	dum_prox_bus	-0.11	0.02	-4.22	0.00 ***	2.0	0.10
Submarket variables	dum_east	0.25	0.08	3.23	0.00 **	9.2	0.19
	dum_west	-0.10	0.04	-2.67	0.01 **	3.3	0.07
	condo_seg	0.07	0.02	3.95	0.00 ***	4.2	0.16
	flat_seg	0.09	0.02	5.01	0.00 ***	4.1	0.20
Neighbourhood Characteristics	nl_pop_dens	0.12	0.03	3.78	0.00 ***	6.4	0.16
	soc_economic	0.04	0.01	3.49	0.00 ***	1.3	0.07
	perc_private	0.33	0.09	3.74	0.00 ***	2.4	0.11
	nl_dens_flats	0.01	0.00	3.59	0.00 ***	2.3	0.09
	nl_dens_condos	0.01	0.01	2.09	0.04 *	2.2	0.06
Plot Characteristics	year	0.09	0.01	17.46	0.00 ***	1.1	0.84
	nl_plot_ar	-0.17	0.03	-6.15	0.00 ***	2.1	0.19
	nl_const_ar	0.14	0.04	3.68	0.00 ***	2.0	0.11
	dum_geometry	0	0	0	0	0	0
	dum_intrablock	0	0	0	0	0	0
	pot	0	0	0	0	0	0
	f_x	0.29	0.07	3.82	0.00 ***	4.4	0.13
f_y	0	0	0	0	0	0	
VIF: Variation inflation factor							
Significance: '****' 0.001 '***' 0.01 '**' 0.05							

We observe a difference between the sub-market variables indicating east and west peripheral municipalities. Assuming that the rest of the variables are held constant, residential land on the eastern periphery is more expensive by 0.25% compared to the rest of the observations, whereas land is 0.10% cheaper in the western peripheral areas. For the average plot, this means a difference of +Q445/m² (\$60) and – Q115/m² (\$15) respectively. Density of new condominiums predominates in both peripheral areas, particularly at the west. However, selling price segmentation clearly points out a more expensive market at the peripheral east side. Selling-price segmentation of residential projects (horizontal and vertical) have both a positive impact of 0.07% and 0.09% on the land-value.

Neighbourhood characteristics all have a positive contribution. An increase of 1% in the population density represents higher land-values by 0.12%. The predominance of different socio-economic groups, incremental from 1 (low incomes) to 5 (high incomes), is associated with a land-value increase of 0.43%. This is equivalent to a difference of Q345/m² (~\$45) for an average plot located in an area classified as 1 or 5. Furthermore, an increase of 1% in the proportion of private vehicle users in an area is associated with a land-value increase of 0.33%. Furthermore, an increase of 1% in density of new vertical and horizontal residential projects has a positive influence of 0.01%. This is equivalent to a difference of ~5-7 new residential projects in a radius of 2 km.

The coefficient of the year predictor reveals that in these data there is an increase in the land-values of 0.09% every year. Such value increase over time is additional to money inflation. This variable could be reflecting the influence of dynamics that are not captured by any other predictor. For example, urban transformations along time (e.g. city expansion, redevelopment) and overall macroeconomic growth. The average annual growth of the gross domestic product (GDP) per capita was 0.93% between 2008 and 2014 (WorldBank, 2017). Furthermore nl_land-value per m² is appraised 0.17% lower per 1% increase in the plot surface area. In turn, a 1% increase in built-up areas has a positive impact of 0.14%. Lastly, the co-ordinate function f_x reflects an increase of 0.29% per one-unit increase, meaning that there is a value increase trend along the east-west direction additional to the access benefits and any other location characteristic addressed by the predictors included.

Finally, the normalized coefficients can be interpreted as changes in standard deviations of the land-values per one standard deviation change in the predictor variable. These coefficients are comparable.

Access to CBD was the most important variable that positively influenced the residential land-values in Guatemala City, thus confirming a mono-centric land-value structure. It is followed by the year predictor, outlining the relevance of land-value gains over the years. Interestingly, geometric via geographic accessibility becomes the third most important predictor that capitalizes residential land, then followed by access to universities, malls and hospitals.

3.3.2 Model performance and diagnostics

Table 3.3 shows a summary of the performance statistics of our original model, contrasted with the alternative models. Those represent existing land-value modelling approaches. The upper part of the table indicates the predictors that were excluded from the initial model formulations, and the predictors kept in each model after the variable selection procedure.

The original model was found to have the highest performance of all the models. It has the highest goodness of fit over the train data explaining up to 72.8% of the land-values variability. It has the lowest AIC value (168), meaning it is the most parsimonious model and closer to reflect the process that generates the land-value observations. Comparison of the AICs reflects an improvement of 30% and 15% over a model based only on public (1st) and private (2nd) mobility correspondingly. Furthermore, it improves by 21% over a model that excludes geometric access information (3rd). We do not observe major differences (up to 4%) in the adjusted R^2 across the models. However, judging by the AIC comparisons, there is an important improvement in the land-value modelling approach when addressing more than one transport mode.

By comparing the original model against the 4th and 5th alternatives, we determined that together both CBD and “jobs” access contribute to a better model. It is clear that CBD access is a dominant predictor when considering its estimated coefficient (table 3.2) and observing its effect on the AIC statistic when excluded in the 5th model, while some caution should be taken when interpreting what is being revealed by “jobs access” given its sign reversal and its multicollinearity with the CBD and the geometric via geographic-access itself.

Furthermore, it seems that addressing geometric accessibility does not only improve the goodness of fit, but also reduces the spatial dependence of the regression residuals by almost 200 m, meaning that less variability in the land-value observations is left unexplained and indicating that geometric access predictors add spatialized information at a localized scale. Such information is comparable to the information

that is contributed to the model by the submarket predictors and the observation co-ordinates. This is also deduced from observing the AIC statistic of the 6th model, and the increase of geometric access predictors being kept in this model.

Table 3.3: Model performance in contrast with alternative models.

Predictors		Original	1. Public	2. Private	3. Geographic	4. CBD	5. Jobs	6. Withouth Sub-Markets
Geographic accessibility	access_groceries	✓	✓	✓		✓	✓	✓
	access_ban_res							
	access_parks							
	access_schools	✓	✓	✓			✓	✓
	access_clinics				✓			✓
	access_markets			✓			✓	
	access_cbd	✓	✓	✓	✓	☐	☐	✓
	access_jobs	✓	✓		✓	☐		
	access_xl_mall	✓	✓	✓	✓	✓	✓	✓
	access_xl_grocerie	✓	✓	✓	✓	✓	✓	✓
Geometric Accessibility	access_university	✓	✓	✓	✓	✓	✓	✓
	access_culture	✓	✓	✓	✓	✓	✓	✓
	access_hospital	✓	✓	✓	✓	✓	✓	✓
	access_xl_sport							
	nl_int_08	✓	✓	✓	☐	✓	✓	✓
	nl_int_15				☐			
	nl_int_25	✓	✓	✓	☐	✓	✓	
	nl_int_50				☐			✓
	nl_int_75				☐			
	nl_int_n (global)	✓	✓	✓	☐	✓	✓	✓
Geomet. via Geog.	nl_nach_08				☐			
	nl_nach_15	✓	✓	✓	☐	✓	✓	✓
	nl_nach_25				☐			
	nl_nach_50				☐			✓
	nl_nach_75				☐			
	nl_nach_n		✓		☐		✓	
	access_int_grav	✓	✓	✓	☐		✓	✓
	dist_mroad							
	dum_prox_bus	✓	✓	✓	✓	✓	✓	✓
	dum_east	✓	✓	✓	✓	✓	✓	☐
Submarket variables	dum_west	✓	✓		✓	✓	✓	☐
	condo_seg	✓	✓	✓	✓	✓	✓	☐
	flat_seg	✓	✓	✓	✓	✓	✓	☐
	nl_pop_dens	✓	✓	✓	✓	✓	✓	✓
Neighbourhood Characteristics	soc_economic	✓	✓	✓	✓	✓	✓	✓
	percent_private	✓	✓	✓	✓	✓	✓	✓
	nl_dens_flats	✓	✓	✓	✓	✓	✓	✓
	nl_dens_condos	✓	✓	✓	✓	✓	✓	✓
Plot Characteristics	year	✓	✓	✓	✓	✓	✓	✓
	nl_plot_ar	✓	✓	✓	✓	✓	✓	✓
	nl_const_ar	✓	✓	✓	✓	✓	✓	✓
	dum_geometry		✓			✓	✓	
	dum_intrablock							
	pot							
	f_x	✓	✓	✓	✓	✓	✓	☐
Goodnes of fit	f_y		✓	✓			✓	☐
	Adjusted R ²	0.728	0.704	0.718	0.711	0.723	0.720	0.687
Auto correlated errors	AIC	167	243	199	213	180	192	274
	Nugget	0.0055	0.0053	0.0059	0.0059	0.0053	0.0049	0.0060
	Partial Sill	0.0611	0.0671	0.0627	0.0675	0.0625	0.0624	0.0678
	Effective range (m)	741	763	771	938	734	703	942
Cross validation	RMSEP	0.3177	0.3184	0.3183	0.3211	0.3203	0.3136	0.3373
	N_RMSEP	0.1127	0.1129	0.1129	0.1139	0.1136	0.1112	0.1196

☐ Predictor excluded ✓ Predictor selected

The cross-validation with the test data shows that variations of prediction accuracy among models are modest. The original model shows the lower RMSEP (0.3150), which is equivalent to 1.37 units in local currency. The N_RMSEP shows that the standard deviation of the prediction errors for all the models remains relatively the same. These statistics were somehow unexpected. However, even at this minor scale, the highest differences are observed when excluding geometric-access and the submarket predictors.

Finally, figure 3.6 shows the regression diagnostics. From left to right, we observe that the dispersion of the residuals follows an approximately constant variance. The QQ plot shows that prediction errors closely follow a normal distribution. The plot on the right shows that errors are mostly centred in 0, with a slight tendency to overestimate extremely low values and underestimate extremely high values. This indicates a slight non-linearity behaviour in the data. However, by visually judging this figure and considering the magnitude of the errors in the cross-validation, we determined that such behaviour has minor effects in the performance of the model.

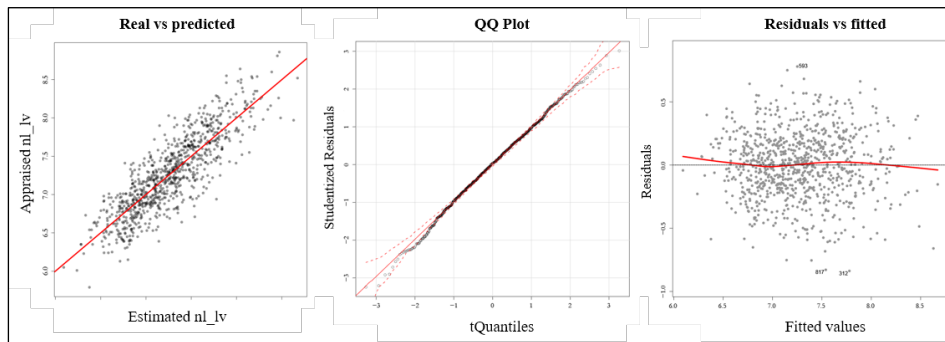


Figure 3.6: Diagnostics of model residuals.

3.4 Conclusions

We implemented a comprehensive operationalization of accessibility indicators to develop a residential land-value model in Guatemala City. The model incorporated metrics of geographic, geometric and geometric via geographic access. Those were computed as indices that incorporate the variation of the access disparity due to current availability of transport modes and adjusted to the predominant use of each mode across the city.

Following our first hypothesis, land-value research would benefit from considering more comprehensive definitions of accessibility. Integrating the available transport modes into the model modestly improved the goodness of fit to the observed data (adjusted R^2) and the prediction accuracy of non-observed data. Nevertheless, it yielded a model that better represents the true process generating the observations (AIC), compared to similar models addressing access by only one transport mode. Furthermore, geometric accessibility brings spatialized and localized information that contributes to a parsimonious model and better explains the variability of land-values. This was deduced from observing a reduction of the spatial dependence in the model residuals.

Following our second hypothesis, we conclude that geometric accessibility does capitalize land at location (to some extent) and as a reachable resource. Similar to the results of Chiaradia et al. (2009), integration (network closeness), analysed at a localized radius of 0.8-1.5 km, has a negative effect on residential land-values. In Guatemala City, local integration outlines areas with compact gridiron layouts in old neighbourhoods. In turn, relative closeness to anywhere in the city in areas with larger blocks seems to capitalize land-values. Only hierarchy of roads on the neighbourhood scale, as analysed by SSx choice (network betweenness), has a positive impact on residential land-values. Furthermore, we determined that the potential reachability of geometric access by available transport modes is positively associated with the value of land. We consider this metric as complementary, as it could reflect access to economically vital areas (or with the potential for such) not addressed explicitly in other access metrics.

From the model results, we found that land-values in Guatemala City differ up to 253% following CBD access, which was the strongest predictor. Thus, we conclude that there is a predominance of a mono-centric land-value structure, which strongly reflects centralized economic activities, as described by Ford (1996). Access to jobs was the most important negative externality. We assume that when other predictors better explain such opportunities, this metric reveals the negative effects of urban centrality (pollution and congestion), also considering that it is computed using trips attraction as a proxy of job opportunities.

The results have some limitations and open a few paths for further research. Geographic access metrics per transport mode might bring multicollinearity problems when using a multivariate regression method. In this research we addressed this situation by computing access indices that, unlike (Adair et al., 2000), are weighted by the predominance or preference for each transport mode across the city. However, it is necessary to consider potential mobility shifts along time, to test improvements by using other approaches to estimate the indices, or to use other regression techniques to deal with multiple collinear predictors. Unlike Bourassa et al. (2007), the use of submarket variables and geographic coordinates did not fully remove spatial dependence, meaning that some unexplained localized information could be further addressed with the help of spatial econometrics of geostatistical approaches. The ability of the geometric via geographic accessibility variable to explain land-values should be investigated in different urban contexts. Different time-decay

parameters could increase or reduce the information that is added to the land-value model.

Several Latin American cities show centralized economic activities reflected in monocentric structures, similar to Guatemala City. Yet, Guatemala City also has its unique characteristics such as the urban form, topography, immediate proximity of the airport with the CBD, arrangement of main infrastructure and particular local values, meaning that the methodological approach presented in this research could be replicated to other Latin American cities and other areas, but not without adjusting it by using local knowledge.

Finally, some implications can be drawn for planning and land administration practice. More robust models for mass valuation purposes, with added transparency to the process, could be achieved by simultaneously addressing public and private accessibility from a geographic perspective. Furthermore, it could be relevant for transport planners to explicitly model the relations between accessibility by public transport infrastructure and the value of land. In developing countries, this could mean producing useful information to help the planning of transport projects in relation to the attractiveness and capitalization of land. Moreover, the approach could be an initial step towards foreseeing the potential impact of mobility-related projects on land-value. This is of great relevance to ground different financial mechanisms to enforce the economic viability and sustainability of projects such as land-value captures.

Chapter 4

Predictive land value modelling using Space Syntax and a geostatistical approach ^{*}

^{*} This chapter is based on: J. Morales, A. Stein, J. Flacke and J. Zevenbergen. Predictive land value modelling in Guatemala City using a geostatistical approach and Space Syntax. International Journal of Applied Earth Observation and Geoinformation, 1-24, 2020.

Abstract

Spatial information of land values is fundamental for planners and policy makers. Individual appraisals are costly, explaining the need for predictive modelling. Recent work has investigated using Space Syntax to analyse urban access and explain land values. However, the spatial dependence or urban land markets has not been addressed in such studies. Further, the selection of meaningful variables is commonly conducted under non-spatialized modelling conditions. The objective of this paper is to construct a land value map using a geostatistical approach using Space Syntax and a spatialized variable selection. The methodology is applied in Guatemala City. We used an existing dataset of residential land value appraisals and accessibility metrics. Regression-kriging was used to conduct variable selection and derive a model for spatial prediction. The prediction accuracy is compared with a multivariate regression. The results show that a spatialized variable selection yields a more parsimonious model with higher prediction accuracy. New insights were found on how Space Syntax explains land value variability when also modelling the spatial dependence. Space Syntax can contribute with relevant spatialized information for predictive land value modelling purposes. Finally, the spatial modelling framework facilitates the production of spatial information of land values that is relevant for planning practice.

4.1 Introduction

In certain Global South regions there is a potential to unlock the economic value of land to finance public investments on important infrastructure to cope with the current challenges of urbanization (Peterson, 2009). But, the feasibility of land-based taxation or value capture mechanisms to support municipal fiscal health is currently challenged by the ability to produce timely and accurate spatial information on land values (Bell et al., 2009; Dye & England, 2010). Land value maps are important for planning and land administration organizations to understand land value structures, as well as to other purposes such as monitoring of real estate and urban studies (Cellmer, 2014; Kuntz & Helbich, 2014). A typical constrain to produce land value maps is the availability of data. Individual appraisals for a city are costly, explaining the need of predictive modelling, i.e. land value estimation at non-appraised locations using a fitted statistical model on a sample of appraised locations.

Literature on predictive modelling of land value is grounded in a larger body of inferential modelling studies. The focus has been to uncover associations between the advantages of location quality with the economic value of land using hedonic regressions (Ahlfeldt, 2007; Iacono et al., 2011; Kuntz et al., 2014; Liu et al., 2010; Paci, Beamonte, Gelfand, Gargallo & Salvador, 2017). Location quality is commonly defined by means of *geographic-access* metrics. These describe the most common understanding of urban access, being defined as the easiness at origin to reach desired locations or opportunities at destinations (Batty, 2009; Curl et al., 2011; Geurs et al., 2004). Following a long standing urban economy theory, most of these studies rely on the assumption that accessibility to a central business districts (CBD) is an important determinant of the value structure (Ahlfeldt, 2007; Alonso, 1964; Ryan, 1999).

Recent research outlines the relevance of using Space Syntax (SSx) metrics to expand the understanding of urban access and its relations with land values (Chiaradia et al., 2009; Giannopoulou, Vavatsikos & Lykostratis, 2016; Morales, Flacke & Zevenbergen, 2019b; Saeid, 2011; Xiao et al., 2016a; Xiao, Webster & Orford, 2016b). SSx analyse *geometric-access* by means of two main metrics, namely integration and choice at various spatial radii (e.g. 0.8 Km, 5 Km and unrestricted radius "rN"). Highly integrated urban areas correlate with presence of economic activities and flows attraction. Choice analyses normally correlate with the hierarchy of urban roads. SSx focuses on the morphological aspects of roads structure that are proven to be associated with various urban phenomena (Morales et al., 2019b; Omer, Rofè & Lerman, 2015; Webster, 2010). As a set of theories and

computational techniques, SSx is based on graph theory (Hillier et al., 2012; Jiang & Liu, 2009; Porta et al., 2006).

Even though spatial dependence, producing autocorrelated errors, is often reported in property value studies (Dubin, Pace & Thibodeau, 1999; Getis, 2007; Krause et al., 2012; McMillen, 2004), this has been largely ignored in SSx related research. Autocorrelated errors may arise from missing locational information and value interdependencies as a function of proximity (Basu et al., 1998; Bourassa et al., 2007; LeSage et al., 2009). Morales et al. (2019b) reported that including SSx metrics and additional submarket information, as suggested by Bourassa et al. (2007), leads to a reduction of spatial dependence but does not completely overcome it. A typical strategy to deal with autocorrelated errors in a multivariate regression (MR) is to consider more predictors, sub-market variables or even the observation coordinates prior to adopting spatial modelling techniques (Bourassa et al., 2010; Des Rosiers et al., 2000; Seya, Tsutsumi, Yoshida & Kawaguchi, 2011; Spinney, Kanaroglou & Scott, 2011). For inferential purposes this might be beneficial at the expense of more complex models and overfitting.

In turn, many studies provide empirical evidence on the importance of modelling the spatial dependence for model estimation (Case, Clapp, Dubin & Rodriguez, 2004; Du et al., 2012; Dubin et al., 1999; Kuntz et al., 2014; Luo & Wei, 2004; McCluskey, McCord, Davis, Haran & McIlhatton, 2013; McCluskey, Deddis, Lamont & Borst, 2000; Seya et al., 2011; Tsutsumi, Shimada & Murakami, 2011; Walacik, Cellmer & Żróbek, 2013; Yoo & Kyriakidis, 2009; Zhang, Du, Geng, Liu & Huang, 2015). Spatial econometrics has been widely used for inferential modelling while geostatistics has been preferred for spatial prediction (Anselin, 2010). Yet, something that has received little attention in spatial statistics literature overall is the selection of variables under spatialized modelling conditions (Hoeting, Davis, Merton & Thompson, 2006). It is of common practice to select variables based on their correlation, significance test, or using stepwise procedures and plugging the resulting formulation into a spatial model. However, the presence of spatial autocorrelation might lead to instability and biased regression coefficients, meaning that selecting variables under this condition might be a less optimum solution. Hence, it is relevant to explore whether variable selection should be conducted under spatialized modelling conditions.

The objective of this chapter is to construct a land value map using geostatistics that considers SSx and a spatialized variable selection in Guatemala City. The contributions of our research are twofold. Firstly,

it extends the literature on Space Syntax; so far only used in non-spatial inferential modelling. Secondly, it presents a spatialized variable selection procedure. We used the data set from Morales et al. (2019b), consisting of point-based observations of residential land value appraisals and associated predictors: property-level and neighbourhood characteristics, submarket dummy variables, and geographic and geometric access metrics. We specified three models. The first model is a multivariate regression (MR) estimated using ordinary least squares (OLS) and an automated procedure to select relevant explanatory variables using Akaike Information Criterion (AIC). In the second model we extended the first model by using the variables selected previously and using regression-kriging (RK) and maximum likelihood estimation (MLE) to solve the coefficients and address autocorrelation. In the third model RK and MLE were used to conduct variable selection and derive a reduced model for spatial prediction. The three models were compared on the grounds of their goodness of fit to observed data and prediction accuracy at unobserved locations.

The content of the chapter is organized as follows. We describe the case study, dataset and methods in section 4.2. Section 4.3 presents the results and discussion. Finally, we provide conclusions of this chapter in section 4.4.

4.2 Study area, data set and methods

4.2.1 Guatemala City

Guatemala City, the capital of Guatemala, has an extension of 996 Km² accommodating approximately 3 million people, 26% of the country's population. It shares characteristics with various of the Latin American cities in terms of urban structure and development dynamics: colonial heritage, centralized economic activities, peripheral unplanned expansion (Ford, 1996). The selection of the case study attempts to contribute to the limited research in this region, as well as to tap on their shared commonalities, which are important for a generic application of our methods.

4.2.2 Data set

The data set was available from research conducted in the study area (Morales et al., 2019a; Morales et al., 2019b), chapters two and three. The data consists of 1,169 observation points of residential land value appraisals and a hexagonal tessellation of the city urban area (Figure 4.1). Appraisals were extracted from a private data base of a real estate company in Guatemala during a fieldwork in 2014-2015 and

georeferenced using the WGS84 coordinate system (Morales et al., 2019b). The land value is expressed in local currency over a unit of surface area (Quetzal/m²). Points are georeferenced using the WGS84 coordinate system and represent the parcels centroids of the appraised residential properties. The observations are randomly split into a training and a test data set, 876 and 293 correspondingly.

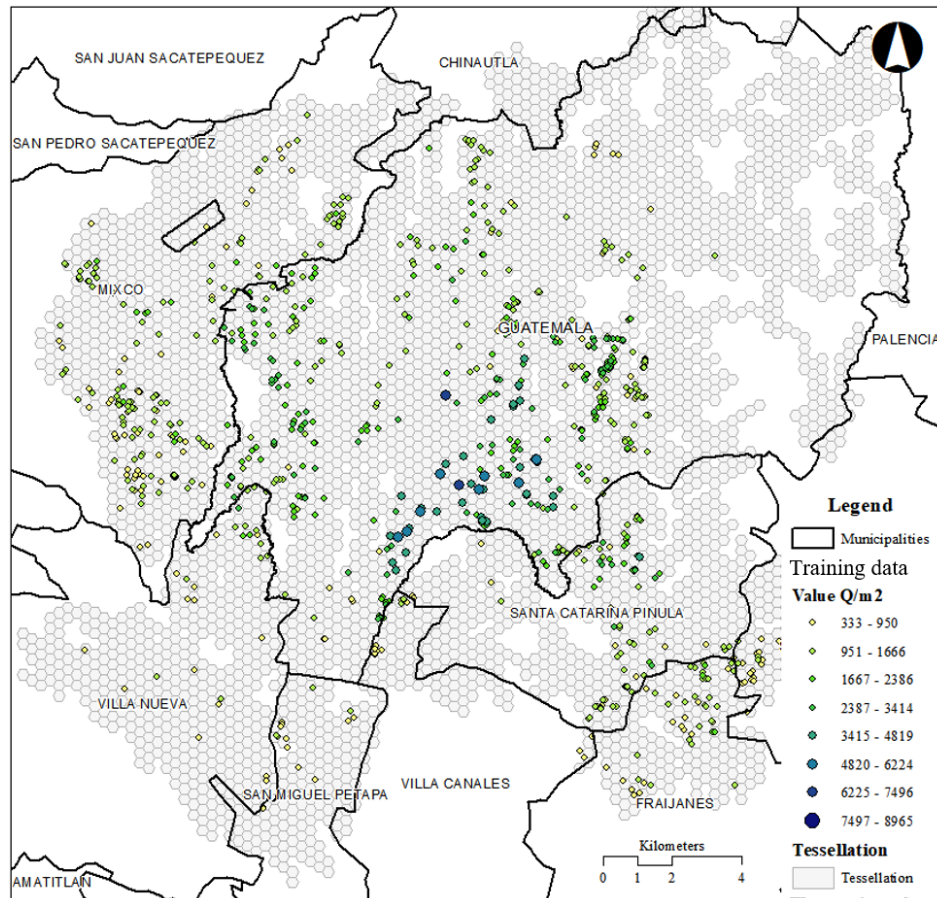


Figure 4.1: Spatial distribution of training observations and the hexagonal tessellation, adapted from (Morales et al., 2019b).

Figure 4.1 shows the spatial distribution of the training observations and the hexagonal tessellation and Table 4.1 (same as Table 3.1) shows descriptive statistics of the land values, our target variable $y(s)$, and the explanatory predictors ($x_k; k = 1, \dots, p$) proposed in (Morales et al., 2019b). The tessellation aggregates information about the spatial distribution of geographic-access to various facility types (e.g. CBD, jobs, malls and education), SSx geometric-access metrics at various spatial radii (i.e. integration and choice), a *geometric via geographic*

access metric, proximity to main infrastructure (e.g. main roads), submarket classification (e.g. market segmentation), and neighbourhood level characteristics (e.g. population density, intensity of new residential developments). Geometric via geographic-access index represents the potential access to SSx integration (rN), hence the acronym *integration_gravity*. As described in Morales et al. (2019b, pp. 762-765), section 3.2.3 of this dissertation, it was computed per hexagon i per transport mode (private and public) using equation (4.1). This is a potential access formulation (Hansen, 1959) where the average global integration (r_N) that is reachable at any hexagon, hex_j , determines the attraction size. This resource is penalized by a function of travel time t and estimated parameters α and β . Then, the *integration_gravity* index combines the two measurements weighted by the percentage of users per transport mode. This index aims to capture the capitalization of urban land as a function of the access to vital urban areas favoured by the presence of economic activities, or potential for such, that are not explicitly addressed by the geographic-access indexes. Geographic-access and geometric via geographic access are both expressed as indexes from 0 (low access) to 1 (high access) and are based on the individual metrics per mobility mode. More details about the computation of the access metrics and indexes are found in published literature (Morales et al., 2019b).

$$integration_gravity\ at\ hex\ i = \sum \left(\frac{\left(\frac{n-1}{\sum_{i=1}^n D_{\theta}(x,i)} \right)^{hex\ j}}{Nseg\ at\ hex\ j} \right) \alpha \exp(-\beta * t_{hex\ i-hex\ j})$$

(4.1)

Lastly, appraisals are associated with information aggregated at the plot-level such as year of the appraisal, plot surface area, built-up area, intra-block location (corner or not) and geometry (regular or not). Unlike the previous study, we leave out the coordinates of the observations as predictors since we address the spatial dependence explicitly by means of a geostatistical method.

Table 4.1: Descriptive statistics accompanied by short descriptions

Group	Acronym	Description	Type	Mean	St. Deviation	Min	Max	% coded 0	% coded 1
Response variable	lv	Land value in Q/m ²	Ratio	1678.45	958.41	325.75	8203.43	<input type="checkbox"/>	<input type="checkbox"/>
Geographic accessibility by public transport	stand_groceries	Standardized access score (from 0-1) based on number of facilities accessible within 10 minutes driving	Ratio	0.08	0.14	0.00	0.80	<input type="checkbox"/>	<input type="checkbox"/>
	stand_bank_rest			0.04	0.09	0.00	0.94	<input type="checkbox"/>	<input type="checkbox"/>
	stand_parks			0.05	0.09	0.00	0.76	<input type="checkbox"/>	<input type="checkbox"/>
	stand_schools			0.03	0.10	0.00	1.00	<input type="checkbox"/>	<input type="checkbox"/>
	stand_clinics			0.06	0.13	0.00	1.00	<input type="checkbox"/>	<input type="checkbox"/>
	stand_markets			0.03	0.13	0.00	1.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_cbd	Travel time to CBD	Ratio	27.21	9.74	2.13	51.59	<input type="checkbox"/>	<input type="checkbox"/>
	time_jobs	Gravity-based access to job opportunities		116375.94	50122.88	1244.84	240368.04	<input type="checkbox"/>	<input type="checkbox"/>
	time_xl_mall	Travel time to closest facility using public transport		39.50	18.56	3.00	109.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_xl_grocery			29.08	15.26	2.00	93.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_university			39.86	22.89	2.00	130.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_culture			42.44	21.23	2.00	111.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_hospital			30.90	19.11	0.00	101.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_xl_sports			50.20	24.12	4.00	136.00	<input type="checkbox"/>	<input type="checkbox"/>
Geographic accessibility by private vehicle	s_groceries	Standardized access score (from 0-1) based on number of facilities accessible within 10 minutes driving	Ratio	0.24	0.20	0.00	0.91	<input type="checkbox"/>	<input type="checkbox"/>
	s_ban_rest			0.10	0.16	0.00	0.99	<input type="checkbox"/>	<input type="checkbox"/>
	s_parks			0.13	0.19	0.00	1.00	<input type="checkbox"/>	<input type="checkbox"/>
	s_schools			0.12	0.18	0.00	1.00	<input type="checkbox"/>	<input type="checkbox"/>
	s_clinics			0.17	0.21	0.00	0.95	<input type="checkbox"/>	<input type="checkbox"/>
	s_markets			0.10	0.18	0.00	1.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_cbd	Travel time to CBD using private vehicle	Ratio	59.72	20.92	8.70	127.80	<input type="checkbox"/>	<input type="checkbox"/>
	time_jobs	Gravity-based access to job opportunities		48502.84	20746.17	9964.52	96362.19	<input type="checkbox"/>	<input type="checkbox"/>
	time_xl_mall	Driving time to closest facility		14.60	6.42	0.00	33.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_xl_grocery			11.36	6.47	1.00	33.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_university			17.02	9.65	1.00	48.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_culture			19.26	9.69	0.00	42.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_hospital			12.56	8.24	0.00	41.00	<input type="checkbox"/>	<input type="checkbox"/>
	time_xl_sports			21.48	9.87	1.00	51.00	<input type="checkbox"/>	<input type="checkbox"/>
Geometric accessibility (Space Syntax)	int_08	Average integRatio r_0.8Km	Ratio	49.95	35.86	6.37	272.05	<input type="checkbox"/>	<input type="checkbox"/>
	int_15	Average integRatio r_1.5Km		106.21	90.15	6.33	558.83	<input type="checkbox"/>	<input type="checkbox"/>
	int_25	Average integRatio r_2.5Km		206.21	191.25	8.80	1060.01	<input type="checkbox"/>	<input type="checkbox"/>
	int_50	Average integRatio r_5Km		524.75	475.18	58.28	2399.83	<input type="checkbox"/>	<input type="checkbox"/>
	int_75	Average integRatio r_7.5Km		922.39	754.09	135.55	3244.11	<input type="checkbox"/>	<input type="checkbox"/>
	int_n (global)	Average integRatio r_N		3063.52	863.52	1396.33	4807.18	<input type="checkbox"/>	<input type="checkbox"/>
	nach_08	Maximum normalised choice r_0.8Km		1.26	0.18	0.00	1.56	<input type="checkbox"/>	<input type="checkbox"/>
	nach_15	Maximum normalised choice r_1.5Km		1.23	0.18	0.00	1.53	<input type="checkbox"/>	<input type="checkbox"/>
	nach_25	Maximum normalised choice r_2.5Km		1.21	0.18	0.00	1.44	<input type="checkbox"/>	<input type="checkbox"/>
	nach_50	Maximum normalised choice r_5Km		1.16	0.18	0.00	1.43	<input type="checkbox"/>	<input type="checkbox"/>
	nach_75	Maximum normalised choice r_7.5Km		1.14	0.19	0.00	1.44	<input type="checkbox"/>	<input type="checkbox"/>
	nach_n	Maximum normalised choice r_N		1.10	0.19	0.00	1.49	<input type="checkbox"/>	<input type="checkbox"/>
Proximity to infrastructure	dist_mroad	Planar distance to main roads as classified in OSM, expressed in meters	Ratio	333.11	313.19	0.19	1492.12	<input type="checkbox"/>	<input type="checkbox"/>
	dum_prox_bus	1 if location is within 500mt distance to any bus line	Dummy	<input type="checkbox"/>	<input type="checkbox"/>	0	1	33%	67%
Sumbarkets	dum_west	1 if area is within the west municipalities	Dummy	<input type="checkbox"/>	<input type="checkbox"/>	0	1	78%	22%
	dum_east	1 if area is within the east municipalities		<input type="checkbox"/>	<input type="checkbox"/>	0	1	84%	16%
	condo_seg	Average total selling price of horizontal housing offer classified from 0 to 3 accord.	Ratio	1.83	1.11	0.00	3.00	<input type="checkbox"/>	<input type="checkbox"/>
	flat_seg	Average total selling price of vertical housing offer classified from 0 to 3 accord.		0.84	1.07	0.00	3.00	<input type="checkbox"/>	<input type="checkbox"/>
	Neighbourhood characteristics	pop_dens	Population density projected to 2015*	Ratio	46.83	21.62	3.38	98.26	<input type="checkbox"/>
soc_economic		Predominant socio economic level (from 1-5) per census block	4.28		0.97	1.00	5.00	<input type="checkbox"/>	<input type="checkbox"/>
percent_priv		Percentage of private-vehicle-based generated trips per TAZ	0.22		0.18	0.03	0.89	<input type="checkbox"/>	<input type="checkbox"/>
dens_flats		Density of recent horizontal housing offer (2013-2014) expressed in units/km ²	Ratio	4785.69	7024.00	0.00	36605.64	<input type="checkbox"/>	<input type="checkbox"/>
dens_condos		Density of recent vertical housing offer (2013-2014) expressed in units/km ²		5367.09	4381.67	0.00	15915.49	<input type="checkbox"/>	<input type="checkbox"/>
Plot characteristics	year	Year when the property was appraised	Ratio	6.46	1.49	4.10	8.98	<input type="checkbox"/>	<input type="checkbox"/>
	plot_area	Plot surface area in m ² without construction		277.24	182.10	100.29	997.00	<input type="checkbox"/>	<input type="checkbox"/>
	const_area	Total construction area in m ² without plot surface area		251.85	111.94	100.48	784.00	<input type="checkbox"/>	<input type="checkbox"/>
	dum_geometry	1 if plot geometry is rectangular, else 0	Dummy	<input type="checkbox"/>	<input type="checkbox"/>	0	1	26%	74%
	dum_intrablock	1 if plot is located on the corner of the block		<input type="checkbox"/>	<input type="checkbox"/>	0	1	86%	14%
	pot	Value indicates building potential according to current policy	Ordinal	3.25	1.08	0	5	<input type="checkbox"/>	<input type="checkbox"/>
	f_x	Estimated quadratic function of 'x'	Ratio	7.29	0.24	6.50	7.50	<input type="checkbox"/>	<input type="checkbox"/>
f_y	Estimated quadratic function of 'y'	7.28		0.21	6.44	7.45	<input type="checkbox"/>	<input type="checkbox"/>	

**Grey highlighted cell indicates measurements that were transformed into natural logarithms (nl)

4.2.3 Predictive modelling

We used RK to model the spatial dependence, conduct variable selection and derive a parsimonious model for spatial prediction. In order to evaluate our modelling strategy against traditional approaches, we used the test data to measure prediction accuracy and compare the results with those of a MR where variable selection was conducted under non-spatialized conditions; and a RK that uses the selected variables via the MR. Then, a land value map and a prediction uncertainty map were computed using our RK model. The predicted log values were transformed back to currency over unit of surface area. All computations were carried out using the statistical application R (R, 2016).

The point-based land value appraisals constitute our sample observations $y(s)$ and hexagon centroids are the target for predictions (s_0). In practice, we would like to predict $\hat{y}(s_0)$ at each parcel addressing the particular property-level characteristics of each. Since parcels data were not available for the entire area, the hexagonal tessellation helped to overcome such limitation. Consequently, we also defined average property-level characteristics at each s_0 . This means that the final land value map applies to average parcels with surface areas of approximately 300m² with built-up areas of approximately 250m² and regular geometries. The limitations imposed by this homogeneous assumption are discussed in section 4.3.3.

Before model fitting, values of each explanatory variable x_k from the training data set, except the dummy variables, were scaled to values between 0 and 1 using the minimum and maximum values of each corresponding x . Then, x_k values in the test data set and in the hexagonal tessellation were scaled conditional to the minimum and maximum values from the training observations. This way, the three data sets (i.e. training, test, hexagons) were numerically linked so the coefficients derived from the training data could be used for predictions. This method was preferred over a Z-score transformation to facilitate coefficient interpretation as the elasticity of land values with respect to the explanatory variables.

4.2.3.1 Multivariate regression (MR)

The first model uses MR which is widely accepted to model property and land value using hedonic pricing principle (Des Rosiers et al., 2000; Law, 2017; Liu et al., 2010). A generalized hedonic formulation is shown in equation (4.2). It assumes that the appraised land value $y(s)$

is equal to the sum of a constant intercept (β_0) plus the sum of the numerical contributions of the spatial and non-spatial characteristics and their corresponding hedonic economic value ($\beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots \beta_k \cdot x_k$). For example, $\beta_1 \cdot x_1$ could express the economic value of CBD access at a given location, with x_1 being the access index. The difference between the appraised and the estimated value is the error $\varepsilon(s) = y(s) - \hat{y}(s)$. Errors are expected to be random with a mean centred at 0 and a constant variance. Errors reflect market imperfections, measurements errors and missing explanatory information. MR's are commonly fitted via ordinary least squares (OLS).

$$y(s) = \beta_0 + \sum_{k=1}^p \beta_k \cdot x_k + \varepsilon(s) \quad (4.2)$$

4.2.3.2 Non-spatial variable selection

It is expected that not all variables contribute with meaningful information to the model, these should be excluded using a selection criterion. We used the Akaike Information Criterion (AIC), equation 4.3, which penalizes the model maximized log-likelihood \hat{L} with the number of parameters k used in the model (Bozdogan, 1987; Held & Sabanés Bové, 2014). AIC can be interpreted as the information loss when fitting an incorrect model from a set of candidate models, hence in model selection the goal is to minimize it (Hoeting et al., 2006).

$$AIC = 2k - 2 \log(\hat{L}) \quad (4.3)$$

Typical strategies to assess candidate models for variable selection are the stepwise regressions: bidirectional, backward or forward. The model reported in Morales et al. (2019b) is based on a bidirectional process. After testing for forward and backward procedures, we found that the selected variables were the same. Thus, our motivation was to stick to the backward procedure in this work, starting from the most complex model towards identifying a more parsimonious model. The algorithm examines the information loss (AIC) of candidate models that are generated by sequentially removing each variable from a fully specified model. The model alternative that minimizes the AIC is selected as the new full model and the process is repeated until minimization of AIC is no longer possible. This was implemented as an automated procedure using the MASS package from R application (Ripley et al., 2016).

4.2.3.3 Regression-kriging (RK) as an extension of Multivariate regression

The second model uses RK as a well-known interpolation technique of the non-stationary hybrid geostatistics (Hengl, Heuvelink & Stein,

2004; Meng, Liu & Borders, 2013; Odeh, McBratney & Chittleborough, 1994; Wackernagel, 2013; Zhu & Lin, 2010). Non-stationarity assumes the presence of a spatial trend (drift) that can be explained by means of auxiliary co-variables to improve spatial prediction. It has been extensively applied in fields such as epidemiology and mineralogy. Literature on its application for predictive modelling of land value is still limited (Basu et al., 1998; Chica Olmo, 1995; Dubin et al., 1999; Kuntz et al., 2014). Yet, RK and other spatialized models might offer adequate and superior alternatives for predictive purposes applied to mass appraisal tasks (Bell et al., 2009; Cellmer, 2014; Jahanshiri, Buyong & Shariff, 2011; McCluskey et al., 2013; Walacik et al., 2013). The work of Luo et al. (2004) and Tsutsumi et al. (2011) reported on the advantages of RK and other geostatistical methods compared to MR to improve prediction accuracy, even in cases when auxiliary variables are poorly correlated with the target variable (Meng et al., 2013). RK is used here as a convenient extension to MR (MR_K) to deal with spatial dependence and estimate the best linear unbiased predictors.

In RK predictions at unobserved locations s_0 follow equation (4.4) (Hengl et al., 2004). This is a summation of the predicted drift using the MR formulation $\hat{\beta}_0 + \sum_{k=1}^p \hat{\beta}_k \cdot x_k$, plus the interpolated residuals that are estimated via ordinary kriging $\sum_{i=1}^n w_i(s_0) \cdot \varepsilon(s_i)$. To do so, empirical semivariograms are sampled to explore the spatial structure of the drift residuals. Semivariograms indicate the degree of dissimilarity among residuals as a function of distance (h) between pairs of points. They are estimated as in equation (4.5) with $N(h)$ being the number of pairs of points separated by approximately the same distance h . Semivariograms are modelled using a semivariogram function, which together with the configuration of the observation points and the location to predict determines the kriging weights $w_i(s_0)$.

$$\hat{y}(s_0) = \hat{\beta}_0 + \sum_{k=1}^p \hat{\beta}_k \cdot x_k + \sum_{i=1}^n w_i(s_0) \cdot \varepsilon(s_i) \quad (4.4)$$

$$\gamma(\bar{h}_g) = \frac{1}{2|N(h)|} \sum_{i=i}^{N(h)} (\varepsilon(s_i) - \varepsilon(s_i + h))^2 \quad (4.5)$$

Common semivariogram functions are the spherical and exponential functions. Based on previous tests with this data, we used the exponential model formulated in equation (4.6) (Cressie, 1992). The model relies on three parameters: the nugget C_0 indicating γ at distance $h = 0$, the partial sill C_1 indicating the difference between C_0 and the value of γ when spatial dependence becomes negligible, and range r indicating distance between sampled errors when $C_0 + C_1$ is reached. Prediction error variance is estimated with equation (4.7). It

can be interpreted as the prediction uncertainty and relies on the distribution of observed locations.

$$\gamma(h) = \begin{cases} 0, & \text{for } |h| = 0 \\ C_0 + C_1 \left[(1 - e^{-\frac{|h|}{\tau}}) \right], & \text{for } |h| > 0 \end{cases} \quad (4.6)$$

$$\sigma^2(s) = \sigma^2 \{ \hat{y}(s) \} + \sigma^2 \{ \varepsilon(s) \} \quad (4.7)$$

The two computation alternatives to solve the parameters in (4.4) are explained in Dubin et al. (1999) and Cressie (2015, pages 91 and 166). The first alternative starts with solving β_k using OLS. Then, a covariance matrix of the residuals \mathbf{C} of $n \times n$ (4.8) is computed using a semivariogram function, e.g. (4.6). Next, the β_k are solved again by means of generalized least squares (GLS) given \mathbf{C} . This process is iterated until convergence occurs, i.e. the estimates β_k stabilize (Chica Olmo, 1995; Opsomer, Ruppert, Wand, Holst & Hössjer, 1999). In practice a single iteration can be sufficient, as demonstrated by Kitanidis (1993). The second alternative is maximum likelihood estimation (MLE), where the parameters are solved simultaneously (Dubin, 1992). MLE relies on the Gaussian assumption. When meeting the conditions of normally distributed and uncorrelated errors, MLE is equivalent to OLS fitting. MLE-based parameters are described as the values that are more likely to be the true parameters in the process generating the data. In this research we implemented the MLE alternative using the "likfit" function in the GeoR package (Ribeiro Jr & Diggle, 2015).

$$\mathbf{C} = \begin{bmatrix} C(s_1, s_1) & \cdots & C(s_1, s_n) \\ \vdots & \ddots & \vdots \\ C(s_n, s_1) & \cdots & C(s_n, s_n) \end{bmatrix} \quad (4.8)$$

4.2.3.4 RK for spatialized variable selection

Our third model, serving the main objective, uses RK for spatialized variable selection. This was conducted by manually implementing a backward stepwise regression as it is done in the MASS package algorithm. We started the selection process with a first round where we fit a "full model" with p number of initial predictors x_k (Table 4.1). Then we fit a number of alternative models equal to p . In each alternative model we removed one variable at the time, as in (4.9). The models were ranked using their AIC statistic as reported by the *likfit* summary function. The best alternative was identified with AIC_{\min} showing no loss of information from all candidate models (Burnham &

Anderson, 2004). Such a model became the new “full model” and the process was repeated again in a second round of models. The process stopped when the AIC_{\min} was found to be the “full model”. The final model, as in equation (4.9) contains the retained predictors and it was used for spatial prediction.

$$\hat{y}(s_0) = \hat{\beta}_0 + \sum_{k=1}^{p-1} \hat{\beta}_k \cdot x_k + \sum_{i=1}^n w_i(s_0) \cdot \varepsilon(s) \quad (4.9)$$

4.2.3.5 Model assessment and cross validation

The three models were assessed and compared using the following statistics. The goodness of fit over the training data is indicated by the adjusted R^2 and root mean squared error (RMSE). The models were cross validated using the test data to evaluate the prediction accuracy at non-observed locations. We computed the mean error, mean squared error, the root mean squared error and the goodness of prediction measure G . This is formulated in (4.10) (Cressie, 2015, p. 164), where $\bar{y} = \sum_{s_0} y(s_0)/n_t$, with n_t being equal to the number of test observations. The measure is expressed as a percentage indicating the explained variability of non-observed data.

$$G = \left[1 - \left\{ \sum_{s_0} (y(s_0) - \hat{y}(s_0))^2 / \sum_{s_0} (y(s_0) - \bar{y})^2 \right\} \right] * 100 \quad (4.10)$$

We omitted multicollinearity tests since our aim is to use the model for prediction. Further, the results of Morales et al. (2019b) suggest that observable multicollinearity among CBD, jobs and integration gravity variables did not affect coefficient signs or significance.

4.3 Results and discussion

This section presents the results and accompanied discussion. The first part focuses on the first two models (MR and MR_K). Then, we focus on the results of RK with a spatialized variable selection. Lastly, we present and discuss the results of the land value map construction.

4.3.1 The MR and the MR_K

We observe in Table 4.2 that by omitting the coordinates in the MR model, the backward stepwise procedure resulted in a very similar set of variables as reported in Morales et al. (2019b). New selected variables are access to clinics, additional geometric-access metrics at different spatial scales (i.e. integration at 1.5 km, 5 km, and normalized choice at radius at 2.5 km) and plot geometry. Further, access to schools was discarded in this model.

Unlike the models previously reported in Morales et al. (2017b), coefficients in the models used here are comparable among each other and same interpretations can be derived from the coefficients. The signs of the coefficients are the same and the mono centricity of the land value structure is observed given the importance of the CBD access. The model performs fairly well with a total of 30 predictors retained explaining 73% of the variability of the training data and a goodness of prediction $G = 63\%$.

We observe an adjustment of the coefficients after extending the MR with RK (MR_K). This was expected as the coefficients are now solved via MLE accounting for the spatial dependence. The hedonic contributions of some variables are now reduced by up to 0.20, with reduction of significance as well: geographic-access to CBD, geometric access at localized scales (i.e. integration at 0.8Km, 1.5Km, normalized choice at 1.5 km, 2.5 km), some submarket and neighbourhood variables (i.e. east, west, population density, percentage of users of private vehicle, density of new residential buildings, density of new condominium projects). Particularly for the geometric-access variables it becomes harder to discuss about their true contribution to the model.

In turn, we observe an increase of the coefficients of the following variables: access to banks and restaurants, access to hospitals and the geometric via geographic access. For the later we notice that its coefficient equals the one of the geographic access to CBD. Meaning that an MLE-based estimation of such coefficients under spatialized model conditions could lead to additional insights about how both variables are complementary and relevant to explain the variability of residential land values.

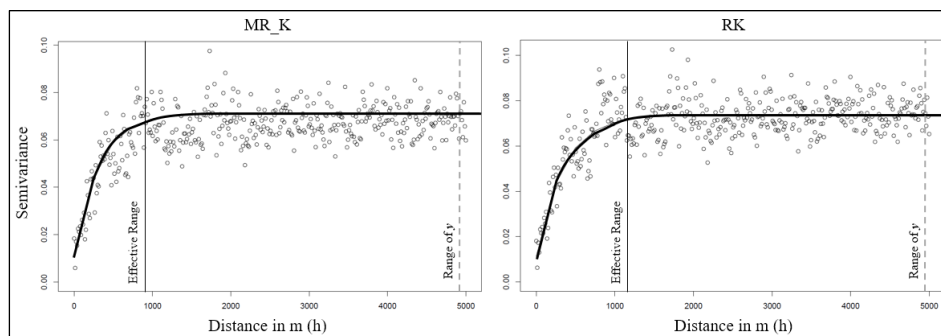


Figure 4.2: MLE fitted exponential semivariogram functions over the model residuals: MR_K on the left side and RK on the right side.

The MR_K explains up to 79% of the training data with a reduction of the RMSE and slightly higher prediction accuracy, $G = 65\%$. The spatial

structure is defined with a nugget C_0 of $\gamma = 0.01$, a partial sill C_1 of $\gamma = 0.06$ and an effective range equals to 0.9 km, see Figure 4.2. We suggest that predictors explain the global trend of the land value structure fairly well, as we already noticed previously in the MR model. However, modelling the spatial structure with RK increases the ability to explain variability within a radius ~ 1 km, similar to a neighbourhood scale. In other words, the coefficients value drop can be partly explained as the interpolated error being able to capture more effectively localized information compared to those predictors.

Table 4.2: Reports of the models' coefficients and assessment statistics

Model parameters		1. MR				2. MR_K				3. RK			
		Coeff.	S. Error	t	Sig.	Coeff.	S. Error	t	Sig.	Coeff.	S. Error	t	Sig.
Geographic accessibility	Intercept	5.22	0.12	42.63	0.000 ***	5.51	0.16	34.41	0.000 ***	5.73	0.13	45.58	0.000 ***
	Groceries	□	□	□	□	□	□	□	□	□	□	□	□
	Bank_restaurants	0.46	0.16	2.81	0.005 **	0.62	0.22	2.82	0.002 **	0.91	0.16	5.55	0.000 ***
	Parks	□	□	□	□	□	□	□	□	□	□	□	□
	Schools	□	□	□	□	□	□	□	□	□	□	□	□
	Clinics	0.45	0.14	3.22	0.001 ***	0.29	0.18	1.61	0.054 *	□	□	□	□
	Markets	□	□	□	□	□	□	□	□	□	□	□	□
	CBD	1.27	0.22	5.68	0.000 ***	0.95	0.33	2.85	0.002 **	0.73	0.32	2.30	0.011 **
	Jobs	-0.64	0.28	-2.27	0.023 *	-0.79	0.45	-1.74	0.041 *	-1.26	0.45	-2.82	0.002 **
	XL_mall	0.75	0.14	5.37	0.000 ***	0.80	0.20	4.05	0.000 ***	0.84	0.20	4.14	0.000 ***
	XL_grocery	-0.50	0.11	-4.70	0.000 ***	-0.54	0.16	-3.37	0.000 ***	-0.60	0.16	-3.73	0.000 ***
	University	0.82	0.13	6.23	0.000 ***	0.50	0.20	2.57	0.005 **	0.41	0.18	2.29	0.011 **
	Culture	-1.03	0.21	-4.93	0.000 ***	-0.78	0.29	-2.64	0.004 **	-0.58	0.29	-1.99	0.024 *
	Hospitals	0.26	0.12	2.20	0.028 *	0.41	0.19	2.17	0.015 *	0.51	0.18	2.87	0.002 **
	XL_sports	□	□	□	□	□	□	□	□	□	□	□	□
Geometric accessibility (Space Syntax)	int_08	-0.66	0.19	-3.50	0.000 ***	-0.39	0.21	-1.84	0.033 *	-0.27	0.12	-2.27	0.012 **
	int_15	0.36	0.24	1.50	0.134 .	0.13	0.28	0.46	0.324 .	□	□	□	□
	int_25	□	□	□	□	□	□	□	□	□	□	□	□
	int_50	-0.78	0.16	-4.82	0.000 ***	-0.71	0.23	-3.04	0.001 ***	-1.07	0.34	-3.16	0.001 ***
	int_75	□	□	□	□	□	□	□	□	0.65	0.35	1.87	0.001 ***
	int_n (global)	0.46	0.15	3.16	0.002 ***	0.34	0.20	1.74	0.041 *	□	□	□	□
	nach_08	□	□	□	□	□	□	□	□	□	□	□	□
	nach_15	1.23	0.37	3.31	0.001 ***	0.57	0.43	1.32	0.094 .	0.32	0.10	3.38	0.000 ***
	nach_25	-0.82	0.35	-2.31	0.021 *	-0.26	0.42	-0.63	0.265 .	□	□	□	□
	nach_50	□	□	□	□	□	□	□	□	□	□	□	□
	nach_75	□	□	□	□	□	□	□	□	□	□	□	□
	nach_n (global)	□	□	□	□	□	□	□	□	□	□	□	□
Geometric via geographic access	Integration_gravity	0.69	0.27	2.60	0.010 **	0.96	0.43	2.24	0.013 **	1.30	0.39	3.30	0.001 ***
Proximity to infrastructure	Dist_mroad	□	□	□	□	□	□	□	□	□	□	□	□
Submarkets	Dum_prox_bus	-0.13	0.03	-4.88	0.000 ***	-0.10	0.03	-3.05	0.001 ***	-0.09	0.03	-2.60	0.005 **
	Dum_west	-0.13	0.04	-3.48	0.001 ***	-0.10	0.06	-1.88	0.030 *	-0.15	0.05	-2.81	0.003 **
Neighbourhood characteristics	Dum_east	0.26	0.07	3.60	0.000 ***	0.15	0.10	1.54	0.062 .	□	□	□	□
	Condo_segment	0.23	0.05	4.59	0.000 ***	0.27	0.07	4.12	0.000 ***	0.30	0.06	4.89	0.000 ***
Plot characteristics	Flat_segment	0.25	0.05	4.80	0.000 ***	0.31	0.08	4.03	0.000 ***	0.30	0.07	4.47	0.000 ***
	Pop_dens	0.20	0.11	1.79	0.074 .	0.05	0.15	0.34	0.366 .	□	□	□	□
Goodness of fit	Soc_economic	0.14	0.04	3.30	0.001 ***	0.13	0.05	2.64	0.004 **	0.12	0.05	2.48	0.007 **
	Percent_priv	0.20	0.07	2.94	0.003 ***	0.14	0.09	1.49	0.069 .	□	□	□	□
Variogram parameters	Flat_density	0.11	0.04	2.76	0.006 **	0.10	0.07	1.46	0.073 .	0.12	0.07	1.82	0.034 *
	Condos_density	0.08	0.05	1.48	0.000 ***	0.05	0.07	0.66	0.256 .	□	□	□	□
Prediction accuracy (cross validation)	Year	0.57	0.03	18.77	0.000 ***	0.59	0.02	25.43	0.000 ***	0.58	0.02	25.29	0.000 ***
	Plot_area	-0.36	0.05	-6.62	0.000 ***	-0.39	0.05	-7.77	0.000 ***	-0.39	0.05	-7.87	0.000 ***
Goodness of prediction G	Const_area	0.29	0.06	4.55	0.000 ***	0.24	0.05	4.44	0.000 ***	0.25	0.05	4.64	0.000 ***
	Dum_geometry	0.04	0.02	1.69	0.091 .	0.03	0.02	1.87	0.031 *	0.03	0.02	1.91	0.028 *
Goodness of fit	Dum_intrablock	□	□	□	□	□	□	□	□	□	□	□	□
	POT	□	□	□	□	□	□	□	□	□	□	□	□
Variogram parameters	Number of variables	30				30				23			
	Number of observations	876				876				876			
Prediction accuracy (cross validation)	AIC	167				-70				-96			
	Adjusted R ²	0.73				0.79				0.81			
Prediction accuracy (cross validation)	Nugget	□				0.01				0.01			
	Partial sill	□				0.06				0.06			
Prediction accuracy (cross validation)	Effective Range (meters)	□				941				1025			
	Mean Error	0.00				0.00				0.00			
Prediction accuracy (cross validation)	Root Mean Squared Error	0.34				0.00				0.00			
	Goodness of prediction G	63%				65%				68%			

Significance: ***0.001, **0.01, *0.05, (.) 0.1
 □ Variable was discharged during the corresponding selection procedure

4.3.2 RK for spatialized variable selection

The spatialized variable selection leads to a reduced and slightly different set of predictors compared to the MR-based procedure. After 21 rounds and fitting 714 models, 23 predictors were selected. The information contribution of some variables tends to be overestimated under non-spatialized conditions. In turn, modelling the spatial

dependence might lead to variable removal. Meaning that for the case of RK, the interpolated errors could explain the variability at a local scale better than some of the predictors considered initially. This becomes important if the objective is to use the model for predictive purposes. Further, this modelling approach yielded a more parsimonious formulation, as we observe a higher fit of the model over the training data (81%) and a higher goodness of prediction with $G = 68\%$. Although, the performance increase is modest, it is achieved with 6 predictors (p) less than the former models. Thereby, the relevance of the spatialized variable selection procedure. A small trade-off is clear by observing the parameters of the semivariance function. The total sill is now reached at longer distance at an effective range of slightly more than one kilometre (1 km), see Figure 4.2. The difference is equivalent to one average city block.

Table 4.2 indicates that predictors from the plot and neighbourhood characteristics, and submarket dummy variables that already have less significance in the MR_K model are finally dropped, as well as some of the geometric-access variables. The RK model only includes four, instead of six variables, from which three are the same as selected in the previous model. However, the retained variables are now highly significant. Which means that SSx geometric-access metrics are significantly adding spatialized information to the model.

Strikingly, the geometric via geographic access now appears to have a higher coefficient (1.30) compared to the, even more reduced, CBD access coefficient (0.73). From the results we infer that geometric via geographic access as defined in Morales et al. (2019b) could be adding highly meaningful spatialized information to explain the land value structure of Guatemala City. Whereas the CBD access metric becomes now only complementary. The mono-centric assumption by means of CBD access metrics has been a dominant explanatory proposition in property value studies (Ahlfeldt, 2007; Bourassa et al., 2010; Chica-Olmo, Cano-Guervos & Chica-Olmo, 2013; Ryan, 1999). Yet, it relies on the predefinition of a focal point based on local knowledge. In turn, the new metric relies on the configuration of the urban layout, which was observed to be associated with the distribution of various economic activities, and the access to highly integrated areas penalized by a time decay function.

Some additional deductions can be inferred from the graphical summary of the spatialized variable selection in Figure 4.3. On the left side the various alternative models are enumerated and named following the variable that was omitted in each version. The first is the full model, containing all the predictors considered. The first column

(1st) shows the AIC-based ranking of the models after fitting the 44 models. Highest information loss is observed after omitting the “plot surface area” predictor. In turn, no information loss is observed when omitting the dummy variable “intra-block”, which indicates whether a parcel is in the corner of a block or not. Hence, such variable was removed in the first round and this model became the new “full model”. As from the 2nd round, the lines allow to trace the information loss when such a variable was omitted. Last column (21st) presents the ranking of the retained predictors. These are the predictors that lead to information loss when omitted. The lower part shows the decrease of the AIC as a result of model complexity reduction.

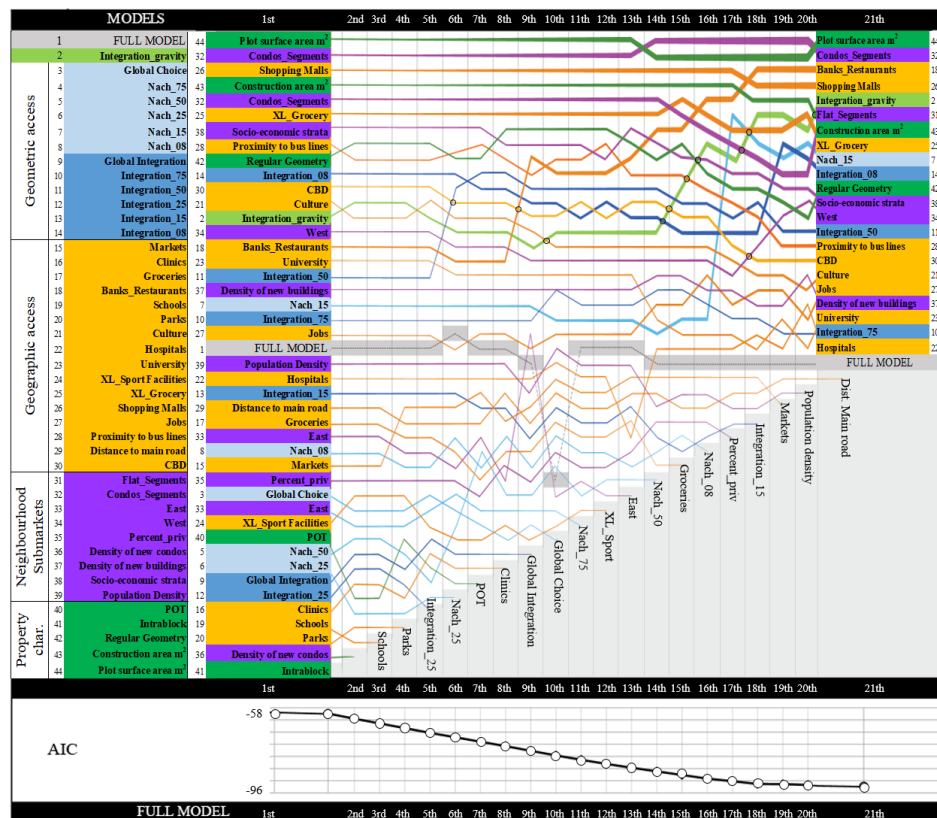


Figure 4.3: Graphic summary of the RK-based variable selection process

Although the AIC is not associated with the coefficient value, it reveals the information loss of each predictor if omitted relative to the candidate models in each round. Information loss if geometric via geographic-access is omitted leads to a higher ranking of this metric after dropping predictors such as: global integration, normalized choice

at 5 km, access to neighbourhood-scale groceries, integration at 1.5 km, and population density. In turn, information loss when CBD access metric is omitted becomes relatively lower as other predictors start to rank higher: integration at 5 km, access to banks and restaurants, the geometric via geographic access and normalized choice at 1.5 km.; after dropping these predictors correspondingly: integration at 2.5 km, access to clinics and normalized choice at 5 km.

4.3.3 Land value map

Following the results of the prediction accuracy assessment, we were confident to use our third model for spatial prediction. Prior to that, Figure 4.4 shows that errors seem to be randomly spatially distributed, meaning that spatial autocorrelation was removed. The residuals are approximately centred on zero. Both the residuals and the predictions are not significantly different from a normal distribution based on a Kolmogorov-Smirnov test ($p > 0.05$).

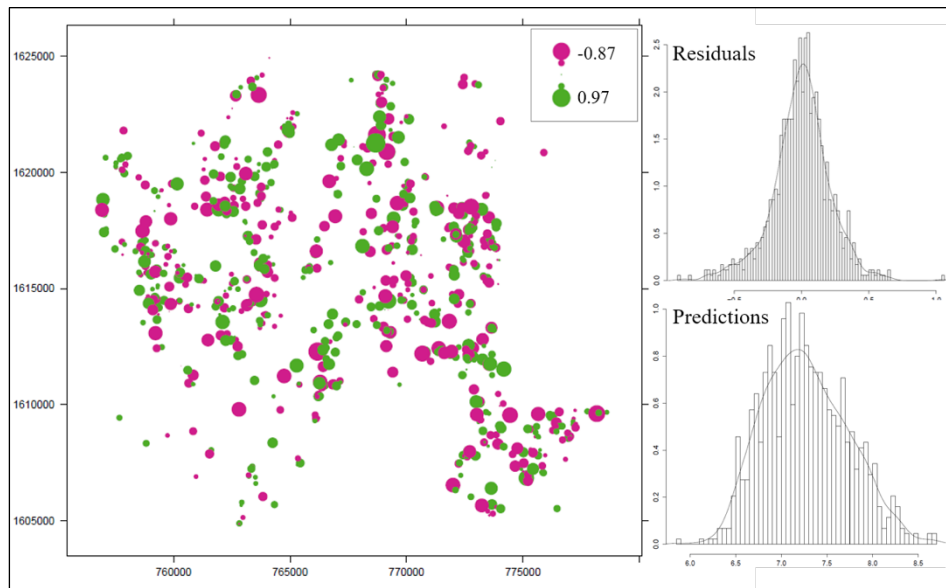


Figure 4.4: Mapping of the residuals (left), green and red for positive and negative residuals correspondingly. Frequency distribution of residuals (top-right) and predictions (bottom-right).

Figure 4.5 shows the constructed residential land value map of average property characteristics according to our sample observations. An important limitation is that in reality residential uses are more heterogeneous across the urban area (i.e. varying sizes and built up areas). Such variability should be taken into account when constructing land values for a specific purpose (e.g. taxation). This could be easily

done by using parcel level data to perform the spatial prediction, or by aggregating average property characteristics at some unit of administrative division (e.g. census tracks) and translate this to the hexagon centroids. Yet, the map provides a plausible visualization to gain insights in the land value structure in Guatemala City.

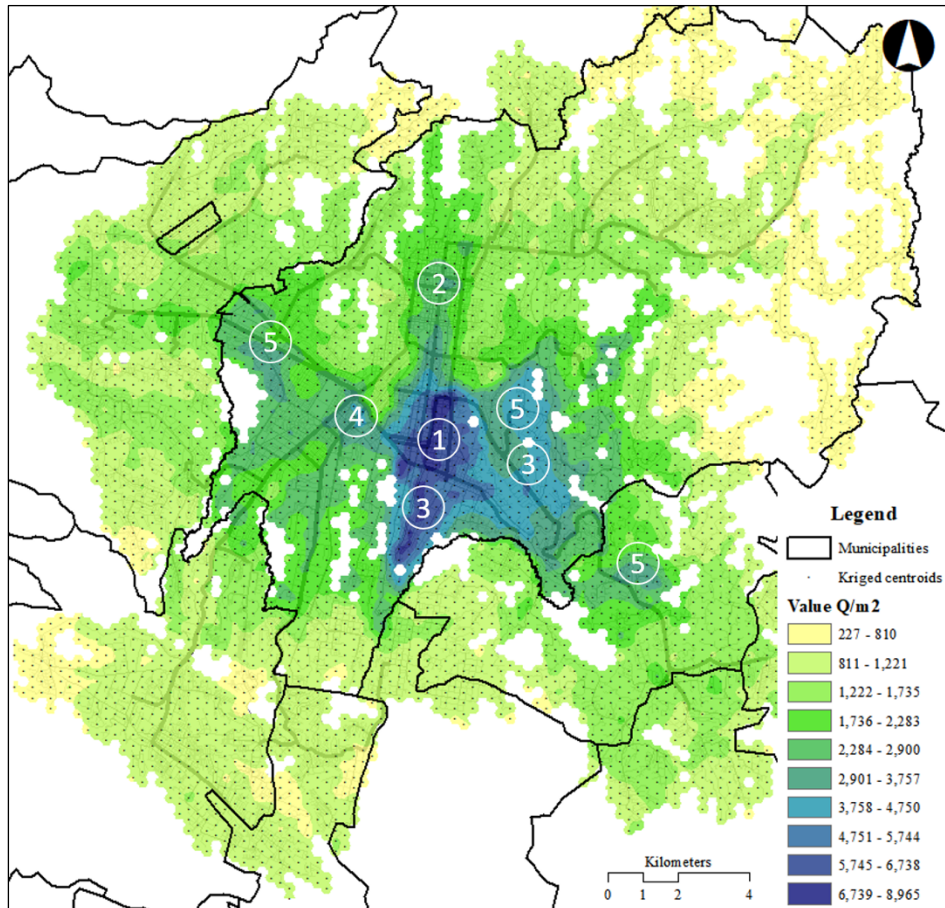


Figure 4.5: Constructed land value map, road structure is overlaid for reference.

We can visualize a clear mono centric structure, in line with the results of Morales et al. (2019b) and the conceptual models by Ford (1996) and Ingram et al. (1981). Most expensive land is observed at the core matching the CBD where various types of commerce and office buildings concentrate (area '1'). As discussed by Morales et al. (2019a) and Morales et al. (2019b), chapters two and three of this dissertation, the CBD benefits the most from geographic-access to several facilities, as well as from geometric-access and geometric via geographic access.

This explains why the highest land values can be expected in this area, even though few observations were available there.

High values slowly decrease towards the historic centre (area '2'), closely shaped by important roads. Even though this area benefits with relatively good access, similar to the CBD, lower land values are observed in this area from the training data. This reflects a known deterioration process of the historical centre that followed after the expansion towards the current CBD. For a long time, the historic centre was associated with street robbery and pollution. Furthermore, the fact that historical buildings are heavily protected has made investment and restoration a cumbersome process, resulting in a reduce bidding for properties in such area.

Areas numbered as '3' are characterized by a combination of high-income residential uses and a mix of compatible commercial uses. Values differ seemingly as a function of their geometric via geographic access and their proximity to the CBD. Number '4' outlines an important urban node, however surrounded by commercial uses of minor scale and also informal economy. Sub-centres of intense commercial development are present in areas numbered as '5'. The land value structure outlines the effects of the connectivity and continuity of central urban areas versus the discontinuity and less consolidated areas at the periphery.

Under spatialized modelling conditions and variable selection, geometric via geographic access has a greater impact in explaining land value than the CBD access metric. By observing the model coefficients and in order to facilitate some visual understanding we included in annex A the visualizations of the following metrics: geometric via geographic-access, CBD access, banks and restaurants, and jobs (Morales et al., 2019a; Morales et al., 2019b). CBD access only reflects generalized concentric patterns of travel times to an assumed focal point. Meanwhile, geometric via geographic access adds more spatial detail by considering SSx integration at the city scale (also called global integration) as the resource to be reached. Cumulative access to concentrations of banks and restaurants, adds an important layer of information that accentuates vital urban areas, also facilitating the identification of the effects of intense commercial developments in nodal areas.

Finally, Figure 4.6 shows the prediction uncertainty expressed as prediction error variance. The uncertainty of prediction closely follows the spatial distribution of the training data and the magnitude of the prediction (\hat{y}). This is shown for example in areas '1' and '4', where

little training data is available, but prediction error variance correlates with the magnitude of predictions (Figure 4.5). Some caution should be taken when interpreting and using the predictions made in areas where the uncertainty is high and amount of observations is low. In order to minimize variance as a function of the spatial sampling, the model would clearly benefit from including additional appraisals at the urban core and at the north-east periphery.

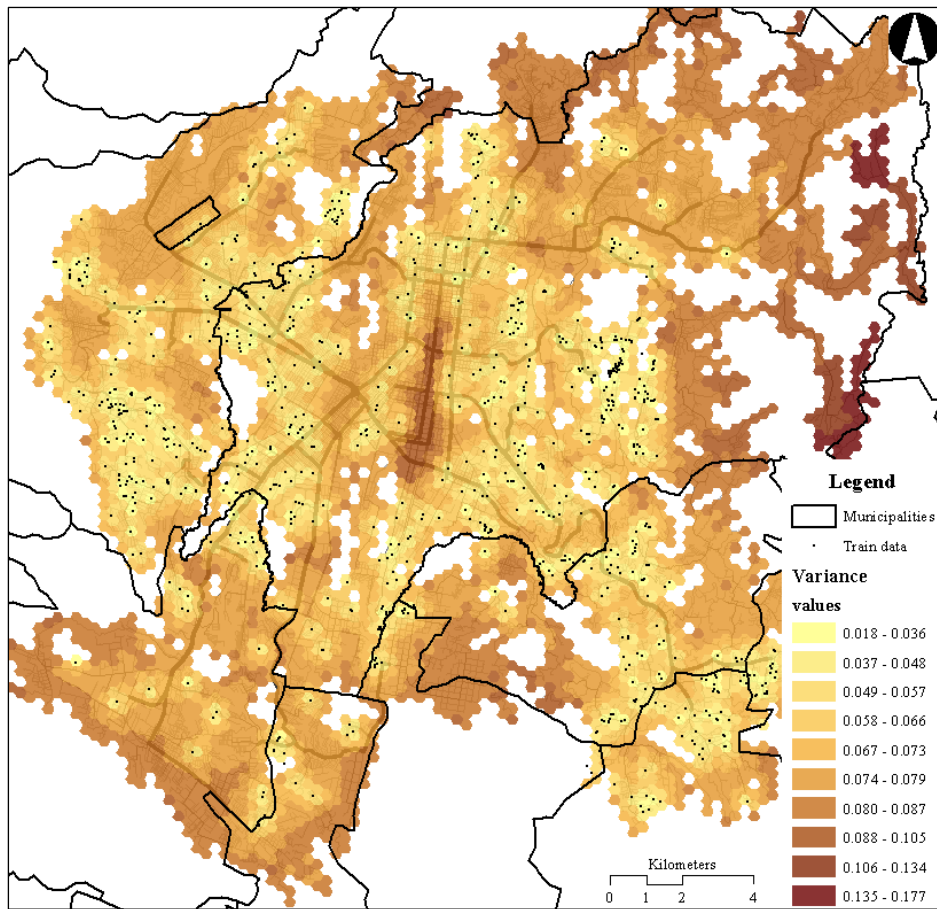


Figure 4.6: Prediction error variance.

4.4 Conclusions

We presented a predictive modelling approach based on regression-kriging (RK), geometric-access metrics as analysed in Space Syntax, and a spatialized variable selection procedure. A land value map was computed for an average observation of residential property use. Our modelling approach points that the advantages of modelling the spatial structure not only to deal with autocorrelated errors and improve

prediction, as suggested in previous research (Cellmer, 2014; Des Rosiers & Recherche, 2001; Dubin et al., 1999; Kuntz et al., 2014), but also as a departing point to achieve parsimonious models with increased accuracy and new inferential insights. We conclude that the amount of information added by a predictor to a non-spatial model might be high. However, information contribution might turn less meaningful and lead to predictor removal in the context of regression-kriging. Meaning that for predictors associated with neighbourhood scale quality, the modelled spatial structure contributes with more meaningful information to explain local land value variability. The RK-based spatialized variable selection leads to a more parsimonious model (24 predictors) compared to a non-spatialized procedure (30). The model performed with a higher accuracy of prediction explaining up to 68% of non-observed locations, compared to 65% of an RK model where variables are selected under non-spatialized conditions.

The results expand those presented in Morales et al. (2019b) by providing new conclusions regarding how Space Syntax metrics contribute to explain residential land values. Geometric-access metrics (integration and choice) do contribute with statistically significant information mostly at local scales. A mono-centric land value structure in Guatemala City is greatly explained by a time-based potential access to highly integrated urban areas (i.e. Space Syntax global integration). In turn, access to CBD as a generalized concentric pattern provides complementary information with less explanatory power reflected in a lower coefficient. We suggest that further research should expand these conclusions as those are so far only limited to the data and case study presented in this paper.

Some limitations can be discussed in relation to our work. The first is the prediction of a continuous land value map while in reality land values are discrete to parcel boundaries. The use of parcel-level data would benefit not only to overcome this limitation but also to incorporate the heterogeneity of residential properties in Guatemala City. In this regard, the model would highly benefit from additional data collection to address this concern as well as to deal with the outputs of the prediction error variance. Second, the findings and conclusions remain somehow limited to our data and case study. Whilst the entire approach is easily replicable to any city, we cannot claim with certainty about how such metrics would add spatialized information in a different context. Consequentially, we recommend that further research should examine the sensitivity of the various geometric-access metrics to explain and predict land value, as well as their relevance to be retained in a model under spatialized modelling conditions.

Our modelling approach outlines the importance on the application of geoinformation science to produce spatial information of land values. This is relevant for the planning and land administration practice in tasks such as mass appraisals. Further, it facilitates insights in the associations between locational characteristics that are addressable through planning with the land value structure. Specifically, the model could be used to estimate the impacts of mobility-related planning interventions on the residential land value structure. This could address current information challenges of unlocking land values to improve fiscal health, optimize value capture and infrastructure investment and manage fast urbanization in Global South regions.

Chapter 5

Where and how much? Predicting the impacts of a Light Rail Transit system intervention on the residential land values of Guatemala City *

* This chapter is based on: J. Morales, J. Flacke and J. Zevenbergen. Where and how much? Predicting the impacts of a Light Rail Transit system intervention on the residential land values in Guatemala City. Computers Environment and Urban Systems (Submitted), 2020.

Abstract

Interest is increasing to evaluate transport investments in the light of their future effects on land values. However, limitations are identified when reviewing the scarce literature dedicated to estimating the potential effects of transport plans on land values. In this chapter we present and implement a modelling framework to estimate the effects of introducing a Light Rail Transit (LRT) system on the residential land values in Guatemala City. We structure data and methods from recent research and extend its applicability in order to estimate land value uplifts at a city-wide scale. The framework addresses a comprehensive concept of geographic and geometric accessibility. We found that the intervention has the potential to significantly increase land values along its corridor and across a distance beyond 4 km to the access stations. Approximately, a fourth of the total land stock value uplift is already equivalent to 130% of the required transport investment. We conclude that the framework outputs could provide input for transport evaluation tools such as CBA's and information to formulate financial mechanisms and facilitate economic viability of transport investments.

5.1 Introduction

The planning of sustainable mobility such as Light Rail Transit (LRT) and Rapid Bus Transit (BRT) systems is of growing interest in many global south countries as a sustainable strategy to face the challenges of rapid urbanization (Banister, 2007; Cengiz et al., 2019; Ferbrache et al., 2017; Gleave, 2005; Ingvardson et al., 2018; Liu et al., 2011). These accessibility improvements can have wider effects such as economic land values uplifts (Ahlfeldt et al., 2011; Banister et al., 2011; Filatova et al., 2009; Giuliano et al., 2010; Mohammad et al., 2013). Land value uplifts are associated with triggering further effects such as economic growth, inward investment and land use transformations (Ferbrache et al., 2017). In the form of rents they are logically desirable for landowners but not for tenants who might be displaced as a consequence of gentrification processes (Jones et al., 2012; Lin, 2002).

There is a growing interest to evaluate transport planning projects in the light of their future effects on the spatial distribution of land values (Banister et al., 2011; Grimes et al., 2010; Li et al., 2020; Lin, 2002; Metz, 2017; Vickerman, 2017). Practitioners are having more motivations to include these effects in Cost Benefit Analyses (CBA's), the main evaluation framework of large transport infrastructure planning (Vickerman, 2007, 2017). The assumption of perfect market equilibrium hardly holds in reality, meaning that direct effects (i.e. time savings) cannot be easily translated into wider economic benefits. It is difficult to capture the added benefit of transport interventions specially when those are extensions in already existing and mature transport networks, hence making it difficult to justify the large amounts of investment required. A direct translation of the monetized accessibility improvements could shed some light on information useful for the design processes of transport infrastructure, plan evaluation/selection processes and objective formulation of mechanisms to finance transport investments (e.g. taxation adjustments, betterment levies, private investment).

However, the task of predicting land value uplifts as an effect of transport investments is not trivial nor easy. Value uplifts can be empirically observed in different stages of the transport investment, namely after project announcement and before construction (ex-ante intervention), during construction and after (Yen et al., 2018). Effects can be heterogeneous along a transport corridor based on the variability on how users do value transit-oriented investments as well as their transport modality preferences (Higgins et al., 2018). Ingvardson et al. (2018), Debrezion et al. (2007), D. Knowles et al. (2016) and Mohammad et al. (2013) provide extensive reviews on

published empirical research about the wide range of proportional value uplifts that can be associated with LRT investments. Ingvardson et al. (2018) extend their review with a comparison of reported modal shift from car ridership to BRT and LRT systems in cities across US and Europe. Overall, the research strand dedicated to analysing value uplifts after LRT interventions are made (post-ante) is quite fertile and widely international. However, results are hardly comparable due to the variability of methods, modelling strategies, used data and the specificities of each case study (Ingvardson et al., 2018; Martínez et al., 2009). Therefore, it is impossible to establish some transferable reference on the expected magnitude and spatial distribution of value uplifts. From published literature effects range from -45% to 100% or more (Cervero et al., 2002; Pan, 2013).

Contrastingly, studies that aim to predict value uplifts before the interventions are implemented (ex-ante) are extremely scarce. Viguie et al. (2014) approach the problem using an urban economic formulation based on household utility maximization. They implement the formulation in a land use transport interaction model calibrated for the city of Paris. They report that accessibility capitalization is sensitive to population expansion. Ahlfeldt (2013) proposes a non-spatial non-linear regression approach to estimate the elasticities of property price as a function of public transport access to labour markets in London area. The model was first calibrated using property transactions and then used to predict value uplift from expected travel time reductions. Gallo (2018) implements a non-spatial linear regression approach to estimate elasticities of average property value per ward as a function of counts of public transport stations. The model was calibrated using observations of asking prices aggregated at district level for the city of Naples. Cengiz et al. (2019) also use a non-spatial linear regression approach to estimate elasticities of property values as a function of Euclidian distance to existing transport stations. The model was calibrated using property values within an assumed 0.5 km catchment area of an existing transport corridor. The model was then used to predict the uplift from future interventions.

Limitations in the aforementioned literature lead us to describe the research gap that is addressed in this paper from different angles. First, markets are far from perfect equilibrium (Vickerman, 2017). In areas where market imperfections are aggravated or unknown, estimating the elasticities of land values as a function of accessibility would be a preferred approach compared to an urban economic formulation. Second, operationalizing public mobility benefits by Euclidian distance to transit stops makes it difficult to associate the value uplift with an interpretable metric of improved geographic access as it accrues to

users. Also, it wouldn't allow to estimate the value uplift effects of accessibility improvements due to new transport technologies relative to the city in question. Introduction of new transport technologies such as LRT are rarely isolated interventions, but rather a composite of urban transformations that commonly include modifications to the existing road network, with implications to private mobility as well.

Modifications to the urban layout would likely have effects on the geometric accessibility (i.e. network centrality) at various spatial scales (e.g. neighbourhood and city-wide level). Geometric accessibility, as analysed in Space Syntax (SSx), has been frequently reported to have correlation with traffic flows, land use patterns and functional hierarchy (Jiang et al., 1999; Kaplan et al., 2020; Karimi, 2012; Li et al., 2019; Serra et al., 2019; van Nes, 2019). Moreover, recent studies suggest that SSx metrics add complementary spatial information to improve the understanding of land and property values variability (Di Pinto et al., 2019; Law et al., 2017; Morales, Stein, Flacke & Zevenbergen, 2020). Therefore, it is essential to account for a comprehensive definition of accessibility not only geographically but also geometrically. Third, we suggest that the spatial scope of analysis cannot be restricted to pre-assumed catchment areas (i.e. the 0.5-1 km buffer around stations). Instead, a city scale approach would allow to calibrate elasticities based on richer datasets whilst gaining broader spatial insights on the potential land value effects of a transport investment. With this consideration it becomes increasingly relevant to address spatial dependence, a common problem in property value studies (Bourassa et al., 2010; Gallo, 2018; Krause et al., 2012).

The objective of this chapter is to present and implement a modelling framework to estimate the effects of introducing an LRT system, *ex-ante* intervention, on the residential land values in Guatemala City. To do so, we first structure methods and data from recent research into an analytical workflow to model land values. We extend its applicability and utilize the workflow to estimate value uplifts from a future transport investment at a city-wide scale. The framework addresses a comprehensive concept of geographic and geometric accessibility as well as the spatial dependence problem.

The remainder of this paper is organized as follows. Section 5.2 introduces the LRT project in the context of the study area and presents the modelling framework. Then, section 5.3 presents the results and section 5.4 provides discussion over the applicability, reproducibility and limitations. We present our conclusions in section 5.5 with suggestions for further research and policy implications.

5.2 Case study and modelling framework

5.2.1 Guatemala City and its LRT project

Guatemala City is the capital of Guatemala (Central America) and it has an extension of 996 km². Similar to other Latin American cities, the concentration of economic activities and jobs are predominantly centralized conforming a monocentric structure in proximity to the historic centre (HC), whilst the city periphery tends to be more sparsely occupied and expanding in an unplanned manner (Ford, 1996; Pacione, 2005). Figure 5.1 shows Guatemala City and the residential land value distribution confirming a monocentric structure. The highest values are observed in what is known as the city's central business district -CBD- (Morales et al., 2020).

Main traffic issues relate to the high mobility needs from people living at the city's periphery towards the CBD. Ongoing planning efforts aim to improve mass transport connectivity across the city towards the city core. TransMetro is a BRT system that started to operate in 2008 and has been progressively adding new corridors until the date. MetroRiel is a proposal of introducing an LRT system and it has been at the centre of attention by planning authorities. MetroRiel is a relevant case study due to the following reasons: 1) it has a metropolitan scale, aiming to improve connectivity along an axis that goes from the South-West to the North-East peripheral areas and the city core; 2) it involves the rehabilitation of the city's old rail system through some important new connections that also favour mobility by private vehicle; 3) information about the project and preliminary planning drafts were available via documentation of the pre-feasibility analyses (IDOM, 2016).

Figure 5.1 shows the 20 km segment of the old rail system that is meant to be used by MetroRiel and the location of 20 proposed access stations. The proposal also includes around 10 km of interventions on roads to facilitate transversal connectivity and two bridges (for both MetroRiel and private vehicles) to overcome the ravines between the stops "Centra Sur"- "JBR" (South-West) and "La Ermita"- "Las Vacas" (North) stations. MetroRiel is expected to have an initial demand of approximately 300,000 passengers per working day projected to increase by 100,000 during the next 20 years. It is expected that people are incentivized to shift their modal preferences towards MetroRiel not only from an affordable travel-time-reduction point of view, but due to other qualities offered by it, e.g. reliability, cleanness and safety. These are particularly important in the local context if people are to choose between using their own vehicle or to become public transport user. The operating speed is set to be at 30 km/h, translated to 45 minutes travel time from "Centra Sur" to "Centra

Norte”. The investment required is estimated to be above \$700 million. Initially the project was set to open in 2021, but it has been delayed due to bureaucratic processes and funding availability.

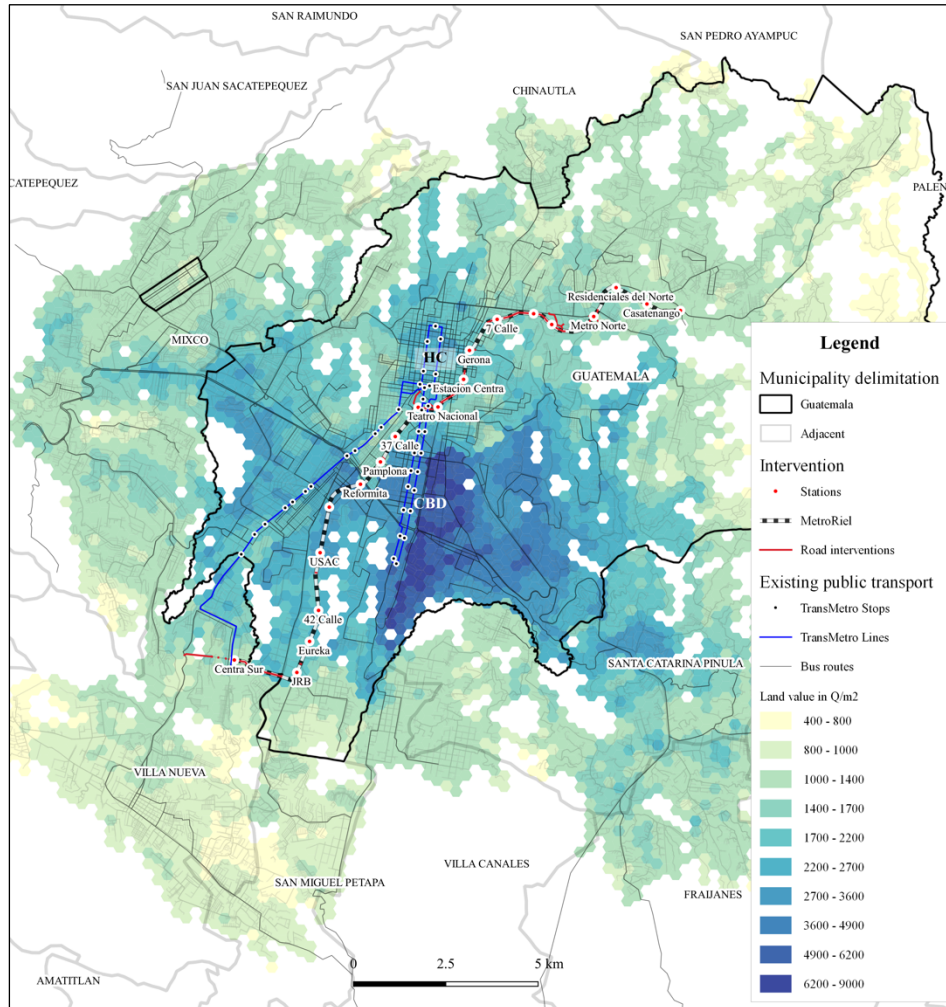


Figure 5.1: Guatemala City, administrative division, TransMetro, MetroRiel and overlay of modelled land values for the year 2014. Adapted from Morales et al. (2020).

5.2.2 Modelling framework

In order to estimate the impacts of MetroRiel, ex-ante intervention, on the residential land values we propose a modelling framework presented in Figure 5.2. The framework structures data and methodologies developed in previous research in Guatemala City (Morales et al., 2019a; Morales et al., 2019b; Morales et al., 2020).

Methods and data to compute geographic (by public and private mobility) and geometric accessibility (as analysed in SSx) were presented in Morales et al. (2019a). Development of a geometric via geographic access metric, accessibility indexes (combining private and public-based measurements into single metrics) and a residential land values dataset was presented in Morales et al. (2019b). Development of a parsimonious geostatistical model, a multivariate regression kriging (MRK), for predictive purposes and the construction of a residential land value map for the year 2014 was presented in Morales et al. (2020).

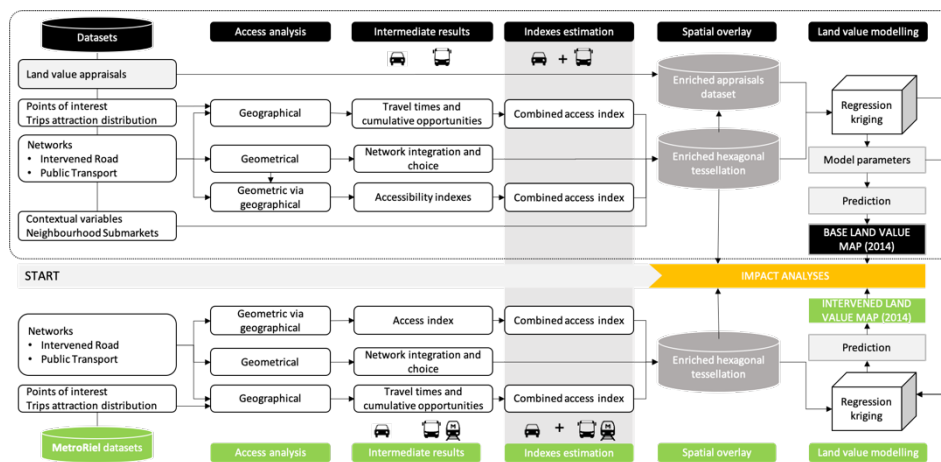


Figure 5.2: Modelling framework, pointed frame outlines methodology from previous literature.

The structuring of the methods can be described as an analytical workflow where geographic accessibility per mode of transport and geometric accessibility analyses are conducted at a city-wide scale using regular units of analysis (i.e. hexagonal tessellation) which are also observable in Figure 5.1. The intermediate results are further processed into combined access indexes that are used to enrich a dataset of land value appraisals (spatial overlay). The enriched appraisals dataset is used to calibrate (train) an MRK model. In this process the model parameters are estimated. These represent the elasticities, or semi elasticities, of land values as a function of the various accessibility values. The enriched hexagonal tessellation can be then used as an input to the calibrated MRK model to construct a map of residential land values for the year 2014.

In this paper we develop further this framework and extend its applicability to analyse potential accessibility improvements of a foreseen transport investment, in this case an LRT line, at a city scale.

An *intervened* enriched hexagonal tessellation is then used with the calibrated MRK model to construct an *intervened* land value map that reflects the improved accessibility capitalization as a function of the known elasticities. In our framework impact is defined as the analysis of the proportional land value uplifts in association with a modelled accessibility improvement due to a transport investment, before such investment is actually done.

5.2.2.1 Data requirements

Table 5.1 shows a summary of the data inputs that are required for the modelling framework operationalization. Cells filled in grey outlines the data inputs produced for this chapter.

Processing of the input data was done using ArcGIS application. Networks are built within a personal database using the same application. The process includes generation of nodes (i.e. intersections and death ends), establishment of connectivity rules between subsystems (e.g. pedestrian network, access stations and BRT) and incorporation of updates in the network.

Table 5.1: Summary of data requirements.

Input data	Description	Source(s)	Processing
Road network	Network layer prepared to analyse travel time by private vehicle and SSx centrality metrics.	Open Street Maps (OSM)	Filtering out non-drivable roads, verifying topological consistency and designation of travel time per segment. For SSx analysis the network is geometrically simplified, then segments are broken at each vertex.
Public transport network	Network layer prepared to analyse travel time by public transport. The layer includes the following interconnected networks: pedestrian network (same as road network), bus routes, the BRT lines and corresponding access stations.	OSM, drafts of bus routes and BRT lines from municipality office.	Digitalization of bus routes, BRT lines and access stations and designation of travel times per segments per system.
MetroRiel (transport service)	Layer prepared be integrated to the public transport network. The layer includes MetroRiel line access stations.	Preliminary design in official report (IDOM, 2016)	Digitalization of service line and designation of travel time per segments between stations.
MetroRiel (road interventions)	Layer prepared to be integrated to the road network. The layer includes new roads (e.g. bridges) or road modifications.		Digitalization and designation of travel time.
Points of interest (POI's)	Layer containing points locations of destinations used for geographic access analyses (i.e. banks, jobs restaurants, CBD, large scale malls, large scale groceries shops, universities, culture and hospitals).	OSM, Google Maps and Wikimapia. Official data on land use and an Origin Destination Matrix (ODM).	POI's (except jobs) were extracted from OSM and consolidated using the other sources. Distribution of jobs was approximated at each hexagon by using the ODM of morning peak hours.
Land value observations	Layer containing 1,169 points locations of real estate appraisals of residential property. Each record has attribute values for year of the valuation, plot surface area, construction area and	Reports database from private company in Guatemala City (IG).	Records were created from individual reports and georeferenced.

5.2.3 Modelling MetroRiel

5.2.3.1 Analysing new accessibility distributions

In order to model the accessibility improvements due to MetroRiel, we created a copy of the existing network layers (i.e. road and public transport network). Then, the corresponding MetroRiel data (transport service and road interventions) was incorporated to both network layers. Accessibility metrics are analysed and classified in three groups as required by the MRK: geographic, geometric and geometric via geographic access. The first one implements location-based methods and analyse the easiness (i.e. travel time) to reach POI's per transport modality (private vehicle or public transport). These analyses are implemented using the ArcGIS Network Analyst extension. Banks and restaurants are analysed as a cumulative opportunity using 10 minutes travel time threshold. Jobs access is analysed using a gravity formulation as defined by (Hansen, 1959), where the resource distribution (jobs) is approximated using the number of trips attracted during morning peak hours per hexagon.

Geometric accessibility is computed using SSx segment angular analysis (Hillier et al., 2012; Turner, 2007). These are implemented using the Space Syntax Toolkit in QGIS (Gil, Varoudis, Karimi & Penn, 2015). Two variables, integration and choice in SSx terminology, are measured at a citywide scale (r_N) and localized scales using metricized radii (0.8 km, 1.5 km, 5 km and 7.5 km). Integration analyses how close is each segment from every other segment in the network using the number of traversed segments as impedance and weighted by the angular deviation between those. Choice analysis determines the frequency of a segment to be used in shortest trips from any segment to every other segment in the network, thus suggesting the hierarchy of a road in an urban system. Geometric via geographic access is defined as access to urban locations that are highly integrated at a city-wide scale (r_N) as analysed by SSx. Similar to jobs, it is also computed using a potential access formulation. We refer to this variable as integration gravity and consistent with the terminology used in Morales et al. (2019b).

Intermediate results of the geographic access analyses per mode of transport are standardized and aggregated using equation (5.1) in order to produce combined access indexes (Morales et al., 2019b). The reference values are the maximums and minimums per access variable, per mode of transport without the intervention that were available from previous research. The metric reflects the easiness to reach a given destination or a resource (i.e. jobs or integration r_N) by the combined availability and travel speed of roads or public transport

(A_{priv_i}, A_{pub_i}). The equation incorporates a ponderation of the transport mode at location ($perc_{priv_i}, perc_{priv_i}$). This approach contextualizes access locally to what users on average benefit the most from. The percentages are based on the modal split of trips generated during peak hours as found in the ODM's data.

$$combined_access_i = (A_{priv_i} * perc_{priv_i}) + (A_{pub_i} * (100 - perc_{priv_i})) \quad (5.1)$$

5.2.3.2 Using an MRK model to construct an intervened land value map

The existing MRK model was calibrated with 876 observations of real estate values. Table 4.2, last column, lists the 23 variables that are required to produce a prediction, a short description of each and the elasticities or semi-elasticities (i.e. model coefficients). The MRK's accuracy was reported to explain 81% of the variability of observed land values (training data) and 68% of the variability at non-observed locations (i.e. land values that were not used to train the model). Prediction of land values (\hat{y}) at a specific location (s_0) is given by equation (4.2) (Morales et al., 2020). The first component is an estimated constant intercept ($\hat{\beta}_0$). The second component is a summation of the products between the values of the improved access variables (x), e.g. access to CBD, and the corresponding estimated elasticity ($\hat{\beta}$). The third component is an interpolation of the residuals from the observed land values weighted by their proximity and spatial distribution. This is equivalent to a universal kriging process. This component is responsible for modelling the spatial dependence.

$$\hat{y}(s_0) = \hat{\beta}_0 + \sum_{k=1}^p \hat{\beta}_k \cdot x_k + \sum_{i=1}^n w_i(s_0) \cdot \varepsilon(s) \quad (4.2)$$

We use the intervened enriched hexagonal dataset which contains the intervened accessibility values and all the other required variables, held constant, to run the calibrated MRK model. The output is an intervened land value map for the year 2014. We considered constructing a land value map for the year 2023 (assuming the intervention will be operational by then). However, this would make it more difficult to differentiate the value gain due to the intervention and the one due to the increment over time, as shown by the "year" variable coefficient.

5.2.3.3 Impact analyses

In order to analyse the impacts for both the accessibility and the land values we decided to compute proportional value changes. We use equation (5.3), where V takes the value of the accessibility metric or the land value. The results are expressed in percentages over the

original value (not intervened). This approach facilitates the visualization and understanding of the benefits of MetroRiel in proportional units.

$$\text{change index} = \left(\frac{V_{ex_MetroRiel} - V_{post_MetroRiel}}{V_{post_MetroRiel}} \right) * 100 \quad (5.3)$$

5.3 Results

5.3.1 Modelled effects on accessibility improvements

Figure 5.3 shows the proportional geographic accessibility improvements to selected POI's (CBD, shopping malls, hospitals, large scale groceries, universities) and the geometric via geographic access improvement (i.e. integration gravity). The figure confirms the obvious expectations that improvements are to be found around the access stations and along the corridor. However, it also reveals that the magnitude and spatial extent of such improvements is heterogeneous. MetroRiel could improve significantly access to the CBD at the areas (beyond 4 km) serviced by the North-East and South-West stations.

Access to other POI's is slightly less, except for access to shopping malls. Access to such destinations seems to improve significantly not only along the corridor and North-East areas, but also at the Historic Centre (HC). The cross-sectional layout of TransMetro and MetroRiel at the core (e.g. "Teatro Nacional" and "Estación Central") offers a type of intermodal centrality that effectively extends the possibilities of public transport modality, facilitating access to facilities such as the shopping malls which are not available at the HC.

Improvement of integration gravity accessibility is tied to the new distribution of geometric access, namely the SSx global integration (r_N). Figure 5.4 shows proportional improvements in terms of SSx integration at selected spatial scales. Improvements are produced by road network modifications or additions that lead to better connectivity at a neighbourhood level and/or at a city-wide scale. At smaller scales, benefits could contribute to stimulate pedestrian movements. At wider scales benefits could stimulate consolidation or allocation of commercial and service activities, hence contributing to the attractiveness of places as destinations for users.

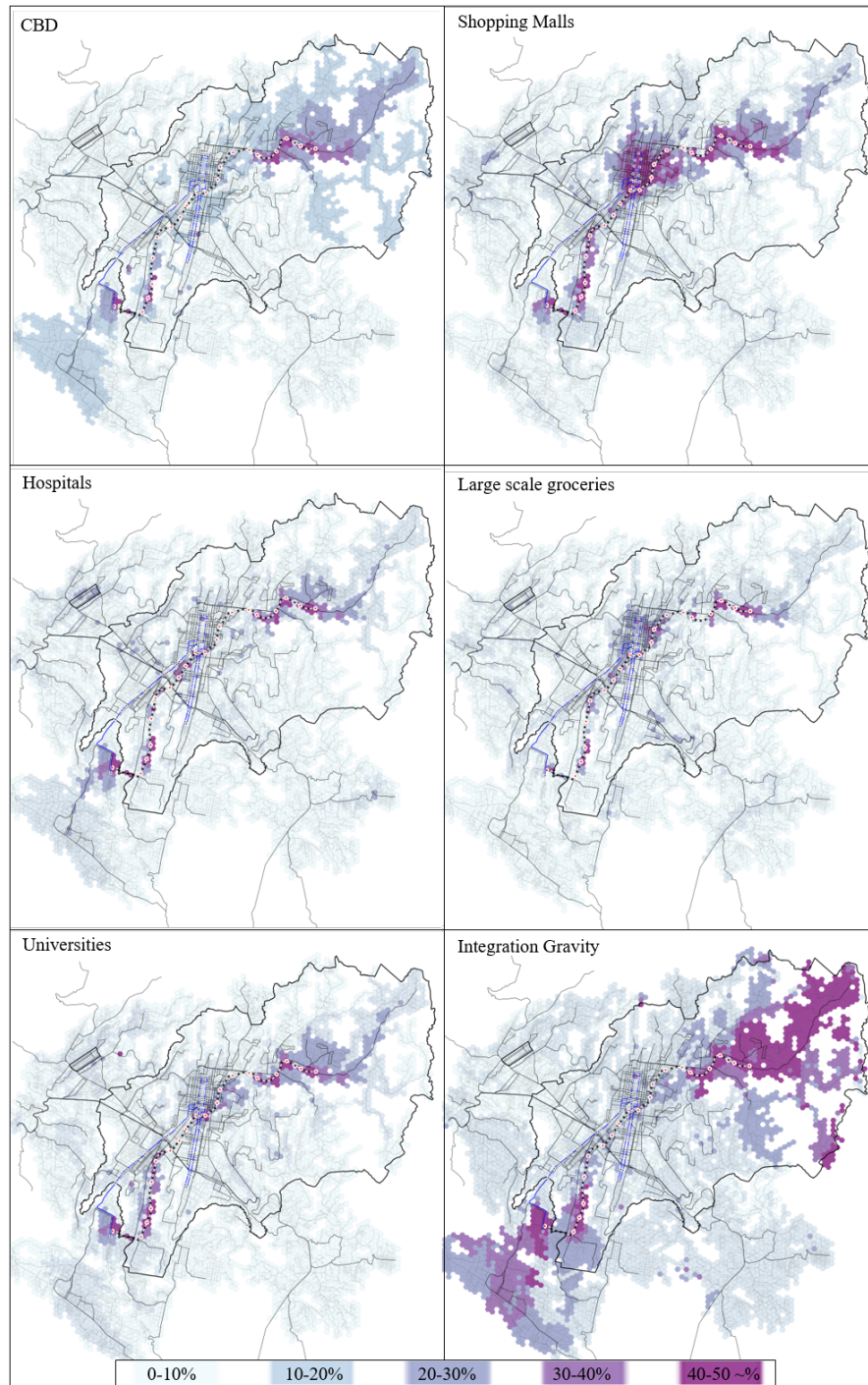


Figure 5.3: Proportional accessibility improvement for selected variables.

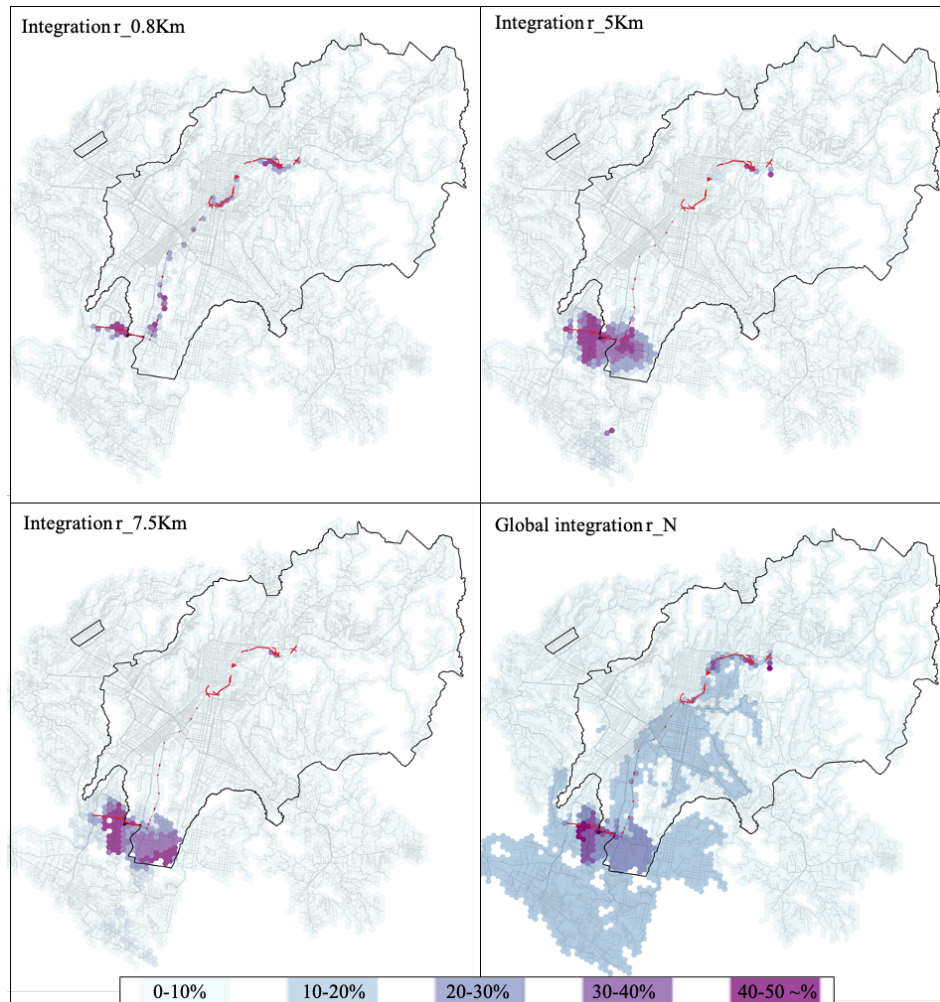


Figure 5.4: Proportional geometric accessibility improvements at selected radii. Road interventions in red.

The new bridges that aim to facilitate access to the station “Centra Sur” and connect with the station “JBR” (South-East extreme) seem to bring significant connectivity benefits across all the spatial scales. Smaller transversal connections along the corridor improve connectivity at a very localized scale ($r_{0.8}$ km). As a whole, the composite of road interventions seems to favour a larger spatial extent at a city-wide scale (r_N). This is reflected in the integration gravity access improvements (Figure 5.3) where both extremes of MetroRiel favour from access to a moderately more integrated city core.

5.3.2 Residential land value uplifts after MetroRiel

Figure 5.5 shows the land value map after modelling the MetroRiel intervention. A monocentric structure of the residential land values prevails. The intervention induces a value uplift around the CBD as some areas are shifted to the next range of values (Q 6,200 – 12,000). The monocentric structure also seems to extend towards the direction of the HC. MetroRiel seems to have a somewhat generalized effect of value uplift in central areas and along its corridor without a noticeable modification of the city-wide structural patterns of residential land values. This observation is similar to those reported in Viguie et al. (2014).

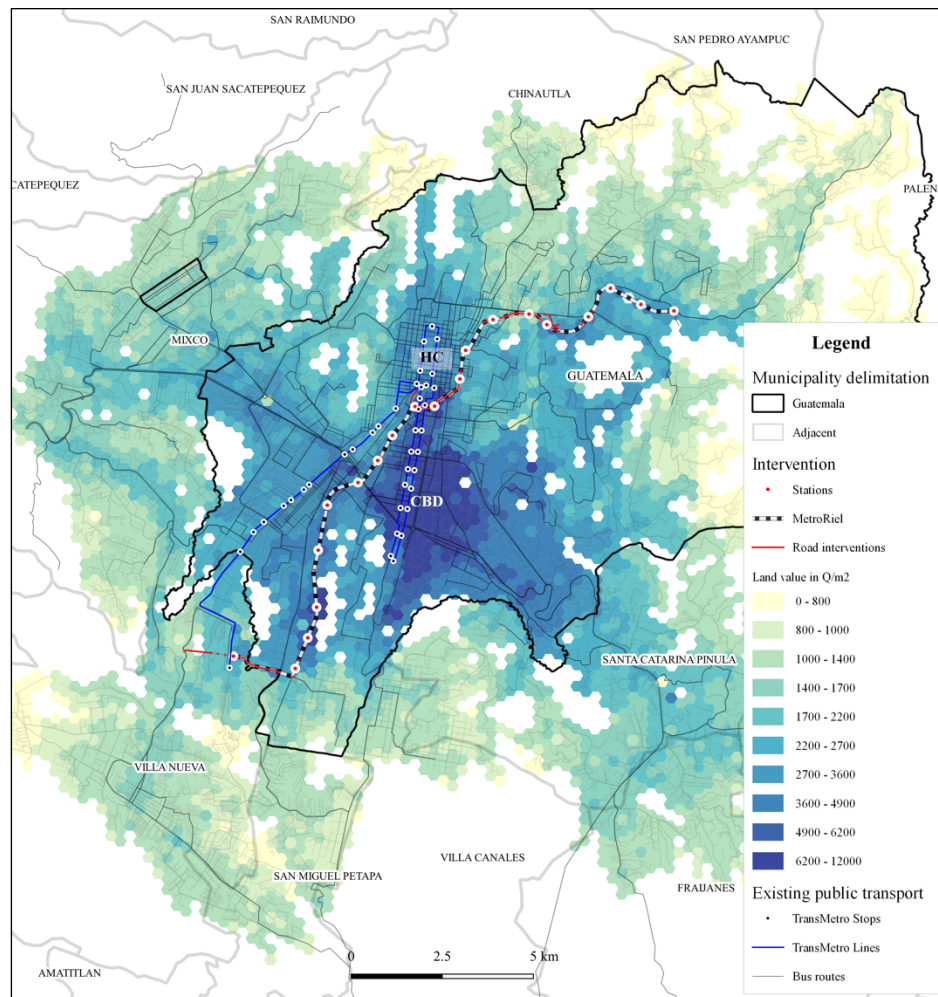


Figure 5.5: Land value map after modelling MetroRiel intervention

Figure 5.6 shows the proportional value uplifts with reference to the land values constructed for the year 2014 (see Figure 5.1). The spatial patterns of such uplifts are plausible since those can be expected to be in close proximity to the MetroRiel particularly at locations neighbouring the stations. However, the spill of the gains goes beyond such proximity in ways that would have been impossible to determine with a limited spatial scope (e.g. 1 km around the intervention) and without the comprehensive accessibility modelling.

The figure shows that the uplift is also shaped by the availability of existing public transport services (i.e. bus routes or TransMetro). Value gains due to MetroRiel cannot be expected due to the availability of the new service alone. But, when observing the spatial pattern of the uplift spill, we deduce that existing infrastructure plays an important role in facilitating access to MetroRiel's stations. Such situation would then indeed reduce travel times in trip segments to reach assumed desired destinations and hence induce a value uplift. This becomes particularly important for the propagation of value gains beyond the stations at the North and South extremes of MetroRiel. Similarly, we observed an interesting interaction between the layout of TransMetro at the HC and the proportional value uplift.

Overall, the proportional value uplift at a city level ranges widely. Peripheral areas on the West (i.e. Mixco municipality) and South-East (e.g. Santa Catarina Pinula) barely benefit from the interventions. There, the proportions range from 0% up to 10% and 20% in some cases where existing public transport services are also available. Areas in proximity to MetroRiel show higher proportional uplifts ranging from 20% up till 100% increase and higher up to 134%.

The spatial distribution and value gains seem to be heterogeneous across the adjacent areas and along MetroRiel. Figure 5.6 shows that the average proportional uplift ranges between 20-40% up to 100%~ for the commonly referred catchment area of 0.5 km around each station. We decided to look into this more in detail by analysing the effects explicitly around the stations. Figure 5.7 shows a comparative plot of curves indicating the average proportional value uplift within different Euclidian distance ranges from MetroRiel stations. The averages were calculated using the prediction locations (hexagon centroids) falling within specified distance thresholds (e.g. 1 - 1.5 km) per station and as a total (i.e. using dissolved buffers). Generally, the curves show a plausible pattern where the value uplift is the highest for locations closest to the stations and then decreases systematically as being farther away. The total average (black dotted line) shows the highest proportional uplift of around 76% for locations within 0.5 km

to the stations. The uplift decreases to almost 40% for locations in the next distance range (0.5 – 1 km). The lowest uplift (~15%) is still observed at the farthest distance range analysed (3.5 – 4 km).

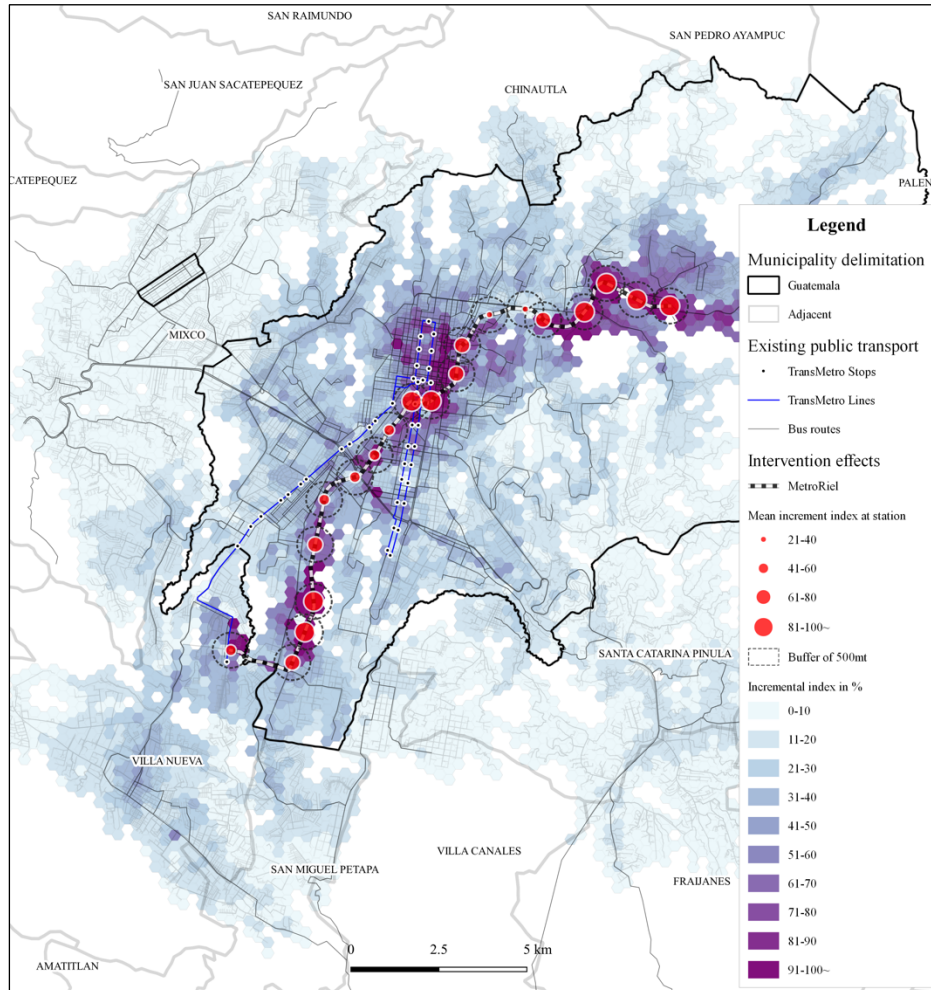


Figure 5.6: Index of residential land value increments.

However, the individual curves show high heterogeneity of the effects along the MetroRiel corridor at different distance scales. This holds for locations within the 0.5 km radius and beyond up to 2 km where the uplift ranges from 17 to 50%. We emphasized the following 7 curves: "Centra Sur" and "42 Calle" located at the South-west side; "La Terminal", "Teatro Nacional" and "Estación Central" located along the middle section; "La Ermita" located at the North-east and "Centra Norte" being the last station. Locations in proximity (0.5 km) to "Centra Sur" on average show relatively medium uplift effect (60%). This might

be unexpected, since it would work as a multimodal node where access to TransMetro is also possible. However, by closer inspection we observe that accessibility benefits for this area are lesser compared to the North-East stations (e.g. "Centra Norte" – 105%) since TransMetro already provides relatively improved access compared to existing bus routes.

In the accessibility modelling, travel times to certain destinations (e.g. shopping malls) and potential access to geometric resources (i.e. integration gravity) might tend to be comparable between TransMetro and MetroRail. Hence, access improvements, translated to value gains, are proportionally less. This is not the case for stations such as "42 Calle", which seems to effect immediate locations with the maximum uplift (134%) followed by the "Eureka" stations. Although, the uplift dramatically decreases as we observe locations in more distant buffers.

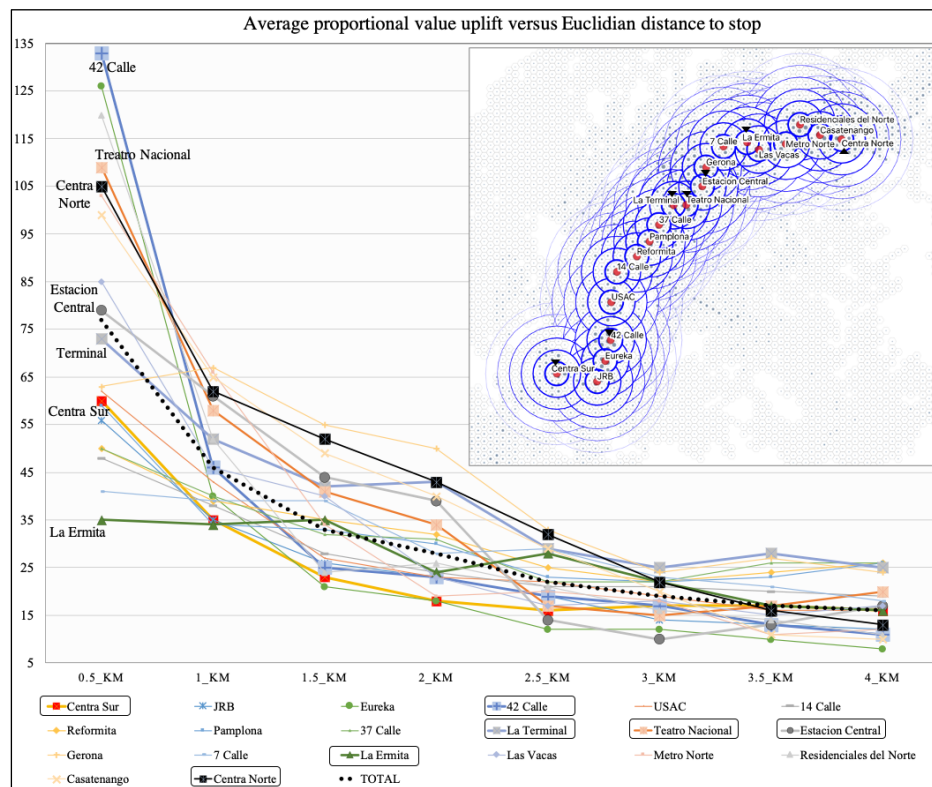


Figure 5.7: Comparison of average proportional land value uplift per station and average total against Euclidian distance ranges from MetroRiel stations

Stations located at central areas show a medium-high uplift effect. Locations within the 0.5 km radius around the station “La Terminal” show an uplift of 73%, while locations around the station “Teatro Nacional” show an uplift of 109%. At these stations the available transport inter-modality (i.e. the MetroRiel and TransMetro interaction outlined earlier) does significantly improve the accessibility of the HC area to facilities such as shopping malls. “La Ermita” shows the lowest proportional uplift (35%) for locations within 0.5 km distance. Yet, it does not decrease beyond that as it can be seen from the locations further away. Figure 5.7 shows great deal of heterogeneity of the effects at locations beyond the 0.5 km service radius as such effects decrease at different rates per station. Yet, we can reiterate that stations on the North-East side such as “Centra Norte” are perhaps the ones that have the highest uplifts followed by stations at central areas.

5.3.3 Comparison of the total land stock value uplift and the required investment for MetroRiel

Figure 5.8 shows the estimated total land stock value before intervention, the estimated stock value uplift and the difference. All values are classified on the horizontal axis by the Euclidian range distances to the MetroRiel access stations. The plot is built assuming that 70% of each hexagon area (54,560 mt²) is dedicated to residential land and it is subdivided in parcels of 350 mt² surface (model assumption). The total stock value of the residential land ascends to Q 20.86 B for parcels within a 0.5 km distance from MetroRiel stations, Q 31.32 B for parcels within a 0.5 to 1 km distance and so on.

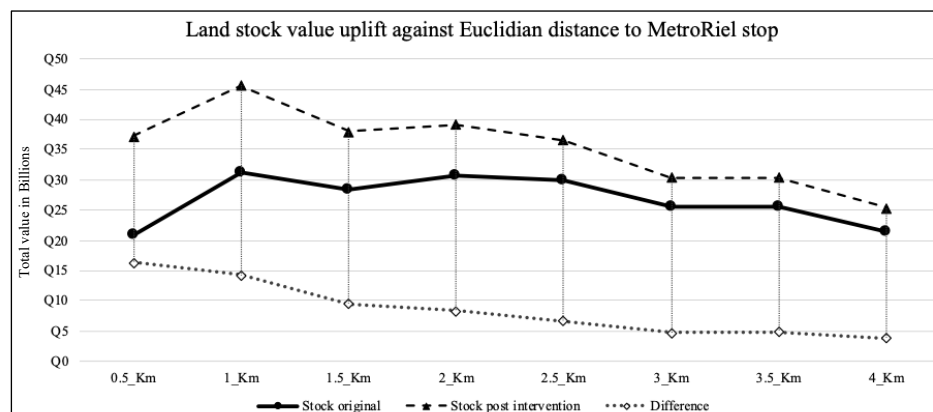


Figure 5.8: Comparison of total land stock value uplift before and after intervention.

The different land stock values, defining the shape of the curve across distances to the stations, are defined by the city-wide structural

pattern of land values. This explains the fact that land stock value in the close proximity to the intervention is relatively cheaper than farther away. The estimated total land stock value without the interventions ascends to Q 214.12 B. The curve plotting the stock value post-intervention shows again higher uplifts on the stock in close proximity to the stations and then the uplifts decrease in magnitude, also shown by the “difference” curve. Yet, the curve is observed as a somewhat upward parallel of the current stock value curve. Hence, the value uplifts do not imply a change in the structural pattern of land values.

The total difference (i.e. net uplift) between the original stock value and the predicted value post-intervention ascends to Q 68.67 B. This is equivalent to \$ 8.87 B at an exchange rate of 7.74. If we consider that the spatial model has an accuracy of 68%, we can penalize the value by multiplying it with a factor of 0.32, which yields a stock value of \$ 6.03 B. Assuming that financial mechanisms could aim to capture at least a fourth of such value, this would ascend to \$ 1.51 B. This is equivalent to 200% of the estimated investment (\$ 700 million). An even more modest scenario would be to assume that only the stock of properties within 1.5 km radius are to be considered for value capture mechanisms, since the total average uplift in this range is higher than 30% (see Fig. 5.7). Using the same logic as before this would yield a value recoverable value of \$ 0.88 B, 126% of the total estimated investment.

5.4 Discussion

Overall, accessibility improvements are observed as a combination of the following: significant travel time reductions from a given location towards the POI’s when comparable public transport alternatives are limited or not available; the distribution of the POI’s itself (e.g. the case of shopping malls); interactions with existing public transport availability when those effectively extend the coverage of public transport network; localized relevance of access improvement (per transport mode) relative to user’s preferences (i.e. access weighting factor in the combined access index calculation). It is worth mentioning that we kept the modal split between private vehicle and public transport users as it was before the intervention (inherited from the observed ODM data). In turn, selection of public transport service is modelled purely as a function of travel time reduction. Yet, incorporating the expectations that users would shift their modal preference towards public transport and particularly to MetroRiel would outline slightly higher accessibility improvements.

The results indicate that the implementation of LRT system in Guatemala City could induce a significant residential land value uplift.

In line with empirical literature analysing data collected post-intervention, the effects seem to be the highest at locations within 0.5 km distance from the LRT stations (35% up to 134%). The effect is less as locations are further away from stations, but it is still observed at a distance range between 3.5 to 4 km and beyond. The spatial pattern of the value uplift spill is partly shaped by the layout of existing alternatives of public transport, namely buses and a Bus Rapid Transit (BRT) system. The heterogeneity of how the LRT line induces value uplifts around the stations is presumably associated with the availability (or lack) of alternative transport choices. In some cases when travel times to certain destinations are comparable between transport modalities (i.e. BRT and MetroRiel), the value uplift might be lesser, than when MetroRiel represent an actual better transport option. In other cases, the interaction between existing BRT and MetroRiel, such as in the historic centre, seems to extend significantly the easiness to reach certain destinations and therefore impacts on the capitalization of land value. Furthermore, the results show that comparatively already one fourth of the estimated total value uplift of the residential land stock is comparable to the required investment to implement the LRT system. In the calculations we take into account the prediction accuracy of the model, meaning that we only consider 68% of the total land stock value uplift.

There are important limitations with the modelling framework that should be borne in mind as part of an integral perspective of how transport investments impact cities. The timings stages (i.e. after announcement, during and post-construction) at which capitalization might occur, as suggested in literature, are not addressed in our approach. Further, other potential wider effects that could be triggered due to access improvement and land value uplift such as land use reorganization (e.g. shifting to commercial land uses and/or densification), inward investment and gentrification processes are also not addressed. However, such effects could eventually also contribute to redraw the land market.

Broadly, the analytical workflow is reproducible to other case studies in the context of some considerations. First, as with any modelling strategy, some decisions need to be made regarding the specificities of each case. A combination of existing literature and localized expert knowledge could help to preliminarily determine a list of destinations (POI's) to which access is expected to explain the land or property value. Second, decisions regarding the statistical modelling approach should be based on the performance, the problems to tackle in the data structures (e.g. the spatial dependence) and the intensions to ease the communication of the model insights (e.g. choosing between a

regression model and a more sophisticated learning algorithm). Thirdly, data availability is paramount. An extensive amount of time was consumed in the collection and preparation of the data inputs particularly for the land value observations. Access to volunteered geographic information (e.g. Open Street Maps) could help greatly to overcome this limitation, with the consideration that its availability varies across regions. Yet, perhaps the readily availability of data on property or land values pose a more restrictive situation in terms of access, quality, number of observations and level of aggregation. This condition would normally have implications in the selection of the statistical method and the reliability of the results. Lastly, the technical operationalization could also pose a time-consuming burden. Some data collection, pre-processing (e.g. network topology) and processing (e.g. responsive accessibility analyses) could be automated by means of an SQL database implementation with a spatial and network routing extension (e.g. PostGIS and PGrouting). This could make possible replicating the framework and applying it for a systematic and dynamic scenario-based evaluation.

We suggest that the applications of the modelling framework might be diverse and could provide a bridging instance between planning practitioners, land administration authorities, landowners and the real estate developers. The framework is suggested to be utilized not only to inform transport design activities (e.g. optimizing accessibility benefits) but to understand one of the wider economic impacts, the land value uplifts (i.e. inputs for a CBA). Later, information and collaborative knowledge production among the aforementioned actors could lead to the construction of favourable setups for the feasibility and sustainability of investments on large transport infrastructure. We outline the formulation of financial mechanisms to recover accessibility capitalization (i.e. collaboration between planners and taxation authorities), facilitating public-private investments and stimulating private investment for redevelopment among others. In parallel, the framework could also be of interest for collaborations between urban designers and real estate developers. The empirical construction of the modelling framework with a city-wide spatial scope allows a “learning from the city” approach (i.e. estimation of elasticities) with application to scenario-based assessment of neighbourhood scale investments. For example, development of master plans with variations of the land uses (e.g. POI’s) and road or public transport connectivity schemes.

5.5 Conclusions

In this chapter we present and implement a modelling framework to estimate the effects of a Light Rail Transit (LRT) system on the residential land values of Guatemala City. Our contribution is threefold.

First, we structure data and methods from recent research into an analytical workflow to model land values. The framework relies on a comprehensive operationalization of accessibility (i.e. geographic access, geometric access as analysed in Space Syntax and geometric via geographic access) and a geostatistical approach (i.e. Multivariate Regression Kriging). Second, we extend its applicability and utilize the workflow to model the expected accessibility improvements of the LRT intervention. Residential land value uplifts are then predicted as a function of such accessibility improvements. Third, by using a city-wide spatial scope, the results of the analyses provide detailed insights on the heterogeneous spatial distribution of the effects (improved accessibility and value uplifts) along the intervention corridor and across the city.

We suggest ideas for future research. Including other land uses in the modelling framework would allow to understand how other land uses capitalize after transport investments. The modelling approach could be a comprehensive reference to continue the development of econometric models embedded in broader modelling architectures such as those operationalized in the land use – transport interaction (LUTI) models. This could be a research direction to look into in order to address how capitalization occurs in time. From a procedural perspective, the operationalization of the entire modelling framework could be further automated aiming to facilitate a dynamic computation of scenario-based results and ease its implementation in practice.

Finally, our modelling framework could bring together the overlapping interests of planning practice and land administration processes (e.g. taxation) when facing the challenges of sustainable transport investment such as constant budget limitations. The analytical approach shows to be an efficient and plausible approach to estimate the effects of city-wide transport investments. Since the approach is mostly based on modelling the value as a function of accessibility, it could allow to preliminarily anticipate land value uplifts of relatively new transport technologies that are not yet implemented in a city, such as the case of MetroRiel. We conclude that the spatial information produced can be further processed and used as an input for a CBA evaluation and the formulation of financial mechanisms to support economic feasibility of transport investments.

Chapter 6

Summary of main findings and reflections

6.1 Introduction

Scholars are becoming increasingly interested in finding the relations between sustainable transport modes and a territory's economic structures. Academics, echoing practitioners, outline various motivations for the evaluation of large transport investments regarding their effects on the economic value of land, particularly value uplifts. While the research strand analysing empirically how such effects, after interventions are delivered, is internationally fertile, development of analytical methods to estimate those effects ex-ante interventions are rare. The research gap is described as the need to adaptively integrate state-of-the-art knowledge from the research domains of urban access and land value modelling while expanding the spatial scope from a transport catchment area to a city-wide scale. The overarching goal of this research was to propose and implement a modelling framework to estimate the spatially distributed effects of future transport infrastructure on land values by means of operationalizing a comprehensive accessibility definition into a predictive land value model.

The remainder of this chapter is organized as follows. Section 6.1 summarizes the main findings from Chapters two, three, four and five in the light of the research objectives and questions, as well as the overarching goal, formulated in Chapter one. Section 6.2 presents reflections about the contributions of this dissertation to the literature and knowledge, to planning practice and to the case study. This section also presents prospect for further research.

6.2 Summary of main findings

6.2.1 Chapter 2 - objective 1: To compare location-based methods and Space Syntax for mapping urban accessibility in two cities in Guatemala.

Chapter two addressed the first research objective and provided answers to the following questions:

- 1 How to measure accessibility at the city and neighbourhood scales while accounting for data scarcity?
- 2 What are the relations between Space Syntax and urban access to various destinations as a first step to evaluating its applicability to explain variations of land values?

In addressing those questions, the chapter set the ground for the adaptable and comprehensive operationalization of an urban accessibility definition. This was essential for the overarching research goal since the multiple access metrics mapped for Guatemala City represent the basis not only to later explain how these determine land values (Chapter 3), but also to construct a land value map (Chapter four) and furthermore to model how transport intervention improves accessibility and consequentially induces land value uplifts. Two concepts of urban access - namely geographic and geometric access - are defined in detail. Correspondingly, location-based and Space Syntax methods were thoroughly described in the context of the state-of-the-art in analysing both types of urban access.

The chapter addressed the first question by describing in detail how geographic accessibility metrics to various Points of Interests (POI's), commonly assumed to be relevant to determine land values, were mapped for Guatemala City. Additionally, a second city in the same country was included in the study, but with different degrees of urbanization and spatial scale. A tailored approach classified the POIs based on their relevance to determine location quality at a city or neighbourhood scale (macro and micro location correspondingly) and implemented location-based analyses per mode of transport. Micro-location POIs were analysed as cumulative access (i.e. number of facilities within a travel time threshold). Macro-location POIs were analysed using minimum travel time. The approach predominantly relies on extracting available data from Volunteered Geographic Information sources (VGI). Analysing geometric accessibility was demonstrated to be significantly less data intense compared to the geographic analyses. Space Syntax metrics were mapped at various spatial scales (city-wide and neighbourhood scales) via metricized radii and using a representation of the road networks as the only input data. A tessellation of hexagons was produced to aggregate all the access metrics (later on predictor variables), which was furthermore also the basis for land value modelling in Chapters three, four and five.

Answers to the second research question were obtained by testing the hypothesis of whether Space Syntax metrics could consistently reflect accessibility conditions that have, so far, only been addressed through location-based methods and using time-based impedances. Pearson correlations, at a 99% confidence interval, revealed that Space Syntax metrics are consistently comparable with location-based access metrics to the selected POIs. This provided first insights about possible application of Space Syntax to modelling land values in Guatemala. It was observed that important city structuring roads within the network, as identified in Space Syntax, do play an important role in the spatial

distribution of geographic access to multiple destinations. Geographic access to POIs relevant at the city scale tends to be associated with areas that are highly integrated from a geometric perspective at radii of 7.5 km or more. In turn, areas that are highly geometrically integrated at smaller radii of analysis (e.g. 0.8 km or 2.5 km) are more associated with geographic access to POIs that are relevant at the neighbourhood scale. Moreover, it was consistently observed that geometric access at localized scales tends to reflect geographic access limitations either imposed by the analysis (i.e. travel time thresholds) or by the limitations of the mobility mode (e.g. longer travel times by public transport).

The results expand the scope of accessibility knowledge derivable from limited data availability using Space Syntax, valuable for its applicability in data-scarce contexts. Significant correlations between methods and across the two cities not only validate the applicability of Space Syntax but also the applicability of VGI data to mapping urban access. However, a Space Syntax metric with an unrestricted metric radius (i.e. city-scale) might not be able to consistently reflect the ongoing land-use processes in less-consolidated areas. It was not possible to conclude from the results which Space Syntax spatial radii best correlates with the distribution of combined geographic access of those POIs relevant to a city-scale. This was one important drawback of SSx analysis, as there is not enough evidence to suggest which spatial radii correlate better with what in the different cities.

6.2.2 Chapter 3 - objective 2: To bridge concepts and definitions to comprehensively address accessibility to uncover its relations with residential land-values in Guatemala City.

Chapter three addressed the second research objective and provided answers to the following questions:

- 1 How to combine Space Syntax and location-based methods to explain the variability of land values in Guatemala City?
- 2 What are the elasticities of land values as a function of urban access in Guatemala City?

In doing so, the chapter brought together state-of-the-art concepts and methods from the fields of urban accessibility and land value modelling. In the context of the overarching goal, this chapter narrowed down the focus to the land value modelling task and introduced the reader to a constructed dataset of residential land

values in Guatemala City that is central to this dissertation. The chapter introduces the following two hypotheses: (a) addressing the disparity of geographic-access opportunities due to available transport modes and the geometric-access at various spatial scales could contribute to an increased capacity to explain land-values variability and prediction accuracy; and (b) geometric- accessibility capitalizes land not on the basis of location, but also as a reachable resource that can be accessed by means of geographic access. Thus, geometric via geographic accessibility was defined and implemented as the potential ease of reaching geometric access by means of private or public transport.

To provide answers to the first question, hypothesis (a) was tested. To do so, the chapter elaborated on the formulation of a multivariate linear regression model that integrates three types of accessibility metrics: time-based geographic access, geometric access and geometric via geographic access. Geographic access metrics per transport mode, already produced in Chapter two, were further standardized and combined per destination (i.e. per POI). The combination considered, in the form of weights, the spatial distribution of modal split. This allowed us to address the assumption that appreciation of access opportunities as a function of transport mode varies across the city given the current modal preferences. Combining the metrics per POI also serves the purpose of reducing the number of variables that could potentially introduce severe problems of multicollinearity. Furthermore, the formulation of the model also considers the inclusion of additional contextual variables reflecting submarket conditions (e.g. concentrations of new residential offer), neighbourhood characteristics (e.g. socioeconomic groups) and even geographical coordinates as a first attempt to deal with a confirmed problem of spatial dependence.

Essential to test the first hypothesis was to focus on the performance of the model in terms of goodness of fit to observed land values (i.e. adjusted R^2 and Akaike Information Criterion), characterization of autocorrelated errors and goodness for prediction using non-observed land values (test data not used for training the model). The model performance was then compared with the performance of alternative models (e.g. leaving out geometric access, using only private mobility-based access metrics, using the potential access to jobs instead of access to CBD and vice-versa, leaving out submarket variables). The findings revealed that the performance of the proposed model was higher than that of the alternative models. The results allowed us to accept hypothesis (a) and the model set a core reference for the research on how to operationalize a comprehensive definition of urban access in a land value model.

The chapter shifted its focus on the inferences that can be drawn from the estimated model in order to provide answers to the second question and to accept hypothesis (b). This was described as the first step in unveiling a data structure that helps us to understand how multiple access metrics explain land values variability and later can be extended for predictive purposes (Chapter 4). The findings show that geographic access to the core business district has the highest impact on land-values, followed by proximity to urban areas with high geometric-access, measured as geometric via geographic access. It is suggested that potential access to vital urban areas as identified via Space Syntax denotes the presence of economic activities, or the potential for such, which were not explicitly addressed through the geographic-access metrics. Furthermore, geometric accessibility at the neighbourhood and city-wide scales add spatialized information that contributes to a parsimonious model. It was concluded that Guatemala City land values have follow a predominantly monocentric structure.

However, the results have some limitations. It is suggested that some parameters used to produce individual access metrics in Chapter two (i.e. distance decay parameters to estimate potential access to jobs) and the combined access metrics presented in this chapter (i.e. weights reflecting modal split across the city) could be further tested systematically to fine-tune the ability of such variables to explain land value variability. Furthermore, similarly to adding submarket variables, it was evident that geometric access metrics produced by Space Syntax do contribute to partially removing the spatial dependence problem. However, more complex model approaches need to be considered to address such problem and before using the model for predictive purposes. Although Guatemala City shares important typological features with other cities in Latin America, it is not suggested that the modelling approach is directly transferable without adjusting its formulation using localized knowledge and addressing each city's particular characteristics.

6.2.3 Chapter 4 - objective 3: To construct a land value map by means of a geostatistical approach using Space Syntax and a spatialized variable selection.

Chapter four addressed the third research objective and provided answers to the following questions:

- 1 How do Space Syntax-based metrics add relevant spatial information to model land values after accounting for spatial dependence?
- 2 What is the spatial distribution of residential land values in Guatemala City?

In doing so, Chapter four described the construction of a land value map by means of a geostatistical approach incorporating Space Syntax metrics and securing model parsimony under spatialized modelling conditions. A geostatistical approach was chosen, rather than a spatial econometric one, since the empirical literature recommends it as a preferred approach for predictive purposes. This chapter addressed two knowledge gaps. First, this was the first time that a land value model incorporating Space Syntax metrics has been approached by explicitly dealing with the spatial dependence problem. Second, model parsimony was achieved by implementing a variable selection procedure via the geostatistical approach. The knowledge produced in this chapter is central to the overarching goal since the predictive model underpins the construction of two important outputs: the base land value map (presented in Chapter 4) and an “intervened” land value map reflecting land value uplifts (Chapter 5).

The implementation of the spatialized variable selection and the estimation of the resulting parsimonious model provided answers to the first question. The findings reveal that a more parsimonious model (i.e. requiring less data) with higher performance, compared to the model presented in Chapter three, can be achieved when addressing spatial dependence during the variable selection process. The amounts of information added by a predictor to a non-spatial model might be high. However, information contribution might turn out to be less meaningful and lead to predictor removal in the context of a geostatistical model. Compared to some predictors associated with neighbourhood scale quality (i.e. micro location variables defined in chapter one), the modelled spatial structure contributes with more meaningful information to explain local land value variability. The summary of the results of the spatialized variable selection procedure also revealed important insights. It was observed that the geometric via geographic access metric contributes with spatial information that seems to be more important to the model than the geographic access to the CBD. This is relevant since such a metric has traditionally underpinned the literature on property and land values. The findings do not suggest that geometric via geographic access, as proposed in this dissertation, should be considered as more important than access to the CBD. However, for the case of Guatemala City, it seems to be capturing additional spatial information embedded in the geometric

characteristics of the urban layout which has greater impact in explaining residential land value variability compared to the CBD access metric. Moreover, the spatial distribution of the new metric, already presented in Chapter three, does not conflict with the conclusion of a monocentric city structure but the other way around. Furthermore, geometric access metrics do contribute with statistically significant information under spatialized modelling conditions, especially at restricted spatial scales (e.g. 0.8 km, 5 km and 7.5 km).

Answers to the second questions were given after utilizing the model for predictive purposes and then constructing a baseline land value map (i.e. reflecting the situation without any transport intervention). The map was constructed assuming average characteristics of a residential property as observed in the input land value dataset. This means that the map shows a prediction of a value per unit of land (m^2) of an average property that has a parcel surface of 300 m^2 , a built-up area of 250 m^2 , has a regular rectangular geometry and it is not a corner property. The constructed map confirms that the spatial distribution of land values in Guatemala City is predominantly monocentric. This observation is aligned with the findings presented in Chapter three, the neo-classic urban economic theory and theoretical city models of other Latin American Cities. Thus, the highest values (around \$ 1200 per m^2) are found around the CBD area. The CBD is the one that benefits the most from all types of accessibility (geographic, geometric and geometric via geographic). The land value map allowed us to recognize some other areas of relatively high value that could suggest some degree of polycentric structure: the areas with high-income residents that benefit from overall good access, nodal areas with intense commercial uses, nodal areas with less intense commercial uses or those known for the concentrations of what is locally known as informal economies (e.g. markets). The land value structure reflects the effects of the geometric connectivity and continuity observed in central urban areas versus the discontinuity and less consolidated areas at the periphery.

The unavailability of data at the parcel level turned out to be one of the major limitations of the results reported in Chapter 4. That is, having to construct a land value map using the average characteristics of the residential properties as observed in the input dataset. This means that the heterogeneity of residential properties across the city is not accounted for. Furthermore, although the predicted map is presented as a continuous surface, in reality variability of land values is tied to the parcel boundaries themselves. In parallel, the inspections of the spatial error variance of the predictions outlines the potential benefits of including more input observations to train the geostatistical

model. Finally, similarly to the limitations of previous chapters, it cannot yet be claimed that similar findings are to be expected in relation to the applicability of Space Syntax metrics until the approach has been replicated in other cities.

6.2.4 Chapter 5 - objective 4: To propose and operationalize a modelling framework to estimate the residential land value uplifts if introducing a Light Rail Transit system in Guatemala City.

Chapter five addressed the fourth research objective and provided answers to the following questions:

- 1 How to model the spatially distributed effects of a Light Rail Transit system on the residential land values?
- 2 Where and what is the potential residential land values uplift that could be expected due to introducing a Light Rail Transit system?

Chapter five was an integrative chapter that articulated the findings and developments from Chapters two, three and four in order to achieve the overarching goal of this dissertation. Unlike previous chapters, this chapter is not focused on the methodological aspects of analysing urban access or building a predictive land value model. Instead, it is grounded in the planning domain, recalling the discussion presented in the introduction to this dissertation. It addressed the limited availability or complete absence of analytical frameworks that could aid practitioners interested in estimating the potential effects of transport investments on the value of land as a function of accessibility improvements.

Answers to the first question were provided by proposing a modelling framework to estimate the effects of a transport investment on land values. The framework applicability was demonstrated by estimating such effects for a possible future intervention of a Light Rail Transit (LRT) system in Guatemala City, the so-called MetroRiel. Data about the proposed intervention was collected from official reports. The framework structured data and methods to comprehensively analyse urban access, as addressed in previous chapters, to model residential land values and extended its applicability to incorporate possible changes or additions in the mobility infrastructures (e.g. new transport corridors, location of access stations, new road connections such as new bridges) and produce inputs to run an existing predictive model

(as presented in Chapter four) to construct an “intervened” land value map. The impacts of the proposed transport infrastructure were computed and expressed as proportional changes of accessibility metrics and of the land values.

The implementation of the framework allowed us to obtain answers to the second question. The findings revealed where and how much the expected accessibility improvements and land value uplifts are, after modelling the implementation of MetroRiel. Overall, significant land value potential increases were identified. Aligned with what is commonly reported in the literature, the effects were higher in areas near to the corridor and gradually lower at locations farther away. Within a distance range of 0.5 km, the proportional uplifts range correspondingly from up to 134% and decrease down to 35% increase. The city-wide modelling scope revealed that such effects spill beyond a distance range of 3.4 km and even beyond 4 km away from the access stations. The spatial distribution of such effects is rather heterogeneous along the corridor and across the city. By close inspection of the results, it can be deduced that the way in which such effects are distributed is associated with the availability, or lack, of other public transport infrastructures. Higher access benefits and value uplifts were identified in areas where either the use of MetroRiel leads to a significant reduction in travel time, or when its interaction with other transport infrastructure (in a perpendicular layout) represents increase in the coverage of the overall transport network.

Furthermore, new road connections in the form of bridges or minor transversal connections along the new transport corridor were shown to have an influence on the city-wide geometric access, as analysed by Space Syntax. This is also reflected in an increased geometric via geographic access that greatly benefits certain peripheral areas of the city. Consequently, this also led to predictions of significant land value uplifts. Strikingly, when translating these value uplifts into total land stock values before and after the interventions, it turns out that this is slightly less than a quarter of the required investment to build MetroRiel. The comparison suggests the feasibility of objectively formulating financial mechanisms that could consider these uplifts capable of generating economic resources (e.g. via taxation adjustments, value capture or attracting private investment) for the planning and construction of MetroRiel, a good example of sustainable transport.

However, on major limitation of the results presented in Chapter 5 was the unavailability of data in Guatemala City to cross-validate the outputs of the modelling approach. Furthermore, there are other

important limitations to the modelling framework that must be borne in mind as part of an integral perspective of how transport investments impact cities. The timing of the stages (i.e. after announcement, during and post-construction) at which capitalization might occur, as suggested in the literature, are not addressed. Furthermore, other potentially wider effects that could be triggered due to access improvement and land value uplift such as land use reorganization (e.g. shifting to commercial land uses and/or densification), inward investment and gentrification processes are also not addressed. However, such effects could eventually also contribute to a redrawing of the land and property market.

6.2.5 Overarching research goal

Meeting the fourth objective allowed us to achieve the overarching goal of this research. The resulting modelling framework incorporates a comprehensive definition of urban access (Chapter two) into a land value model (Chapters three and four) that is used for predictive purposes to construct a baseline land value map (Chapter four) and to compute the spatially distributed effects of future transport infrastructure (Chapter five).

The analytical approach has been shown to be an efficient and plausible approach to estimating the effects of transport investments. The modelling framework presented relies on estimating elasticities and semi-elasticities of land values (against analysed predictors) adopting a city-wide spatial scope. Broadly, the applicability of the framework is relevant for the domains of land administration, land use and transport planning. We elaborate about this claim in section 6.3.3.

In particular, the utilization of the land value predictive model could support mass-valuation practices. Analysis of the land value uplifts due to new transport infrastructure, discussed in Chapter 5, would contribute to addressing the need for more comprehensive CBA evaluations. Moreover, the land value model on its own could directly support automated land valuation for tax-related purposes, exploiting the model's predictive capabilities. The implementation of the modelling framework could allow planners and tax authorities to anticipate value uplifts, both citywide and at a neighbourhood scale, during the planning of new transport infrastructure and any further expansions. The outputs of such analyses can be then used to address the motivation of implementing more comprehensive CBA assessments as outlined by Banister et al. (2011) and (Vickerman, 2017) in sections 1.2 and 1.3.

6.3 Reflections

6.3.1 Contributions to scientific research

This dissertation has contributed to the currently very scarce literature dedicated to anticipating value uplifts as a result of future transport interventions. We have presented and implemented a modelling framework that allows us to estimate the spatially distributed potential effects of transport investments on land values at a city-wide scale, namely the value uplifts. The development of the framework addressed limitations identified in the existing literature and described in section 1.3. The framework comprehensively bridges state-of-the-art concepts and methods to quantitatively analyse urban access and conduct land value predictive modelling. The adopted city-wide spatial scope translates into a modelling approach that first “learns” from the relationships between access and distribution of land values (i.e. elasticities and semi-elasticities). This approach then allows exploration of the effects of hypothetical changes on the availability of mobility infrastructure.

Our methodology differs from existing studies reviewed in section 1.2 (e.g. Cengiz et al., 2019; Gallo, 2018; Pettit et al., 2020). The land value model proposed in chapter three, further developed in Chapter four and utilized in Chapter five relies on explicit accessibility as it accrues to users and not as a function of aerial proximity to transport infrastructure. Under this consideration, the actual benefit of a new transport infrastructure and proximity to the access points is strictly measured as travel time reductions (if any) towards a set of desired destinations (e.g. shopping facilities, education) selected based on their actual statistical contribution to explain land value variability (Chapter four). This approach overcomes the simplistic assumption, already outlined by Ryan (1999), that aerial proximity to new access points to transport infrastructure is translatable to improved access to the CBD. Chapter three already presented evidence that proximity to transport stations or main road infrastructure does not contribute statistical information in the presence of travel time-based metrics by both private vehicle and public transport. Moreover, similarly to the findings of Higgins et al. (2018) and Pettit et al. (2020), Chapter five argues that land value uplifts are heterogeneous at areas around the stations and along new transport corridors. In this regard, the travel-time analysis and incorporation of users preferred modality (see equation 3.1) along with the integrated modelling of the existing public transport modalities allows a more realistic analysis of the actual benefits of new transport infrastructure at both city and neighbourhood scales. Furthermore, the resulting modelling approach makes it possible to analyse land value uplifts during the planning of new

transport technologies relative to the city in question, such as the case of MetroRiel in Guatemala City.

The findings presented in this dissertation expand the existing academic knowledge about the applications of Space Syntax, as one of the foremost methodologies for analysing geometric access or network centrality. In particular, it was demonstrated for the case of Guatemala City that Space Syntax metrics, which are based on the topological and geometrical properties of road networks, are significantly and consistently correlated to common geographic access metrics that are based on travel time. New interpretations were revealed in Chapter two about what Space Syntax metrics can reveal about a city's urban access, which are also relevant to contexts in which data availability is restricted. Also, new to the literature is that the findings reported in Chapters three and four demonstrated that Space Syntax metrics do add spatial information that contributes to explaining the variability of land values at a city-wide scale next to other multiple geographic access metrics - and even when addressing spatial dependence explicitly. These results expand the growing literature strand of Space Syntax in the context of the urban economy (e.g. Di Pinto et al., 2019; Enström et al., 2008; Law et al., 2017; Matthews et al., 2007; Netzell, 2012; Xiao et al., 2016a). Moreover, the hypothesis that geometric access not only capitalizes land at its location but also as a type of resource that can be reached by common transport modes was demonstrated to hold. Under spatialized modelling conditions, it was shown to contribute with more statistical information to a residential land value model than the traditional access metric to a Central Business District.

New to the literature dedicated to land value modelling are the two following contributions. First, Chapter two provides evidence about how incorporating the concept of urban accessibility more thoroughly could lead to models that better explain the variability of land values. In particular, emphasis is placed on the way in which the actual access opportunities of public transport are addressed as they accrue to users. Second, Chapter four demonstrated the relevance of developing parsimonious models by means of implementing variable selection processes while addressing the spatial dependence problem. The results provide evidence that an explicit model component addressing spatial dependence was shown to be better at explaining variability of land values at a localized scale, leading to the removal of some geographic and geometric access variables. Not considering spatial dependence in the variable selection process might lead to a less optimum selection of predictor variables, more complex models and perhaps misleading conclusions. This finding is relevant in the context

of the high likelihood of coming across the dependence problem in any real spatial dataset as discussed by Kuntz et al. (2014). Beyond the land value modelling domain, the findings of Chapter four add knowledge to the current almost non-existent literature, as outlined by (Hoeting et al., 2006), that investigates the effects of the spatial dependence problem in variable selection procedures.

6.3.2 Contributions to the case study area

While achieving the overarching goal of this research, contributions that benefit the case study area were made. In Chapter two, structuring data from VGI sources such as Open Street Maps allowed us to consolidate the required data for the implementation of detailed accessibility analyses that was not previously readily available. The results of the analyses shed light for the first time on the current distribution of urban access and its disparity per mobility mode (i.e. private vehicle or public transport) for two cities in Guatemala. The readily available mobility networks are being implemented in ongoing research to map the ecological benefits, supply and demand for recreational green spaces in Guatemala City*. The findings of Chapters two, three and four have been translated into Spanish and shared on a few occasions with local professionals by means of public presentations and through a web platform available to a professional audience[†].

Chapters three and four presented valuable new knowledge about the conditions under which the accessibility advantages of location and other contextual characteristics determine residential land value. Furthermore, in combination with the results presented in Chapter two, various interpretations were produced of relevant structural characteristics of the city such as its predominantly monocentric structure, the observation of secondary sub-centres characterized by various types of economies and the role of city-structuring infrastructure to shape the distribution of urban access and land values. The empirical-based insights are a contribution to the existing, but limited, literature that has theorized land use models of Latin American cities (e.g. Ford, 1996; Griffin et al., 1980).

Detailed spatial information was presented in Chapter five on how Guatemala City's land values could potentially be uplifted as the

* This is a continuation of the following research: F. C., Cabrera, & D., Haase. Guatemala City: A socio-ecological profile. *Cities*, 72, 379-390, 2019.

[†] Content available at: <http://criticarq.gt/articulos/urbanismo/analizando-accesibilidad-urbana-en-contextos-de-datos-escasos-guatemala>

consequence of implementing a Light Rapid Transport system. Such information was missing in the available preliminary government reports. However, it is relevant for the ongoing discussions of what could be the implications of building such a system in terms of improved urban access. Simultaneously, the value uplifts analyses provide initial data-driven inputs for the ongoing discussions about financial mechanisms aiming for long-term sustainable transport investment. Furthermore, planning authorities are encouraged to use the insights of this research in managing desired and non-desired effects of transport investments as described in sections 1.1 and 1.2. (see Alade et al., 2020; Borras Jr et al., 2012; Jones et al., 2012; Lin, 2002); for example by analysing future viability of affordable residential development, foreseeing and avoiding potential processes of gentrification and land grabbing, and assessing how the new investment could effectively persuade private vehicle users to adopt modal shift. Moreover, authorities should not ignore how investments such as MetroRiel could also positively impact informal economies and informal settlements.

Finally, an updated version of the land value model has been developed in collaboration with private parties and the provider of residential land valuations data (see section 3.2.2). The modelling framework has been incorporated into an experimental data-driven approach for the formulation and evaluation of land-use master plans in two areas in Guatemala City. Accessibility and value uplift analyses of a many initial master plans schemes have helped stakeholders in the evaluation and design of possible forms of urban development. Outputs of the analysis have helped stakeholders in the preliminary design of master plans that seek high connectivity with neighbouring areas and aim for a diverse spectrum of medium-dense compatible land uses (including affordable housing). Moreover, presenting the results of these analyses in design meetings has aided communication and collaborations between local planning authorities, landowners, investors and designers in the strategic and sustainable development of privately-owned land.

6.3.3 Contributions to planning practice and policy

The applications of the modelling framework are diverse. Its implementation could provide a bridge for collaborations between policy makers, planners and land administration practitioners. First, its operationalization could deliver useful inputs during the “intelligence” and “design” phases of strategic transport planning, as described in section 1.5. Comprehensive accessibility analyses (presented in Chapter two and three), that are required to train the land value model, are the starting point to envisioning how new transport infrastructure

could improve accessibility. Chapters two, three and four show that visualizing city-wide distributions of geographic access, geometric access and distribution of land values provides data-driven insights about relevant structural features of a city (e.g. identifying city-structuring roads, monocentric or polycentric patterns).

Second, aligned with the contributions of other scholars (e.g. Bell et al., 2009; Cellmer, 2014; Dye et al., 2010; Gallo, 2018; Jahanshiri et al., 2011; Kuntz et al., 2014; Pettit et al., 2020), the information derived from a land value geostatistical model has relevant implications for both planners and land administrators. Statistical inferences reported in Chapters three and four provide insights about the drivers (e.g. access to a shopping mall, a hospital or level of urban connectivity) that influence the determination of land value. In this case, for Guatemala City. Understanding of these drivers provides a richer local perspective about aspects such as: (a) compatibility and value appreciation of various non-residential land uses into residential ones; (b) potential identification of negative externalities (e.g. traffic attracted to central areas during peak hours); and (c) valorisation of highly interconnected urban layouts within a neighbourhood and on a city scale, as identified via Space Syntax. Such value drivers can be then considered during land-use and mobility planning practices. Furthermore, the worth of constructing land value models does not lay only in their inferential opportunities as has been the case for many spatial econometric practices. A predictive implementation of a geostatistical model for the construction of a land value map was demonstrated in Chapters four and five. Following further refinement (e.g. using parcel-level data instead of a hexagonal tessellation), the geostatistical model could support the timely production of transparent land value data at the practical expense of limited observations of real estate valuations. This ability is fundamental for the implementation of land-based mass valuation practices (McCluskey et al., 2013), and essential for the implementation of land value capture mechanisms (Bell et al., 2009; Dye et al., 2010).

Third, the modelling framework echoes the emerging motivations to “evaluate” future transport plans by also addressing their effects on land values, as outlined by Vickerman (2017) and Banister et al. (2011). The analytical workflow, as operationalized in Chapter five, could supply inputs for both the traditional CBA (i.e. travel time reduction) as well as for cases where land value uplifts analyses are required. Furthermore, current CBA limitations in dealing with spatially distributed effects could be partially overcome by adding detailed spatial information as it is produced by the framework presented in this dissertation. Furthermore, utilization of the framework could underpin

collaborations between planners and land administration authorities. Understanding of land value uplifts is relevant for both type of practitioners on their own. However, such shared interest informed by a data-driven approach could provide objective grounds for advancing the formulation and adoption of policies such as land value capture. In this regard, the modelling framework addresses the technical difficulty inherent in estimating value uplifts and the scarcity of tools available to planners for that. Such limitations have formed a major barrier to the adoption of financial mechanisms to secure long-term funding for sustainable transport investments (Smolka, 2012).

Some considerations should be taken into account for the adoption, reproducibility and applicability of the modelling framework in the context of the applications described above. The framework can be seen as the overarching logic for quantitative analyses that would lead to produce insights not only on the current state of land (or even property) values, but further on the potential uplifts as a function of access improvements (or losses). Varying components of the model should be adapted depending on data availability and the particularities of the city under study. Hence, practitioners are recommended to either simplify or improve on the selection of variables as well as of the quantitative and correlational methods. For example, for a given city or application the selection of land uses as destinations of interest might need to be expanded or simplified. If data availability allows, some more advanced (or simpler) access metrics could be put in place. Other statistical methods might be needed to tackle more restrictive availability of valuation data. Moreover, other statistical methods could be used to improve the performance of the model and overcome the likelihood of objections to its utilization in taxation-related activities. Finally, refining the level of aggregation, from uniform areas of analysis to real property boundaries, and incorporating other land uses would lead to a more realistic perspective by incorporating the heterogeneity of localized urban contexts.

6.3.4 Prospects for future research

The findings and limitations derived from this research suggest prospects for future research. It is important to advance the validation and reproducibility of the framework. To achieve that, it is recommended to replicate the approach in a city where data on land values is available before, during and after a major transport intervention. This would provide the means to compare the results of the modelling framework against actual value uplifts of land in terms of spatial distribution, proportional magnitude and in association with already realised accessibility improvements.

From a procedural perspective, the operationalization of the entire modelling framework would still be a task to be conducted by a specialist with high technical skills. The modelling approach could be further automated and developed through an end-user approach. Automation could be achieved by implementation of an application architecture aiming for dynamic and efficient numerical computation, while facilitating certain user interaction features and an effective visualization. From experience, we suggest exploring the possibility of centralizing all the analyses post-processing into a PostgreSQL database with PostGIS and pgRouting extensions. Database functions could be set in place to automate the validations of user inputs (e.g. new roads, transport links or access points) and to trigger required accessibility analyses. The resulting “tool” should aim to facilitate a dynamic computation of scenario-based results and ease its implementation into practice by less specialised users. That would leverage an efficient utilization of data in planning practices whilst opening opportunities to engage in multi-stakeholder participation (Hall et al., 2000; Pettit et al., 2020).

Value uplifts of land as a function of accessibility are a result of a rather complex, and not fully measurable process that is also interlinked with other dynamics over time and with feedback loops (Alade et al., 2020; Higgins et al., 2018; Kii et al., 2019; Pettit et al., 2020; Yen et al., 2018), for example: redistribution of land uses, inward investment, modal shift, alleviation of traffic and capitalization in time stages (i.e. before, during and after transport interventions are made). Therefore, the modelling framework could be seen as a comprehensive reference that should be further embedded in more complex modelling architectures, such as those dedicated to model Land Use and Transport Interactions (LUTI).

Naturally, the understanding of how new transport infrastructure influences cities should not remain focussed on purely economic aspects such as land value uplifts. As argued by other authors (i.e. Jones et al., 2010; Jones et al., 2012; Lin, 2002), transport investments have also socially transformative effects. We consider it relevant that the application of the modelling framework and its outputs to be extended by investigating the implications of value uplifts for the suitability and viability of sustainable land use transformations. Of particular interest to the academic research and practice are the social and environmental impacts that value uplifts could have on land. For example, value uplifts in the form of rents only benefit landowners and not tenants who could potentially be displaced by the stimulation of gentrification processes, also increasing the pressure for land consumption in peripheral areas. It is therefore paramount to link the

findings and the opportunities that the contributions of this dissertation offers towards an integral and balanced perspective of sustainable transport planning.

Summary of main findings and reflections

Bibliography

- Adair, A., McGreal, S., Smyth, A., Cooper, J., & Ryley, T. (2000). House prices and accessibility: The testing of relationships within the Belfast urban area. *Housing studies*, 15(5), 699-716.
- Adams, D. (1994). *Urban Planning And The Development Process*. London: UCL Press Ltd.
- Ahlfeldt, G. M. (2007). *If Alonso was right: accessibility as determinant for attractiveness of urban location*: Hamburg contemporary economic discussions.
- Ahlfeldt, G. M. (2013). If we build it, will they pay? Predicting property price effects of transport innovations. *Environment Planning A*, 45(8), 1977-1994.
- Ahlfeldt, G. M., & Wendland, N. (2011). Fifty years of urban accessibility: The impact of the urban railway network on the land gradient in Berlin 1890–1936. *Regional Science and Urban Economics*, 41(2), 77-88.
- Ahlström, A., Pilesjö, P., & Lindberg, J. (2011). Improved accessibility modeling and its relation to poverty – A case study in Southern Sri Lanka. *Habitat International*, 35(2), 316-326.
- Alade, T., Edelenbos, J., & Gianoli, A. (2020). A sustainable approach to innovation adoption in light-rail transport. *Sustainability*, 12(3), 1262.
- Albacete, X., Olaru, D., Paül, V., & Biermann, S. (2015). Measuring the Accessibility of Public Transport: A Critical Comparison Between Methods in Helsinki. *Applied Spatial Analysis and Policy*, 1-28.
- Alonso, W. (1964). *Location and land use: towards a general theory of land rent*. Cambridge, MA: Harvard University Press.
- Anantsuksomsri, S., & Tontisirin, N. (2015). The Impacts of Mass Transit Improvements on Residential Land Development Values: Evidence from the Bangkok Metropolitan Region. *Urban Policy and Research*, 33(2), 195-216.
- Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1), 3-25.
- Arsanjani, J. J., Zipf, A., Mooney, P., & Helbich, M. (2015). An Introduction to OpenStreetMap in Geographic Information Science: Experiences, Research, and Applications. In *OpenStreetMap in GIScience* (pp. 1-15): Springer.
- Bafna, S. (2003). Space Syntax: A Brief Introduction to Its Logic and Analytical Techniques. *Environment and Behavior*, 35(1), 17-29.
- Banister, D. (2007). Cities, mobility, and climate change. *Journal of Industrial Ecology*, 11(2), 7-10.
- Banister, D., & Thurstain-Goodwin, M. (2011). Quantification of the non-transport benefits resulting from rail investment. *Journal of Transport geography*, 19(2), 212-223.

- Bank, W. (2017). GDP per capita growth in Guatemala. Retrieved from <http://data.worldbank.org/indicator/NY.GDP.PCAP.KD.ZG?end=2015&locations=GT&start=2001&view=chart>
- Basu, S., & Thibodeau, T. (1998). Analysis of Spatial Autocorrelation in House Prices. *The Journal of Real Estate Finance and Economics*, 17(1), 61-85.
- Batty, M. (2004). A new theory of space syntax. In. London: Centre of Advanced Spatial Analysis UCL.
- Batty, M. (2009). Accessibility: in search of a unified theory. *Environment and Planning B: Planning and Design*, 36(2), 191-194.
- Batty, M. (2013). The new science of cities. In: Mit Press.
- Bell, M., Bowman, J., & German, J. (2009). The assessment requirements for a separate tax on land. *Land Value Taxation: Theory, Evidence, and Practice*. Lincoln Institute of Land Policy, Cambridge, Mass, 129-194.
- Birch, C. P., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological modelling*, 206(3-4), 347-359.
- Birch, E. L., & Wachter, S. M. (2011). World Urbanization: The Critical Issue of the Twenty-First Century. In *Global Urbanization* (pp. 3-23). Philadelphia: University of Pennsylvania Press.
- Black, J. (2018). Fundamentals of land use - transport planning. In *Urban transport planning: Theory and practice* (Vol. 4). London, UK: Routledge.
- Borras Jr, S. M., Franco, J. C., Gómez, S., Kay, C., & Spoor, M. (2012). Land grabbing in Latin America and the Caribbean. *The Journal of Peasant Studies*, 39(3-4), 845-872.
- Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). Spatial dependence, housing submarkets, and house price prediction. *The Journal of Real Estate Finance and Economics*, 35(2), 143-160.
- Bourassa, S. C., Cantoni, E., & Hoesli, M. (2010). Predicting house prices with spatial dependence: A comparison of alternative methods. *Journal of Real Estate Research*, 32(2), 139-159.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345-370.
- Burdziej, J. (2012). A Web-based spatial decision support system for accessibility analysis—concepts and methods. *Applied Geomatics*, 4(2), 75-84.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, 33(2), 261-304.
- Case, B., Clapp, J., Dubin, R., & Rodriguez, M. (2004). Modeling spatial and temporal house price patterns: A comparison of four

- models. *The Journal of Real Estate Finance and Economics*, 29(2), 167-191.
- Cellmer, R. (2014). The possibilities and limitations of geostatistical methods in real estate market analyses. *Real Estate Management and Valuation*, 22(3), 54-62.
- Cengiz, E. C., & Çelik, H. M. (2019). Investigation of the impact of railways on housing values; the case of Istanbul, Turkey. *International Journal of Transport Development*, 3(4), 295-305.
- Cervero, R., & Duncan, M. (2002). Transit's value-added effects: Light and commuter rail services and commercial land values. *Transportation Research Record*, 1805(1), 8-15.
- Chiaradia, A., Hillier, B., Barnes, Y., & Schwander, C. (2009, 2009). *Residential Property Value Patterns*. Paper presented at the 7th International Space Syntax Symposium, Stockholm.
- Chica Olmo, J. (1995). Spatial estimation of housing prices and locational rents. *Urban Studies*, 32(8), 1331-1344.
- Chica-Olmo, J., Cano-Guervos, R., & Chica-Olmo, M. (2013). A coregionalized model to predict housing prices. *Urban Geography*, 34(3), 395-412.
- Colonna, P., Berloco, N., & Circella, G. (2012). The interaction between land use and transport planning: a methodological issue. *Procedia-Social and Behavioral Sciences*, 53, 84-95.
- Condat, L., Van De Ville, D., & Blu, T. (2005). *Hexagonal versus orthogonal lattices: A new comparison using approximation theory*. Paper presented at the Image Processing, 2005. ICIP 2005. IEEE International Conference.
- Cooper, C. (2014). Using OpenStreetMap in spatial network analysis. In. Cardiff University: sDNA.
- Cooper, C., & Chiaradia, A. (2015). sDNA: how and why we reinvented Spatial Network Analysis for health, economics and active modes of transport.
- Cressie, N. (1992). Statistics for spatial data. *Terra Nova*, 4(5), 613-617.
- Cressie, N. (2015). *Statistics for spatial data*: John Wiley & Sons.
- Curl, A., Nelson, J. D., & Anable, J. (2011). Does Accessibility Planning address what matters? A review of current practice and practitioner perspectives. *Research in Transportation Business & Management*, 2, 3-11.
- Curtis, C., & Scheurer, J. (2010). Planning for sustainable accessibility: Developing tools to aid discussion and decision-making. *Progress in Planning*, 74(2), 53-106.
- Curtis, E. A., Comiskey, C., & Dempsey, O. (2016). Importance and use of correlational research. *Nurse Researcher*, 23(6).

- D. Knowles, R., & Ferbrache, F. (2016). Evaluation of wider economic impacts of light rail investment on cities. *Journal of Transport Geography*, 54, 430-439.
- Dale-Johnson, D., & Jan Brzeski, W. (2001). Land Value Functions and Land Price Indexes in Cracow, 1993–1999. *Journal of Housing Economics*, 10(3), 307-334.
- Dalton, R. C. (2003). The Secret Is To Follow Your Nose. *Environment and Behavior*, 35(1), 107-131.
- Debrezion, G., Pels, E., & Rietveld, P. (2007). The impact of railway stations on residential and commercial property value: a meta-analysis. *The Journal of Real Estate Finance Economics*, 35(2), 161-180.
- Decker, B. L. (1986). *World geodetic system 1984*. Retrieved from St. Louis, Missouri:
- Des Rosiers, F., & Recherche, U. L. F. d. s. d. l. a. D. d. l. (2001). *Neighborhood profiles and house values: dealing with spatial autocorrelation using kriging techniques* (289524121X). Retrieved from Québec:
- Des Rosiers, F., Thériault, M., & Villeneuve, P.-Y. (2000). Sorting out access and neighbourhood factors in hedonic price modelling. *Journal of Property Investment & Finance*, 18(3), 291-315.
- Desyllas, J. (1997). *Berlin in Transition: using space syntax to analyse the relationship between land use, land value and urban morphology*. Paper presented at the Proceedings.
- Devaux, N., Dubé, J., & Apparicio, P. (2017). Anticipation and post-construction impact of a metro extension on residential values: The case of Laval (Canada), 1995–2013. *Journal of Transport geography*, 62, 8-19.
- Di Pinto, V., & Rinaldi, A. M. (2019). A configurational approach based on geographic information systems to support decision-making process in real estate domain. *Soft Computing*, 23(9), 2853-2862.
- Diao, M., Leonard, D., & Sing, T. F. (2017). Spatial-difference-in-differences models for impact of new mass rapid transit line on private housing values. *Regional Science Urban Economics*, 67, 64-77.
- DICABI. (2005). Manual de Valuación Inmobiliaria. In (pp. 119). Guatemala.
- Drakakis-Smit, D. W. (2000). *Third world cities* (Second edition ed.). London etc.: Routledge.
- Du, H., & Mulley, C. (2006). Relationship between transport accessibility and land value: Local model approach with geographically weighted regression. *Transportation Research Record*, 1977(1), 197-205.

- Du, H., & Mulley, C. (2012). Understanding spatial variations in the impact of accessibility on land value using geographically weighted regression. *Journal of Transport and Land Use*, 5(2), 46-59.
- Dubin, R., Pace, K., & Thibodeau, T. (1999). Spatial autoregression techniques for real estate data. *Journal of Real Estate Literature*, 7(1), 79-95.
- Dubin, R. A. (1992). Spatial autocorrelation and neighborhood quality. *Regional Science and Urban Economics*, 22(3), 433-452.
- Dye, R. F., & England, R. W. (2010). *Assessing the theory and practice of land value taxation*. Cambridge, MA: Lincoln Institute of Land Policy
- Enström, R., & Netzell, O. (2008). Can Space Syntax Help Us in Understanding the Intraurban Office Rent Pattern? Accessibility and Rents in Downtown Stockholm. *The Journal of Real Estate Finance and Economics*, 36(3), 289-305.
- Evans, A. W. (1987). *Urban economics : an introduction*. Oxford: Blackwell.
- Ferbrache, F., & Knowles, R. D. (2017). City boosterism and place-making with light rail transit: A critical review of light rail impacts on city image and quality. *Geoforum*, 80, 103-113.
- Filatova, T., Parker, D., & van der Veen, A. (2009). Agent-based urban land markets: agent's pricing behavior, land prices and urban land use change. *Journal of Artificial Societies and Social Simulation*, 12(1), 3.
- Fletcher, T. (2015). Quantitative Psychology Tools, version 1.5. In.
- Ford, L. (1996). A New and Improved Model of Latin American City Structure. *Geographical Review*, 86(3), 437-440.
- Fox, J., & Weisberg, S. (2016). Companion to Applied Regression. Version 2.1-2. In.
- Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35-41.
- Fujita, M., & Krugman, P. (2004). The new economic geography: Past, present and the future. *Papers in Regional Science*, 83(1), 139-164.
- Gallo, M. (2018). The impact of urban transit systems on property values: A model and some evidences from the city of Naples. *Journal of Advanced Transportation*, 2018, 1-22.
- Gellert, G. (1990). Desarrollo de la estructura espacial en la ciudad de Guatemala: Desde su fundación hasta la revolución de 1944. *Anuario de Estudios Centroamericanos*, 16(1), 31-55.
- Getis, A. (2007). Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, 37(4), 491-496.
- Geurs, K., & van Eck, R. (2001). Accessibility measures, review and applications. In *RIVM rapport = RIVM report* (Vol. 4085050, pp.

- 265). Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu (RIVM).
- Geurs, K., & Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport geography*, 12(2), 127-140.
- Giannopoulou, M., Vavatsikos, A. P., & Lykostratis, K. (2016). A Process for Defining Relations between Urban Integration and Residential Market Prices. *Procedia-Social and Behavioral Sciences*, 223(2016), 153-159.
- Gil, J. (2015). Building a Multimodal Urban Network Model Using OpenStreetMap Data for the Analysis of Sustainable Accessibility. In Z. A. Jokar Arsanjani J., Mooney P., Helbich M. (Ed.), *OpenStreetMap in GIScience* (pp. 229-251). Cham: Springer.
- Gil, J., Varoudis, T., Karimi, K., & Penn, A. (2015). *The space syntax toolkit: Integrating depthmapX and exploratory spatial analysis workflows in QGIS*. Paper presented at the SSS 2015-10th International Space Syntax Symposium.
- Giuliano, G., Gordon, P., Pan, Q., & Park, J. (2010). Accessibility and residential land values: Some tests with new measures. *Urban Studies*, 47(14), 3103-3130.
- Gleave, S. D. (2005). *What Light Rail can do for Cities: A review of the evidence*. Retrieved from Leeds, UK:
- Glebbeeck, M.-L., & Koonings, K. (2015). Between Morro and Asfalto. Violence, insecurity and socio-spatial segregation in Latin American cities. *Habitat International*, 54(1), 3-9.
- Goodchild, B. (1994). Housing design, urban form and sustainable development: reflections on the future residential landscape. *Town Planning Review*, 65(2), 143.
- Griffin, E., & Ford, L. (1980). A Model of Latin American City Structure. *Geographical Review*, 70(4), 397-422.
- Grimes, A., & Liang, Y. (2010). Bridge to somewhere: Valuing Auckland's northern motorway extensions. *Journal of Transport Economics and Policy (JTEP)*, 44(3), 287-315.
- Hajrasouliha, A., & Yin, L. (2015). The impact of street network connectivity on pedestrian volume. *Urban Studies*, 52(13), 2483-2497.
- Hall, P., & Pfeiffer, U. (2000). *Urban future 21 : a global agenda for 21st century cities*. London: E & FN SPON.
- Handy, S. L., & Niemeier, D. A. (1997). Measuring accessibility : an exploration of issues and alternatives. *Environment and Planning A*, 29(7), 1175-1194.
- Hansen, W. G. (1959). How accessibility shapes land use. In: *Journal of the American Institute of Planners*, (1959) pp. 73-76.

- Heikkilä, E., Gordon, P., Kim, J. I., Peiser, R. B., Richardson, H. W., & Dale-Johnson, D. (1989). What happened to the CBD-distance gradient?: land values in a polycentric city. *Environment and planning A*, 21(2), 221-232.
- Held, L., & Sabanés Bové, D. (2014). Model Selection. In *Applied Statistical Inference: Likelihood and Bayes* (pp. 221-245). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Hengl, T., Heuvelink, G. B. M., & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120(1), 75-93.
- Hewko, J., Smoyer-Tomic, K. E., & Hodgson, M. J. (2002). Measuring neighbourhood spatial accessibility to urban amenities: does aggregation error matter? *Environment and planning A*, 34(7), 1185-1206.
- Higgins, C., & Kanaroglou, P. (2018). Rapid transit, transit-oriented development, and the contextual sensitivity of land value uplift in Toronto. *Urban Studies*, 55(10), 2197-2225.
- Hillier, B. (2007a). Cities as movement economies. In *Space is the machine: a configurational theory of architecture* (pp. 111-137): Space Syntax.
- Hillier, B. (2007b). *Space is the machine: a configurational theory of architecture* (Digital edition ed.). London: Space Syntax.
- Hillier, B. (2009). *Spatial sustainability in cities: Organic patterns and sustainable forms*. Paper presented at the 7th International Space Syntax Symposium, Stockholm.
- Hillier, B., Greene, M., & Desyllas, J. (2000). Self-generated Neighbourhoods: the role of urban form in the consolidation of informal settlements. *Urban Design International*, 5(2), 61-96.
- Hillier, B., & Iida, S. (2005). *Network and psychological effects in urban movement*. Paper presented at the International Conference on Spatial Information Theory.
- Hillier, B., Leaman, A., Stansall, P., & Bedford, M. (1976). Space syntax. *Environment and Planning B*, 3(2), 147-185.
- Hillier, B., & Penn, A. (2004). Rejoinder to Carlo Ratti. *Environment and Planning B-Planning & Design*, 31(4), 501-511.
- Hillier, B., Penn, A., Hanson, J., Grajewski, T., & Xu, J. (1993). Natural Movement-or, configuration and attraction in urban pedestrian movement. *Environ Plann B*, 20(1), 29-66.
- Hillier, B., Turner, A., Yang, T., & Park, H.-T. (2010). Metric and topogeometric properties of urban street networks: some convergences, divergences and new results. *Journal of Space Syntax Studies*, 1(2), 1-21.
- Hillier, B., Yang, T., & Turner, A. (2012). Normalising least angle choice in Depthmap-and how it opens up new perspectives on the

- global and local analysis of city space. *Journal of Space Syntax*, 3(2), 155-193.
- Hoeting, J. A., Davis, R. A., Merton, A. A., & Thompson, S. E. (2006). Model selection for geostatistical models. *Ecological Applications*, 16(1), 87-98.
- Iacono, M., & Levinson, D. (2011). Location, regional accessibility, and price effects: Evidence from home sales in Hennepin County, Minnesota. *Transportation Research Record: Journal of the Transportation Research Board*, 2245(1), 87-94.
- Iacono, M., & Levinson, D. (2015). Accessibility dynamics and location premia: Do land values follow accessibility changes? *Urban Studies*, 54(2), 364-381.
- Ibeas, Á., Cordera, R., dell'Olio, L., Coppola, P., & Dominguez, A. (2012). Modelling transport and real-estate values interactions in urban systems. *Journal of Transport geography*, 24, 370-382.
- IDOM. (2016). *Interconexión vial y transporte ferroviario interurbano de pasajeros eje Norte-Sur de la Ciudad de Guatemala*. Guatemala
- INE, I. N. d. E. (2016). Índice de Precios al Consumidor. Retrieved from <https://www.ine.gob.gt/index.php/estadisticas-continuas/indice-de-precio-al-consumidos>
- Ingram, D. R. (1971). Concept of accessibility : a search for an operational form. *Regional Studies*, 5(2), 101-107.
- Ingram, G. K., & Carroll, A. (1981). The spatial structure of Latin American cities. *Journal of Urban Economics*, 9(2), 257-273.
- Ingvarðson, J. B., & Nielsen, O. A. (2018). Effects of new bus and rail rapid transit systems—an international review. *Transport Reviews*, 38(1), 96-116.
- Jahanshiri, E., Buyong, T., & Shariff, A. R. M. (2011). A review of property mass valuation models. *Pertanika Journal of Science and Technology*, 19(1), 23-30.
- Jenkins, P., Smith, H., Wang, Y.P. (2007). *Planning and housing in the rapidly urbanising world*. London: Routledge.
- Jenks, M., & Jones, C. (2010). *Dimensions of the sustainable city* (Vol. 2). Dordrecht: Springer.
- Jiang, B., & Claramunt, C. (2002). Integration of space syntax into GIS: new perspectives for urban morphology. *Transactions in GIS*, 6(3), 295-309.
- Jiang, B., Claramunt, C., & Batty, M. (1999). Geometric accessibility and geographic information: extending desktop GIS to space syntax. *Computers, Environment and Urban Systems*, 23(2), 127-146.
- Jiang, B., & Liu, C. (2009). Street-based topological representations and analyses for predicting traffic flow in GIS. *International*

- Journal of Geographical Information Science*, 23(9), 1119-1137.
- Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, 30(2), 3-13.
- Jones, C., Leishman, C., MacDonald, C., Orr, A., & Watkins, D. (2010). Economic Viability. In M. Jenks & C. Jones (Eds.), *Dimensions of the Sustainable City* (Vol. 2, pp. 145-162). Dordrecht: Springer.
- Jones, P., & Lucas, K. (2012). The social consequences of transport decision-making: clarifying concepts, synthesising knowledge and assessing implications. *Journal of Transport geography*, 21, 4-16.
- Kaplan, N., Burg, D., & Omer, I. (2020). The spatial organization of accessibility and functional hierarchy: The case of Israel. *Computers, Environment and Urban Systems*, 80, 1-14.
- Karimi, K. (2012). A configurational approach to analytical urban design: 'Space syntax' methodology. *Urban Design International*, 17(4), 297-318.
- Kennedy, C., Miller, E., Shalaby, A., Maclean, H., & Coleman, J. (2005). The four pillars of sustainable urban transportation. *Transport Reviews*, 25(4), 393-414.
- Kii, M., Moeckel, R., & Thill, J.-C. (2019). Land use, transport, and environment interactions: WCTR 2016 contributions and future research directions. *Computers, Environment and Urban Systems*, 77, 1-4.
- Kim, H.-K., & Sohn, D. W. (2002). An analysis of the relationship between land use density of office buildings and urban street configuration: Case studies of two areas in Seoul by space syntax analysis. *Cities*, 19(6), 409-418.
- Kitanidis, P. K. (1993). Generalized covariance functions in estimation. *Mathematical Geology*, 25(5), 525-540.
- Kivell, P. (1993). *Land and the city : patterns and processes of urban change*. London etc.: Routledge.
- Koenig, J.-G. (1980). Indicators of urban accessibility: theory and application. *Transportation*, 9(2), 145-172.
- Kourtit, K., Nijkamp, P., & Partridge, M. D. (2015). Challenges of the New Urban World. *Applied Spatial Analysis and Policy*, 8(3), 199-215.
- Krause, A., & Bitter, C. (2012). Spatial econometrics, land values and sustainability: Trends in real estate valuation research. *Cities*, 29, 19-25.
- Kuntz, M., & Helbich, M. (2014). Geostatistical mapping of real estate prices: an empirical comparison of kriging and cokriging. *International Journal of Geographical Information Science*, 28(9), 1904-1921.

- Law, S. (2017). Defining Street-based Local Area and measuring its effect on house price using a hedonic price approach: The case study of Metropolitan London. *Cities*, 60, 166-179.
- Law, S., Penn, A., Karimi, K., & Shen, Y. (2017). *The economic value of spatial network accessibility for uk cities: a comparative analysis using the hedonic price approach*. Paper presented at the Proceedings-11th International Space Syntax Symposium, SSS 2017.
- LeSage, J., & Pace, R. K. (2009). Motivating and interpreting Spatial Econometric Models. In *Introduction to spatial econometrics* (pp. 25-43). Boca Raton: CRC Press.
- Li, Q., Zhou, S., & Wen, P. (2019). The relationship between centrality and land use patterns: Empirical evidence from five Chinese metropolises. *Computers, Environment and Urban Systems*, 78, 1-13.
- Li, X., & Love, P. E. (2020). Employing land value capture in urban rail transit public private partnerships: Retrospective analysis of Delhi's airport metro express. *Research in Transportation Business Management*, 32, 100431.
- Lin, J. (2002). Gentrification and transit in northwest Chicago. *Transportation Quarterly*, 56(4), 175-191.
- Lin, T.-C., & Evans, A. W. (2000). The relationship between the price of land and size of plot when plots are small. *Land Economics*, 76(3), 386-394.
- Litman, T. (2007a). Developing indicators for comprehensive and sustainable transport planning. *Transportation Research Record*, 2017(1), 10-15.
- Litman, T. (2007b). Land use impacts on transport: How land use factors affect travel behavior. Canada: Victoria Transport Policy Institute.
- Litman, T., & Burwell, D. (2006). Issues in sustainable transportation. *International Journal of Global Environmental Issues*, 6(4), 331-347.
- Liu, Zheng, B., Turkstra, J., & Huang, L. (2010). A hedonic model comparison for residential land value analysis. *International Journal of Applied Earth Observation and Geoinformation*, 12, 181-193.
- Liu, C., & Shen, Q. (2011). An empirical analysis of the influence of urban form on household travel and energy consumption. *Computers, Environment and Urban Systems*, 35(5), 347-357.
- Lungo, M., & Smolka, M. (2005). Land value and large urban projects: The Latin American experience. *Land Lines*, 17(1), 3-6.
- Luo, J., & Wei, Y. D. (2004). A geostatistical modeling of urban land values in Milwaukee, Wisconsin. *Geographic Information Sciences*, 10(1), 49-57.

- Malczewski, J. (1999). *GIS and multicriteria decision analysis*. New York etc.: Wiley & Sons.
- Martínez, L. M., & Viegas, J. M. (2009). Effects of transportation accessibility on residential property values: Hedonic Price Model in the Lisbon, Portugal, metropolitan area. *Transportation Research Record*, 2115(1), 127-137.
- Matthews, J., & Turnbull, G. (2007). Neighborhood Street Layout and Property Value: The Interaction of Accessibility and Land Use Mix. *The Journal of Real Estate Finance and Economics*, 35(2), 111-141.
- Mavoa, S., Witten, K., McCreanor, T., & O'Sullivan, D. (2012). GIS based destination accessibility via public transit and walking in Auckland, New Zealand. *Journal of Transport geography*, 20(1), 15-22.
- McCluskey, W., McCord, M., Davis, P., Haran, M., & McIlhatton, D. (2013). Prediction accuracy in mass appraisal: a comparison of modern approaches. *Journal of Property Research*, 30(4), 239-265.
- McCluskey, W. J., Deddis, W. G., Lamont, I. G., & Borst, R. A. (2000). The application of surface generated interpolation models for the prediction of residential property values. *Journal of Property Investment & Finance*, 18(2), 162-176.
- McMillen, D. P. (2004). Employment Densities, Spatial Autocorrelation, and Subcenters in Large Metropolitan Areas. *Journal of Regional Science*, 44(2), 225-244.
- Meng, Q., Liu, Z., & Borders, B. E. (2013). Assessment of regression kriging for spatial interpolation – comparisons of seven GIS interpolation methods. *Cartography and Geographic Information Science*, 40(1), 28-39.
- Metz, D. (2017). Valuing transport investments based on travel time saving: Inconsistency with United Kingdom policy objectives. *Case studies on transport policy*, 5(4), 716-721.
- Meyer, M. D. (2000). Transport planning for urban areas: A retrospective look and future prospects. *Journal of Advanced Transportation*, 34(1), 143-171.
- Milder, J. (2012). Sustainable Urban Form. In E. van Bueren, H. van Bohemen, L. Itard, & H. Visscher (Eds.), *Sustainable Urban Environments* (pp. 263-284): Springer Netherlands.
- Mohammad, S. I., Graham, D. J., Melo, P. C., & Anderson, R. J. (2013). A meta-analysis of the impact of rail projects on land and property values. *Transportation Research Part A: Policy and Practice*, 50, 158-170.
- Morales, J., Flacke, J., Morales, J., & Zevenbergen, J. (2019a). Mapping Urban Accessibility in Data Scarce Contexts Using Space Syntax

- and Location-Based Methods. *Applied Spatial Analysis and Policy*, 12(2), 205-228.
- Morales, J., Flacke, J., & Zevenbergen, J. (2019b). Modelling residential land values using geographic and geometric accessibility in Guatemala City. *Environment and Planning B: Urban Analytics and City Science*, 46(4), 751-776.
- Morales, J., Stein, A., Flacke, J., & Zevenbergen, J. (2020). Predictive land value modelling in Guatemala City using a geostatistical approach and Space Syntax. *International Journal of Applied Earth Observation and Geoinformation*, 1-24.
- Morales, J. A. (2013). *Sustainable development by means of a collaborative planning framework, the case of municipal housing in Guatemala city*. University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede.
- Municipality, G. C. (2009). Guía de Aplicación. Plan de ordenamiento Territorial. In (pp. 58). Guatemala.
- Munroe, D. K. (2007). Exploring the determinants of spatial pattern in residential land markets: amenities and disamenities in Charlotte, NC, USA. *Environment and Planning B: Planning and Design*, 34(2), 336-354.
- Muth, R. F. (1969). *Cities and Housing, The Spatial Pattern of Urban Residential Land Use*. Chicago: The University of Chicago Press.
- Netzell, O. (2012). The effect of accessibility on retail rents: testing integration value as a measure of geographic location. *Journal of Property Research*, 30(1), 1-23.
- Nyerges, T. L., & Jankowski, P. (2009). *Regional and urban GIS: a decision support approach*. New York: Guilford Press.
- Odeh, I. O., McBratney, A., & Chittleborough, D. (1994). Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma*, 63(3-4), 197-214.
- Oliveira, V., & Pinho, P. (2010). Evaluation in Urban Planning: Advances and Prospects. *Journal of planning literature*, 24(4), 343-361.
- Omer, I., Rofè, Y., & Lerman, Y. (2015). The impact of planning on pedestrian movement: contrasting pedestrian movement models in pre-modern and modern neighborhoods in Israel. *International Journal of Geographical Information Science*, 29(12), 2121-2142.
- Opsomer, J. D., Ruppert, D., Wand, M. P., Holst, U., & Hössjer, O. (1999). Kriging with nonparametric variance function estimation. *Biometrics*, 55(3), 704-710.
- Orford, S. (2002). Valuing locational externalities: a GIS and multilevel modelling approach. *Environment and Planning B*, 29(1), 105-128.

- Osland, L., & Thorsen, I. (2013). Spatial impacts, local labour market characteristics and housing prices. *Urban Studies*, 50(10), 2063-2083.
- Ottensmann, J. R., Payton, S., & Man, J. (2008). Urban location and housing prices within a hedonic model. *Journal of Regional Analysis and Policy*, 38(1), 19-35.
- Paci, L., Beamonte, M. A., Gelfand, A. E., Gargallo, P., & Salvador, M. (2017). Analysis of residential property sales using space-time point patterns. *Spatial Statistics*, 21, 149-165.
- Pacione, M. (2005). *Urban geography : a global perspective* (Second edition ed.). London etc.: Routledge.
- Pan, Q. (2013). The impacts of an urban light rail system on residential property values: a case study of the Houston METRORail transit line. *Transportation Planning and Technology*, 36(2), 145-169.
- Pebesma, E., & Graeler, B. (2016). Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation. Version 1.1-3. In.
- Peiser, R. B. (1987). The determinants of nonresidential urban land values. *Journal of Urban Economics*, 22(3), 340-360.
- Penn, A. (2003). Space Syntax and Spatial Cognition or Why the axial line? *Environment and Behavior*, 35(1), 30-65.
- Peterson, G. E. (2009). *Unlocking land values to finance urban infrastructure*. Washington, D.C.: The World Bank.
- Pettit, C., Shi, Y., Han, H., Rittenbruch, M., Foth, M., Lieske, S., . . . Christensen, B. (2020). A new toolkit for land value analysis and scenario planning. *Environment and Planning B: Urban Analytics and City Science*, 1-18.
- Porta, S., Crucitti, P., & Latora, V. (2005). The network analysis of urban streets: a primal approach. *Environment and Planning B*, 33(5), 705-725.
- Porta, S., Crucitti, P., & Latora, V. (2006). The network analysis of urban streets: a dual approach. *Physica A: Statistical Mechanics and its Applications*, 369(2), 853-866.
- Porta, S., Latora, V., & Strano, E. (2010). Networks in urban design. Six years of research in multiple centrality assessment. In *Network science* (pp. 107-129): Springer.
- Porta, S., Latora, V., Wang, F., Rueda, S., Strano, E., Scellato, S., . . . Latora, L. (2012). Street Centrality and the Location of Economic Activities in Barcelona. *Urban Studies*, 49(7), 1471-1488.
- Pujol, R., Pérez, E., & Sánchez, L. (2013). Estimaciones Del Impacto Del Límite De Crecimiento (Anillo De Contención) Sobre Los Valores Del Suelo En El Norte De Heredia, 1997-2007. *Revista de Ciencias Económicas*, 31(2), 117-134.

- Ratti, C. (2004). Urban texture and space syntax: some inconsistencies. *Environment and Planning B: Planning and Design*, 31(4), 487-499.
- Ribeiro Jr, P. J., & Diggle, P. J. (2015). geoR: Analysis of Geostatistical Data. R package version 1.7-5.1. In.
- Rietveld, P. (2015). Flying Carpets and Urban Development Trajectories. *Applied Spatial Analysis and Policy*, 8(3), 217-230.
- Rietveld, P., & van Wee, B. (2008). Ex ante evaluation of railway station development projects: issues still to be solved. In *Railway Development* (pp. 147-169). The Netherlands: Physica-Verlag HD.
- Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2016). Support Functions and Datasets for Venables and Ripley's MASS. version 7.3-45. In.
- Rodríguez, D. A., & Mojica, C. H. (2009). Capitalization of BRT network expansions effects into prices of non-expansion areas. *Transportation Research Part A: Policy and Practice*, 43(5), 560-571.
- Ross, C. L., Orenstein, M., & Botchwey, N. (2014). HIA, EIA, SIA and Other Appraisals. In *Health Impact Assessment in the United States* (pp. 33-43). Ney York: Springer.
- Ryan, S. (1999). Property values and transportation facilities: finding the transportation-land use connection. *Journal of planning literature*, 13(4), 412-427.
- Saeid, A. (2011). Space Syntax as a tool to assess land value. In R. Sietchiping (Ed.), *Innovative Land and Property Taxation* (pp. 172-190). Nairobi: UN-HABITAT.
- Seitzinger, S. P., Svedin, U., Crumley, C. L., Steffen, W., Abdullah, S. A., Alfsen, C., . . . Dearing, J. A. (2012). Planetary stewardship in an urbanizing world: beyond city limits. *Ambio*, 41(8), 787-794.
- Serra, M., & Hillier, B. (2019). Angular and Metric Distance in Road Network Analysis: A nationwide correlation study. *Computers, Environment and Urban Systems*, 74, 194-207.
- Seya, H., Tsutsumi, M., Yoshida, Y., & Kawaguchi, Y. (2011). Empirical Comparison of the Various Spatial Prediction Models: in Spatial Econometrics, Spatial Statistics, and Semiparametric Statistics. *Procedia-Social and Behavioral Sciences*, 21, 120-129.
- Sharifi, M. A., & Zucca, A. (2009). Integrated planning and decision support systems : concepts and application to a site selection problem. In *Spatial decision support for urban and environmental planning : a collection of case studies* (pp. 5-31). Kuala Lumpur: Arah Publications.

- Sharma, R., & Newman, P. (2018). Does urban rail increase land value in emerging cities? Value uplift from Bangalore Metro. *Transportation Research Part A: Policy Practice*, 117, 70-86.
- Smolka, M. O. (2012). A new look at value capture in Latin America. *Land Lines*, 24(3), 10-15.
- Spinney, J., Kanaroglou, P., & Scott, D. (2011). Exploring Spatial Dynamics with Land Price Indexes. *Urban Studies*, 48(4), 719-735.
- Steadman, P. (2004). Developments in space syntax. *Environment and Planning B: Planning and Design*, 31(4), 483-486.
- Steinebach, G. (2009). Planning Sustainable Living . In H. Hagen, S. Guhathakurta, & G. Steinebach (Eds.), *Visualizing Sustainable Planning* (pp. 3-36). Berlin: Springer Berlin Heidelberg.
- Stewart, J. Q., & Warntz, W. (1958). Physics of population distribution. *Journal of Regional Science*, 1(1), 99-121.
- Team, R. C. (2016). R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing.
- Tsutsumi, M., Shimada, A., & Murakami, D. (2011). Land price maps of Tokyo Metropolitan Area. *Procedia - Social and Behavioral Sciences*, 21, 193-202.
- Turner, A. (2007). From axial to road-centre lines: a new representation for space syntax and a new model of route choice for transport network analysis. *Environment and Planning B: Planning and Design*, 34(3), 539-555.
- UN-HABITAT. (2005) Promoting Local Economic Development through Strategic Planning. In, *The Local Economic Development Series: Vol. 1* (pp. 24). Nairobi: UN-HABITAT
- EcoPlan International Inc.
- URBANISTICA. (2009). Estratificación Socioeconómica por Sector Censal. [Socio Economic Groups per Sensus Track].
- Van der Molen, P. (2002). The dynamic aspect of land administration: an often-forgotten component in system design. *Computers, Environment and Urban Systems*, 26(5), 361-381.
- van Nes, A. (2019). Applied Mathematics on Urban Space. In L. D'Acci (Ed.), *The Mathematics of Urban Morphology* (pp. 253-270). Cham: Springer International Publishing.
- van Nes, A., Berghauser-Pont, M., & Mashhoodi, B. (2011). *Combination of Space Syntax with Spacematrix and the Mixed Use Index*. Paper presented at the 8th International Space Syntax Symposium, Santiago, Chile.
- Vickerman, R. (2007). Cost—Benefit analysis and large-scale infrastructure projects: State of the art and challenges.

- Environment and Planning B: Planning and Design*, 34(4), 598-610.
- Vickerman, R. (2017). Beyond cost-benefit analysis: the search for a comprehensive evaluation of transport investment. *Environment Research in Transportation Economics*, 63, 5-12.
- Viguie, V., & Hallegatte, S. (2014). *Urban infrastructure investment and rent-capture potentials*: The World Bank.
- Volchenkov, D. (2019). Assessing Complexity of Urban Spatial Networks. In L. D'Acci (Ed.), *The Mathematics of Urban Morphology* (pp. 173-197): Springer.
- Wackernagel, H. (2013). *Multivariate geostatistics: an introduction with applications*: Springer Science & Business Media.
- Waddell, P., Berry, B. J., & Hoch, I. (1993). Residential property values in a multinodal urban area: New evidence on the implicit price of location. *The Journal of Real Estate Finance and Economics*, 7(2), 117-141.
- Walacik, M., Cellmer, R., & Żróbek, S. (2013). Mass appraisal-international background, Polish solutions and proposal of new methods application. *Geodetski list*, 67(4), 255-269.
- Walker, G., Fay, H., & Mitchell, G. (2005). *Environmental Justice Impact Assessment: An evaluation of requirements and tools for distributional analysis*. Retrieved from Stoke-on-Trent:
- Waze. (2015). Community-based Mapping, Traffic and Navigation. Retrieved from www.waze.com
- Webster, C. (2010). Pricing accessibility: Urban morphology, design and missing markets. *Progress in Planning*, 73(2), 77-111.
- Wegener, M., & Fürst, F. (2004). Land-use transport interaction: state of the art. Available at SSRN 1434678.
- World_Pop. (2014). Guatemala Population Map. Retrieved from <http://www.worldpop.org.uk/data/summary/?contselect=America&countselect=Guatemala&typeselect=Population>. Retrieved March 2016 <http://www.worldpop.org.uk/data/summary/?contselect=America&countselect=Guatemala&typeselect=Population>
- Xiao, Y., Orford, S., & Webster, C. J. (2016a). Urban configuration, accessibility, and property prices: a case study of Cardiff, Wales. *Environment and Planning B: Planning and Design*, 43(1), 108-129.
- Xiao, Y., Webster, C., & Orford, S. (2016b). Identifying house price effects of changes in urban street configuration: An empirical study in Nanjing, China. *Urban Studies*, 53(1), 112-131.
- Yan, S., Delmelle, E., & Duncan, M. (2012). The Impact of a New Light Rail System on Single Family Property Values in Charlotte, North Carolina. *The Journal of Transport and Land Use*, 5(2), 60-67.

- Yeh, & Gar-On, A. (1991). The development and applications of geographic information systems for urban and regional planning in the developing countries. *International Journal of Geographical Information System*, 5(1), 5-27.
- Yeh, A. (1999). Urban Planning and GIS. *Geographical Information Systems*, 2, 877-888.
- Yen, B. T., Mulley, C., Shearer, H., & Burke, M. (2018). Announcement, construction or delivery: When does value uplift occur for residential properties? Evidence from the Gold Coast Light Rail system in Australia. *Land Use Policy*, 73, 412-422.
- Yoo, E.-H., & Kyriakidis, P. C. (2009). Area-to-point Kriging in spatial hedonic pricing models. *Journal of Geographical Systems*, 11(4), 381.
- Zhang, R., Du, Q., Geng, J., Liu, B., & Huang, Y. (2015). An improved spatial error model for the mass appraisal of commercial real estate based on spatial analysis: Shenzhen as a case study. *Habitat International*, 46, 196-205.
- Zhu, Q., & Lin, H. S. (2010). Comparing Ordinary Kriging and Regression Kriging for Soil Properties in Contrasting Landscapes. *Pedosphere*, 20(5), 594-606.

Summary

Sustainable mobility, that provided by Light Rail Transit (LRT) and Bus Rapid Transit (BRT) systems, is of growing interest in many Global South countries as a sustainable strategy to face the challenges of rapid urbanization. Sustainable mobility is fundamental to boosting economies and promoting equality. It is well known that accessibility improvements also have wider effects such as land values uplifts. Thus, practitioners are calling for options to include these effects in Cost Benefit Analyses (CBAs). Direct effects (i.e. travel-time savings) cannot be easily translated into wider often economic benefits. This makes it difficult to justify the high investment required. In contrast, a direct translation of the monetized accessibility improvements could provide useful information for design processes and for plan evaluation/selection processes. Moreover, land value uplift analyses could provide the grounds for the formulation of mechanisms (e.g. land value capture) to secure funding for sustainable transport investments. However, practitioners continue to face methodological difficulties in estimating land value uplifts reliably.

Studies that aim to predict value uplifts before the interventions are implemented are scarce. Some limitations are identified in such literature as there is. First, estimating the elasticities of land values as a function of accessibility would be the preferred approach rather than an urban economic formulation in areas where market imperfections are accentuated or even unknown. Second, a comprehensive concept and operationalization of urban accessibility as it accrues to users and the valorisation of land is missing. Accounting for public mobility benefits by Euclidian distance to transit stops makes it impossible to associate the value uplift with an interpretable metric of improved geographic access (i.e. travel times reduction). Introduction of new transport technologies commonly include modifications to existing road networks. Hence, impacts on the geometric accessibility (i.e. network centrality) can also be expected. This is important since recent studies have suggested that geometric accessibility, as analysed in Space Syntax (SSx), adds spatial information that improves the understanding of land and property values. Third, the spatial scope of analysis must not be restricted to pre-assumed catchment areas. Instead, a city scale approach would allow the calibration of elasticities based on richer datasets while gaining broader spatial insights into the potential land value effects of a transport investment.

The objective of this research was to propose and implement a modelling framework to estimate the spatially distributed land value uplifts of future transport infrastructure by means of operationalizing a comprehensive accessibility definition into a predictive model. To do so, we implemented a quantitative correlational research using

Guatemala City as the case study area. We first compared location-based and Space Syntax methods to map urban accessibility in data scarce contexts. Multiple data layers were obtained from official and Volunteered Geographic Information (VGI) sources and then consolidated. Second, we incorporated a comprehensive definition of accessibility and uncovered its relations with residential land values by means of a multivariate regression model. Third, we used a geostatistical method to develop a parsimonious land value model. The predictive model was then used to construct a land value map for Guatemala City. Finally, we structured data and methods from previous steps into a modelling framework. The framework's applicability is demonstrated by estimating the future impacts of an LRT system on urban access and residential land values.

The findings revealed that SSx metrics consistently reflect urban accessibility conditions that have had previously only been addressed using location-based methods. Furthermore, an increased ability to explain residential land-values and prediction accuracy was achieved when including geometric-access metrics and addressing the disparity of geographic-access opportunities due to available transport modes. From the results, it appears that geometric- accessibility capitalizes land not only at its location, but also as a resource that is reachable by means of geographic-access.

Implementing a geostatistical method was demonstrated to be important in developing a parsimonious model. Some SSx metrics were confirmed to contribute significant statistical information, even under spatialized modelling conditions. Moreover, it was observed that geographic access to urban centrality contributed more statistical information compared to access to the central business district. A land value map, in combination with inferential modelling and accessibility analyses, allowed us to conclude that Guatemala City is predominantly monocentric. The road network hierarchy identified via SSx was observed to play an important role in the distribution of urban access and land values.

Operationalizing the modelling framework revealed where and how much the accessibility improvements and land value uplifts can be expected after introducing an LRT system. The effects were observed to be higher in areas near to the corridor but gradually lower at locations farther away. However, the city-wide modelling scope revealed that such effects spill beyond a distance of 3.4 km and even beyond 4 km from the access stations. The spatial distribution of such effects is rather heterogeneous along the corridor and across the city. The way in which such effects are distributed is associated with the

availability, or its lack, of other public transport infrastructures. Higher access benefits and value uplifts were identified in areas where either MetroRiel's use leads to a real reduction in travel time, or when its interaction with other transport infrastructure (in a perpendicular layout) represents increase in the coverage of the overall transport network. Furthermore, new road connections in the form of bridges or minor transversal connections along the new transport corridor have a positive influence on the city-wide access. This improvement, as measured by means of geometric via geographic access, has been shown to greatly benefit certain peripheral areas of the city. Consequently, this also leads to significant land value uplifts. Strikingly, less than a quarter of the net land stock value uplifts would meet the required investment to build the LRT system.

Our findings and discussions expand the literature available about Space Syntax from two perspectives: first, about the applicability of such a method in the Global South context and derivable knowledge to map urban access; second, its contribution to improving land value modelling while spatial dependence is not ignored. The framework demonstrates how a comprehensive operationalization of urban accessibility allows the estimation of impacts of future transport infrastructure, even when such infrastructure is new to the city being considered. The modelling workflow offers a reference for practitioners in overcoming the difficulties of estimating land value uplifts. Overcoming such difficulty contributes to the ongoing regional discussions about the feasibility of adopting financial mechanisms to secure funding for sustainable transport investments. It is this aspect that means the framework could provide a bridge for collaborations between planners and land administration practitioners.

The limitations reported in this dissertation suggest directions for future research. It is vital to advance on the validation and reproducibility of the framework. The modelling approach could be further automated and developed focussed, and be more centred, on an end-users approach. Capitalization of land as a function of accessibility is a result of a rather complex, and not fully measurable, process that is also interlinked with other dynamics over time and with feedback loops. The modelling framework could be further embedded in more complex modelling architectures, such as those dedicated to modelling the Land Use and Transport Interactions (LUTI). Naturally, the understanding of how new transport infrastructure influences cities should not remain narrowly focused on economic benefits. It is therefore paramount to link the findings and the opportunities that the contributions of this research offer towards achieving an integral and balanced perspective within sustainable transport planning.

Samenvatting

Duurzame mobiliteit, middels Light Rail Transit (LRT) en Bus Rapid Transit (BRT), is van toenemend belang voor veel landen op het zuidelijk halfrond, als een duurzame strategie om de uitdagingen van de snelle urbanisatie het hoofd te bieden. Duurzame mobiliteit is van fundamenteel belang voor economische groei en de bevordering van gelijkheid. Het is algemeen bekend dat verbetering van de toegankelijkheid verstrekkende effecten heeft zoals toenemende waardes van grond. Daarom pleiten professionals voor mogelijkheden om deze effecten in kosten-baten analyses (KBA's) op te nemen. De directe effecten (o.a. besparingen in reistijd) zijn niet eenvoudig om te vertalen naar bredere economische voordelen. Dit maakt het lastig om de benodigde hoge investeringen te rechtvaardigen. Een directe vertaling van de te gelde gemaakte verbeteringen in toegankelijkheid kan daarentegen veel bruikbare informatie bieden voor ontwerpprocessen en voor de processen die ten grondslag liggen aan de selectie van projecten.

Analyses van waardeontwikkeling van kavels kunnen bovendien een basis vormen voor de formulering van mechanismen (o.a. grondwaarde projectie) waarmee fondsen verworven kunnen worden ten behoeve van duurzame mobiliteitsinvesteringen. Praktijkbeoefenaars lopen echter nog continue aan tegen methodologische uitdagingen op het gebied van het betrouwbaar inschatten van de toename van landwaardes.

Er zijn maar weinig studies die beogen om landwaarde stijgingen te prognosticeren voordat de interventies geïmplementeerd zijn. Sommige beperkingen zijn aangestipt in de beschikbare literatuur. Allereerst, zou het schatten van de elasticiteiten in waarde van grond als een functie van de toegankelijkheid de voorkeursaanpak zijn, in plaats van het formuleren van een stedelijke economische functie in gebieden waar imperfecties in de markt aangemerkt of zelfs onbekend zijn. Ten tweede is er een gebrek aan een uitgebreid concept en op rationalisatie van stedelijke toegankelijkheid en de wijze hoe deze doorwerkt op gebruikers en de waardebepaling van grond. Verantwoording voor de voordelen van openbaar vervoer op basis van Euclidische afstanden naar de doorvoerhaltes maakt het onmogelijk om de waardestijging te koppelen d.m.v. een interpreteerbare meeteenheid voor toegenomen geografische toegankelijkheid (o.a. vermindering van reistijden). Introductie van nieuwe transporttechnologieën vereisen vaak aanpassingen van de bestaande weginfrastructuur. Daarom liggen impacts op de geometrische toegankelijkheid (o.a. netwerk centraliteit) in de lijn der verwachting. Dit is van belang omdat recente onderzoeken hebben voorgesteld dat geometrische toegankelijkheid, zoals onderzocht in Space Syntax

(SSx), spatiale informatie toevoegt die het begrip van grond- en vastgoedwaardes verhoogt. Ten derde, de spatiale scope van het onderzoek niet moet worden beperkt tot vooronderstelde en vooraf gedefinieerde verzorgingsgebieden. In plaats daarvan, zou een aanpak op stadsschaal kalibratie van elasticiteiten toelaten, gebaseerd op rijkere datasets en tegelijkertijd ervoor zorgen dat er bredere ruimtelijke inzichten worden opgedaan op het gebied van potentiële grondwaarde effecten van een transportinvestering.

Het doel van dit onderzoek was om een modeleringsraamwerk voor te stellen en te implementeren, waarmee de ruimtelijk gedistribueerde grondwaardestijgingen als gevolg van toekomstige investeringen in transportinfrastructuur, door de operationalisering van een uitgebreide toegankelijkheidsdefinitie in een voorspellend model. Om dit te bereiken hebben we een kwantitatief co-relatieve onderzoek opgezet waarbij Guatemala-stad als onderzoeksgebied is gebruikt. We hebben eerst locatie gebonden en Space Syntax methoden gehanteerd om de stedelijke toegankelijkheid in data-arme situaties in kaart te brengen. Meerdere datalagen werden verkregen van officiële en vrijwillig verzamelde geografische informatiebronnen (VGI) en zijn daarna samengevoegd. Daarna hebben we een uitgebreide definitie van toegankelijkheid opgenomen en hebben we de relaties van deze definitie met grondwaardes in woongebieden bepaald met behulp van een multivariabel regressiemodel. Ten derde hebben we een geo-statistische methode gehanteerd om een spaarzaam grondwaardemodel te ontwikkelen. Het voorspellende model hebben we daarna gebruikt om een grondwaardekaart voor Guatemala stad te genereren. Tenslotte hebben we de data en methoden van de voorgaande stappen gestructureerd en in een modeleringsraamwerk omgezet. De toepasbaarheid van het raamwerk wordt aangetoond door het te gebruiken om de toekomstige effecten van een LRT-systeem op het gebied van stedelijke toegankelijkheid en grondwaardes in woongebieden te bepalen.

Het onderzoek heeft onder meer aangetoond dat SSx statistieken consequent de stedelijke toegankelijkheidssituatie weergeeft, die daarvoor enkel werden geadresseerd door gebruik van locatie gebonden methoden. Bovendien werd een groter vermogen om de grondwaarden van woonwijken en voorspellingsnauwkeurigheid te verklaren bereikt door meetkundige meetgegevens op te nemen en de ongelijkheid van geografische toegangsmogelijkheden als gevolg van beschikbare transportmogelijkheden aan te pakken. Uit de resultaten blijkt dat geometrische toegankelijkheid niet alleen de grond op zijn locatie kapitaliseert, maar ook als een hulpbron die via geografische toegang bereikbaar is.

Het implementeren van een geo-statistische methode bleek belangrijk te zijn bij het ontwikkelen van een zo efficiënt mogelijk model. Van sommige SSx-getallen werd bevestigd dat ze significante statistische informatie opleveren, zelfs onder ruimtelijke modelleringsomstandigheden. Bovendien werd opgemerkt dat geografische toegang tot stedelijke centraliteit meer statistische informatie bijdroeg in vergelijking met toegang tot het centrale zakendistrict. Een grondwaardekaart, in combinatie met inferentiële modellering en toegankelijkheidsanalyses, lieten ons concluderen dat Guatemala-Stad overwegend monocentrisch is. De via SSx geïdentificeerde wegnethiërarchie blijkt een belangrijke rol te spelen bij de verdeling van stedelijke ontsluiting en grondwaarden.

Door het modelleerraamwerk te operationaliseren, is gebleken waar en in hoeverre de toegankelijkheidsverbeteringen en grondwaarde verhogingen kunnen worden verwacht na de introductie van een LRT-systeem. De effecten bleken duidelijker waarneembaar te zijn in gebieden nabij de corridor, maar geleidelijk minder sterk op locaties verder weg. De stadsbrede modellering bracht echter aan het licht dat dergelijke effecten zich over een afstand van 3,4 km en zelfs meer dan 4 km van de MetroRiel-stations uitstrekken. De ruimtelijke spreiding van dergelijke effecten is nogal heterogeen langs de corridor en over de gehele stad. De manier waarop dergelijke effecten worden verdeeld, hangt samen met de beschikbaarheid of het ontbreken van andere openbaar vervoerinfrastructuren. Hogere toegangsvoordelen en waardestijgingen werden vastgesteld in gebieden waar het gebruik van MetroRiel leidt tot een echte verkorting van de reistijd, of waar de interactie met andere transportinfrastructuur (in een loodrechte layout) een toename van de dekking van het totale transportnetwerk vertegenwoordigt. Daarnaast hebben nieuwe wegverbindingen in de vorm van bruggen of kleinere transversale verbindingen langs de nieuwe transportcorridor een positief effect op de gehele ontsluiting van de stad. Het onderzoek toont aan dat deze verbetering bepaalde perifere delen van de stad enorm ten goede komt. Bijgevolg kan dit ook leiden tot aanzienlijke stijging van de grondwaarde. Het is extra opvallend dat minder dan een kwart van de stijging van de netto grondwaardetoename de vereiste investering voor de ontwikkeling van het LRT-systeem zou kunnen dekken.

Onze bevindingen en discussies dragen bij aan de beschikbare literatuur over Space Syntax vanuit twee perspectieven: ten eerste over de toepasbaarheid van een dergelijke methode in de context van het zuidelijk halfmond en daaruit voortvloeiende kennis om stedelijke toegang in kaart te brengen; ten tweede aan het verbeteren van grondwaardemodellering, terwijl de ruimtelijke afhankelijkheid niet

wordt genegeerd. Het raamwerk laat zien hoe een alomvattende operationalisering van de stedelijke bereikbaarheid het mogelijk maakt de effecten van toekomstige transportinfrastructuur te schatten, zelfs wanneer een dergelijke infrastructuur nieuw is voor de betreffende stad. De modelleringssamenhang biedt een referentie voor praktijkbeoefenaars, en helpt hen bij het overwinnen van de moeilijkheden die ontstaan bij het schatten van de stijging van de grondwaarde. Het aanpakken van dergelijke problemen draagt bij aan de lopende regionale discussies over de haalbaarheid van het invoeren van financiële mechanismen om financiering voor duurzame transportinvesteringen veilig te stellen. Dit aspect kan betekenen dat het raamwerk een brug zou kunnen slaan in samenwerkingen tussen stedelijke planners en kadastrale organisaties (bijvoorbeeld middels een grondbelasting).

De beperkingen die in dit proefschrift worden vermeld, geven tegelijkertijd de mogelijke richtingen aan voor toekomstig onderzoek. Het is van essentieel belang om vooruitgang te boeken bij de validatie en reproduceerbaarheid van het raamwerk. De modelleringsaanpak zou verder geautomatiseerd en ingericht kunnen worden, en bovendien meer op de eindgebruiker gericht ontwikkeld kunnen worden. Grondkapitalisatie als functie van toegankelijkheid is het resultaat van een vrij complex en niet volledig meetbaar proces dat in de loop van de tijd ook verweven is met andere systeemdynamieken en feedback loops. Het modelleerraamwerk zou verder kunnen worden ingebed in meer complexe modelarchitecturen, zoals degene die gewijd zijn aan het modelleren van de Landgebruik- en Transportinteracties (LUTI). Uiteraard mag het begrip van de invloed van nieuwe transportinfrastructuur op steden niet strikt gericht blijven op enkel economische voordelen. Het is daarom van het grootste belang om de bevindingen en de mogelijkheden die dit onderzoek bieden met elkaar te verbinden om te komen tot een integraal en evenwichtig perspectief binnen duurzame openbaarvervoersplanning.

Samenvatting

About the author



Andrés Morales was born in Guatemala City, Guatemala, on April 28th, 1986. He completed his Bachelor of Architecture Design in 2009 at Istmo University in the same city. Immediately after he joined the planning bureau of Guatemala City Hall. He was highly motivated by wanting to get more knowledge on how to understand cities and generate creative inputs that lead to integrative architectural and urban solutions. Andrés participated in the implementation of the city's new territorial ordinance plan (POT); the formulation of localized plans aiming for the redevelopment and the rehabilitation of Guatemala City's central areas and in the implementation of collaborative approaches to conceptualize and design public spaces.

In 2011 he travelled to The Netherlands to advance his studies at the Faculty of Geo-information Sciences and Earth Observation (ITC), University of Twente. In 2013 he was awarded with an MSc in Geo-information Sciences and Earth Observation with application to urban and regional planning. During his dissertation he developed a collaborative framework in Guatemala City to identify the suitability and feasibility to building affordable residence in central areas. Same year he was granted a scholarship through the NICHE project, funded by Nuffic, and began his doctoral research.

Since 2017 he is based in The Hague and is the geodata specialist lead at MoViCi (formerly known as SIMCI), now a subdivision of Next Generation Infrastructures. There he applies technical skills and domain knowledge, next to his colleagues, in developing a simulation platform that addresses the challenges of transforming interdependent critical Dutch infrastructures into resilient systems. Andrés is engaged in actively developing the simulation platform from a conceptual and visionary perspective as well as facilitating an agile and performant team workflow. In 2019 and 2020 he collaborated with IBI Group (USA-Canada teams) to develop an innovative data-driven approach to optimize the design and land value uplift strategies of master plans for urban developments in Guatemala City.

Andrés is highly motivated to strengthening the intersection between technology and domain knowledge to disruptively leverage the transformation of cities towards sustainable processes, resilient systems and responsiveness.

UNIVERSITY
OF TWENTE.



ISBN: 90-978-365-5094-9
DOI: 10.3990/1.9789036550949
Diss.no: 387