
Predicting accessibility in Los Angeles County:
A comparative study using a multilevel model and an artificial neural network

By

CHRISTINA LAST



Department of Geographical Sciences
UNIVERSITY OF BRISTOL

Presented as part of, and in accordance with, the requirements
for the Final Degree of B.Sc. at the University of Bristol, School
of Geographical Sciences, April 2019 (Geography with Study
Abroad).

APRIL 2019

Word count: 9979





CERTIFICATION OF OWNERSHIP OF THE COPYRIGHT IN A TYPESCRIPT OR MANUSCRIPT

Dissertation presented as part of, and in accordance with, the requirements for the Final Degree of B.Sc/M.Sci (delete as appropriate) at the University of Bristol, School of Geographical Sciences.

I hereby assert that I own exclusive copyright in the item named below. I give permission to the University of Bristol Library to add this item to its stock and to make it available for use and re-copying by its readers.

AUTHOR	Christina Kimberley Last
TITLE	<p>Predicting accessibility in Los Angeles County:</p> <p>A comparative study using a multilevel model and an artificial neural network</p>
DATE OF SUBMISSION	4 th April 2019

Signed:

A handwritten signature in black ink, appearing to read "Christina Kimberley Last".

Full name: Christina Kimberley Last

Date: 04/04/19

ACKNOWLEDGMENTS

This study forms a part of my B.Sc. degree at the University of Bristol. I am grateful to Dr. Levi Wolf for his unwavering support and technical guidance, he is one of the most proficient and gracious mentors a student could hope to have. He was endlessly generous with his time, expertise and network, and has provided unparalleled advice, not only on this study but on the applications of these techniques in industry, and has become a good friend.

I extend my thanks to Professor Kostas Goulias for providing the neighbourhood accessibility data from the Southern California Association of Governments (SCAG). Prof. Goulias is a lynchpin of the transportation research community, and during my sabbatical as an exchange scholar at the University of California Santa Barbara, he provided opportunities to work on national transportation planning projects and present research at an international travel behaviour conference. Because of this, I am an indebted member of his legion of fans and friends in the academic community.

Furthermore, I welcome the introduction of transportation indicators to the Los Angeles Sustainability Plan (pLAN), thanks to the incredible work from former colleagues Heidi Creighton and Chris Rhie, and the task force of cities consultants at BuroHappold Engineering. Both Heidi and Chris are driven, tenacious and - above all - incredibly intuitive. Their intellect and work ethic has placed transportation centre stage of the pLAN, and I am proud to have worked with them during the project - I am sure it will not be the last time.

I gratefully acknowledge the support of Peter Chen, Laurence Wright and Miranda Kombert, as their technical excellence, huge determination and eye for detail significantly improved the quality of this research.

Predicting accessibility in Los Angeles County: a comparative study using a multilevel model and an artificial neural network

Christina Last

School of Geographical Sciences, University of Bristol, UK
cl15540@bristol.ac.uk

Abstract

The social exclusion of minority groups from areas with good accessibility to employment opportunities has long characterised large metropolitan centres, and thus has attracted a high level of academic attention. However, previous transportation research aggregates accessibility to large spatial units such as the state or metropolitan level, and fails to illustrate the differential spatial accessibility for certain minority groups. Comparing two advanced modelling procedures, a multilevel model (MLM) and an artificial neural network (ANN), I take advantage of granular neighbourhood accessibility data to show that Hispanic communities experience lower equity and lower accessibility in Los Angeles County. The results from this novel methodology indicate that the artificial neural network offers better predictive accuracy than multilevel models, especially when the ethnic composition of communities is introduced (ANN $R^2 = 0.42$ and MLM $R^2 = 0.15$). I conclude that advancements in the predictive capabilities of machine learning techniques offer a promising entry point to further research on mechanisms that drive accessibility in Los Angeles County.

Word count: 9979

Keywords and phrases accessibility, multilevel modelling, neural networks

Supplement Material All documentation and code to run this analysis can be found at
<https://github.com/ChristinaLast/Equity-Accessibility-Los-Angeles>

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

Contents

1	Introduction	7
2	Literature Review	10
2.1	Defining Accessibility	10
2.2	Assessing accessibility	11
2.3	Representation	12
2.3.1	Representing Space	13
2.3.2	Representing Individuals	14
2.3.3	Representing Space and Time	15
2.4	Methodology	15
2.4.1	Aggregated Modelling	15
2.4.2	Multilevel Modelling	16
2.4.3	Machine Learning	17
2.5	Applications	17
2.5.1	Accessibility as an economic indicator	18
2.5.2	Accessibility as a social indicator	19
2.6	Evaluating the Research Gap	20
2.7	Narrowing the Research Gap	21
2.8	Research Questions	21
3	Data and Methodology	23
3.1	Data overview and Setting	23
3.2	Variables	26
3.2.1	Response Variable	26
3.2.2	Equity Predictors	26
3.2.3	Ethnicity Predictors	27
3.2.4	Spatial Weights Matrix	29
3.3	Methodology Overview	30
3.3.1	Multilevel Model	30

Christina Last

3.3.2 Artificial Neural Network	31
3.4 Testing Predictions	33
3.5 Software	33
4 Results	34
4.1 Descriptive Statistics	34
4.2 Performance Measures	34
4.2.1 Goodness of Fit	34
4.2.2 Predictive Accuracy	36
4.2.3 Explanatory Capacity	36
4.3 Null Model	37
4.4 To what extent do household and transportation costs as a percentage of total income matter in determining accessibility to employment opportunities?	39
4.5 Is there evidence of spatial spillovers that explain differences in accessibility to employment opportunities?	42
4.6 What is the relative effect of being African American or Hispanic in determining the level of accessibility to employment opportunities at the block level?	45
5 Discussion	49
6 Limitations and implications for future research	51
7 Conclusion	52
8 Technical Appendix	55
8.1 Data Preparation	55
8.2 Spatial Autocorrelation	56
8.2.1 Global Binary Spatial Autocorrelation	56
8.2.2 Local Continuous Spatial Autocorrelation	58
8.3 Model Specifications	60

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

8.3.1	Specification of the Null Model	61
8.3.2	Specification of Model 1	62
8.3.3	Specification of Model 2	63
8.3.4	Specification for Model 3	65
8.3.5	Specification of Neural Network	66
8.4	Results	67
8.4.1	Multilevel Modelling	67
8.4.2	Artificial Neural Network	70

List of Figures

1	Literature Analysis Methodology	11
2	Literature Timeline	12
3	Representation scatterplot	13
4	Methodology scatterplot	16
5	Application scatterplot	18
6	Combined 3D scatterplot	20
7	Total local Accessibility	24
8	Household and Transportation costs	25
9	Figure showing the distribution of (a) African American and (b) Hispanic population in Los Angeles County.	26
10	Multilevel modelling process	27
11	Artificial Neural Network analysis process	28
12	Null Model Trace and Posterior plots	33
13	Null Model predictions	34
14	Model 1 Trace and Posterior plots	35
15	Model 1 predictions	36
16	Model 1 error	37
17	Model 2 Trace and Posterior plots	38
18	Model 2 predictions	39
19	Model 2 error	40
20	Model 3 Trace and Posterior plots	41
21	Model 3 predictions	42
22	Model 3 error	43
23	The significance, visualisation and covariance of the multilevel model predictions.	56
24	The significance, visualisation and error history of the Artificial Neural Network predictions.	57

List of Tables

Table		Page
1	Hypothesis and equation for each model	22
2	Descriptive statistics for all variables introduced into the models . . .	31
3	Table comparing the goodness of fit and accuracy measurements for the MLM and ANN using Mean Absolute error (MAE), Mean Absolute Percentage Error (MAPE) and R^2	32
4	Results from all Multilevel models	54
5	Results from all Artificial Neural Networks	57
6	Performance measures for all Artificial Neural Networks	57

1 Introduction

The uneven spatial distribution of access to employment opportunities has long characterised American cities. This has directly contributed to segregation and income inequality in urban areas (Glaeser *et al.* 2008, Chetty *et al.* 2014). Illustrating this, over 19 percent of Americans living in urban areas are below the poverty level¹, whereas in suburban locations this decreases to 7.5 percent (Glaeser *et al.* 2008: 3). Lower income households, in particular those lacking car ownership, are more likely to live in inner city areas proximate to public transportation. Brueckner and Rosenthal (2006) indicate that lower quality housing, and a lack of investment in transportation infrastructure in central city areas further explains the central location of low income households. To understand this relationship, it is essential to introduce the concept of accessibility, defined as the “ease of contact” to spaces of activity (Hansen 1959: 3).

Accessibility to employment opportunities is often the most influential factor in determining where people live. However, patterns of accessibility in the USA are changing. Before the 21st century, living proximate to public transportation was a successful predictor of locating the urban poor in American cities, yet more recently, rising household and transportation costs in American metropolitan regions has led to the “Suburbanisation of Poverty” (Kneebone and Garr 2010: 1). Between 1985 and 2000 in Los Angeles, a typical middle income household would spend 36 percent of their income on rent; whereas since 2000, rent costs increased to 47 percent of income for a typical middle income earner. Furthermore, households in the bottom third of income earners need to spend 121 percent of their income on rent to live proximate to jobs in Los Angeles County (Terrazas 2018). The exclusionary effect of rapidly increasing cost of living in Los Angeles has led to low income households locating further from the city centre.

As of 2008, the suburban poor were the largest growing community, increasing by 25 percent between 2000 and 2008. This an important dynamic to understand, as accessibility to

¹ The definition of poverty level uses monetary income taxation, not including capital gains and noncash benefits. For single-occupancy households in 2008 the poverty threshold was \$11,201, in contrast, for a family of two children and two adults the threshold was \$21,834 (Kneebone and Garr 2010: 20).

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

employment opportunities remain a “significant predictor of upward mobility” (Chetty *et al.* 2014: 1611). As such, this study attempts to understand the characteristics of communities with low access to spatially distributed activities, in order to target transportation policy that directly improves accessibility and the chance of upward mobility. Nowhere in America is this analysis more important than in Los Angeles County, whose citizens experience the most time-consuming commute in the USA (Carter 2019).

However, there is lacking consensus on the best method to identify the characteristics of neighbourhoods with low accessibility. Furthermore, there is enthusiasm in the academic community to adopt contemporary techniques such as machine learning, for example, (Chen *et al.* 2016, de Bruijn *et al.* 2018) to analyse patterns of accessibility using “big spatio-temporal datasets” (Chen and Kwan 2018: 1). In an attempt to consolidate the methodological approaches to measuring accessibility, I compare the results of a multilevel model (MLM) and an artificial neural network (ANN) to explore the relationship between equity and accessibility for Hispanics and African Americans in Los Angeles County. For the multilevel model, I adopt a cross-sectional, three level structure with census blocks as the primary sampling unit, block groups as the secondary unit and tracts the third, to capture the neighbourhood level effects on accessibility to employment opportunities. For the artificial neural network, I build a sequential, feed-forward neural network using two dense layers to capture the spatial pattern of accessibility. I introduce a novel statistical tool into the MLM and ANN to model the spatial spillover effects of low equity communities in Los Angeles County. Alongside a broader investigation into the comparison between multilevel modelling and neural networks, the following three questions are considered:

1. To what extent do household and transportation costs as a percentage of total income matter in determining accessibility to employment opportunities?
2. Is there evidence of spatial spillovers that explain differences in accessibility to employment opportunities?
3. What is the relative effect of being a Hispanic or African American in determining the level of accessibility to employment opportunities?

Christina Last

The contributions of this paper are as follows. Foremost, Chapter 1 contextualises the literature on the low upward mobility of the urban poor in American cities. In Chapter 2, I foreground the multiple definitions and measures of accessibility, using the framework of representation, methodology and application (Kwan *et al.* 2003). After evaluating recent attempts to measure accessibility within the framework, I present the need to consolidate approaches to measuring accessibility, whilst incorporating novel statistical techniques that identify spatial clusters of high and low accessibility. To do so, in Chapter 3, I outline a comparative methodology for a multilevel model and an artificial neural network. I present and discuss the results in Chapter 4. In Chapter 5, I compare the results from the MLM and ANN. I report the study's limitations and suggest implications for further research in Chapter 6. Chapter 7 concludes.

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

2 Literature Review

To unpack the literature, I trace the definitions of accessibility, starting from Hansen (1959: 3) who defines accessibility as the “ease of contact” to spatially distributed activities. After selecting a definition of accessibility, I draw on Kwan *et al.’s* (2003) framework to evaluate the literature’s “simultaneous treatment of representation, methodology and application” (2003: 129). I average each score and visualise the results with a 3D scatterplot, using the highest scoring literature to closely inform this study (figure 1). The review highlights that accessibility literature has failed to develop a methodology that represents accessibility accurately and has a clear application in transportation planning (figure 6a). After identifying this research gap, I develop a comparative analysis of multilevel modelling and artificial neural networks to understand the spatial barriers to accessibility for ethnic groups.

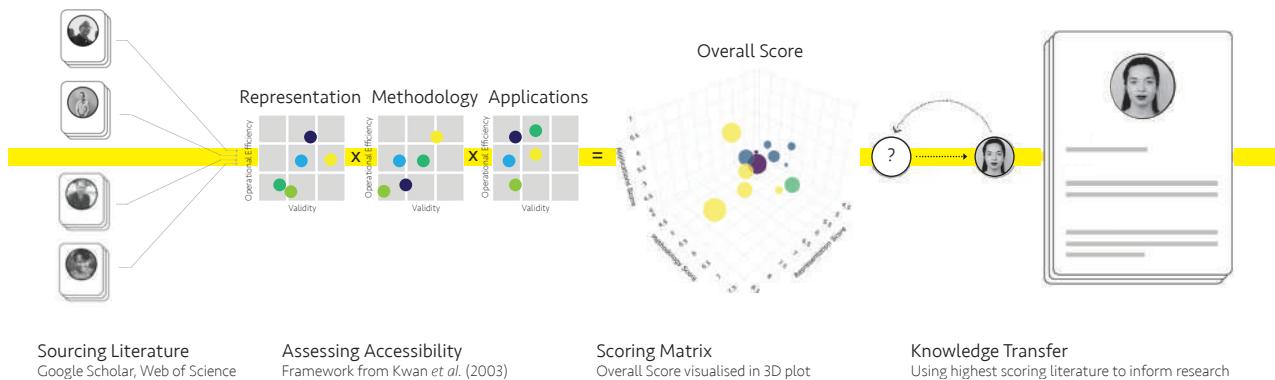


Figure 1 Process of analysing and evaluating literature on accessibility.

2.1 Defining Accessibility

The definition of accessibility previously depended on the context of its application. In early literature, accessibility was measured purely as a function of utility, defined as the likelihood of an individual choosing a particular destination based on the utility value of that choice compared to other perceivable choices (Bogue 1949). Since then, academics such as Hansen (1959) have measured accessibility geographically and argued that accessibility should be understood as a function of the potential opportunities for spatial interaction “adjusting for

the ability and desire of people or firms to overcome spatial separation” (1959: 5). Although a seminal paper, Hansen’s argument narrows the definition of accessibility to employment and retail opportunities, perhaps showing a limited understanding of spatially distributed activities. In addition, Hansen’s method removes spatial granularity by categorising land as inter-urban (between urban areas) and intra-urban (within urban areas), which fails to indicate the nuanced differences between rural, suburban and central city accessibility patterns. A more pragmatic understanding of accessibility has been introduced by Van Wee and Geurs (2011), who defined accessibility as “the extent to which land use and transport systems enable individuals to reach activities and destinations” (2011: 351). They argue that the definition of accessibility should combine location based and utility based theory, yet be computed with models using spatially granular data. I adopt this definition as a foundation to develop an approach that combines a location based measure of accessibility with a novel methodology, to analyse the neighbourhood characteristics that determine local access to employment opportunities.

2.2 Assessing accessibility

Although the components of accessibility are difficult to define, they are crucial to develop an understanding of the level of access individuals have to employment opportunities. I summarise the historical and contemporary approaches of measuring accessibility in figure 2. The timeline shows that both the theoretical approach and modelling strategies evaluating accessibility change from spatial aggregations of accessibility (Hansen 1959) to individual, real-time measurements (Kim and Kwan 2019).

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

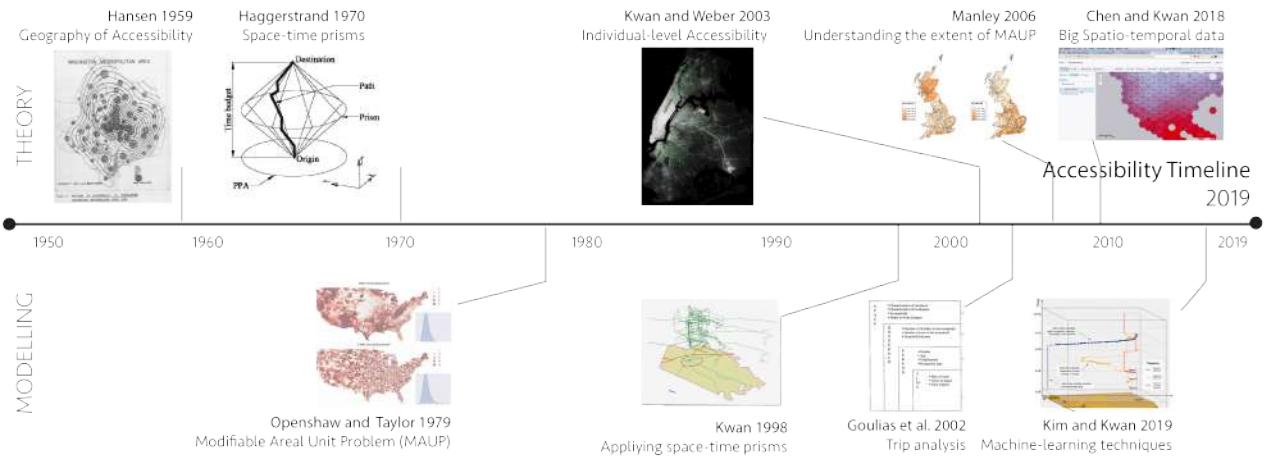


Figure 2 The notable academics and their key approaches to measuring accessibility from 1950 to 2019.

The key literature can be categorised as two approaches; location based measures and person based measures. On the one hand, location based measures describe the relationship between accessibility and residential segregation, such as Glaeser *et al.*'s (2008) analysis of the suburbanisation of poverty, and lower rates of upward mobility for Hispanic and African Americans (Chetty *et al.* 2014). On the other hand, person based measures of accessibility capture its gender differences (Kwan 1998: 191). Regardless of whether the literature reflects the spatial distribution of accessibility or individual accessibility, the academic field is developing an appetite for new methodologies using contemporary techniques and novel datasets (Chen *et al.* 2016, de Bruijn *et al.* 2018).

I develop a systematic literature review based on Kwan *et al.*'s (2003) framework which assesses the “simultaneous treatment of representation, methodology and application” (2003: 129). I adopt this evaluative structure in my analysis of accessibility to evaluate whether there is an accurate representation of accessibility, whether accessibility modelling is valid, and if there is a clear application of accessibility research.

2.3 Representation

I will analyse representations in three respects, spatially aggregated accessibility, individual accessibility, and space-time representations of accessibility. I summarise these three representations of accessibility in terms of their operational efficiency (ability to simplify the

representation of accessibility) and validity (accurate representation of accessibility) (figure 3).

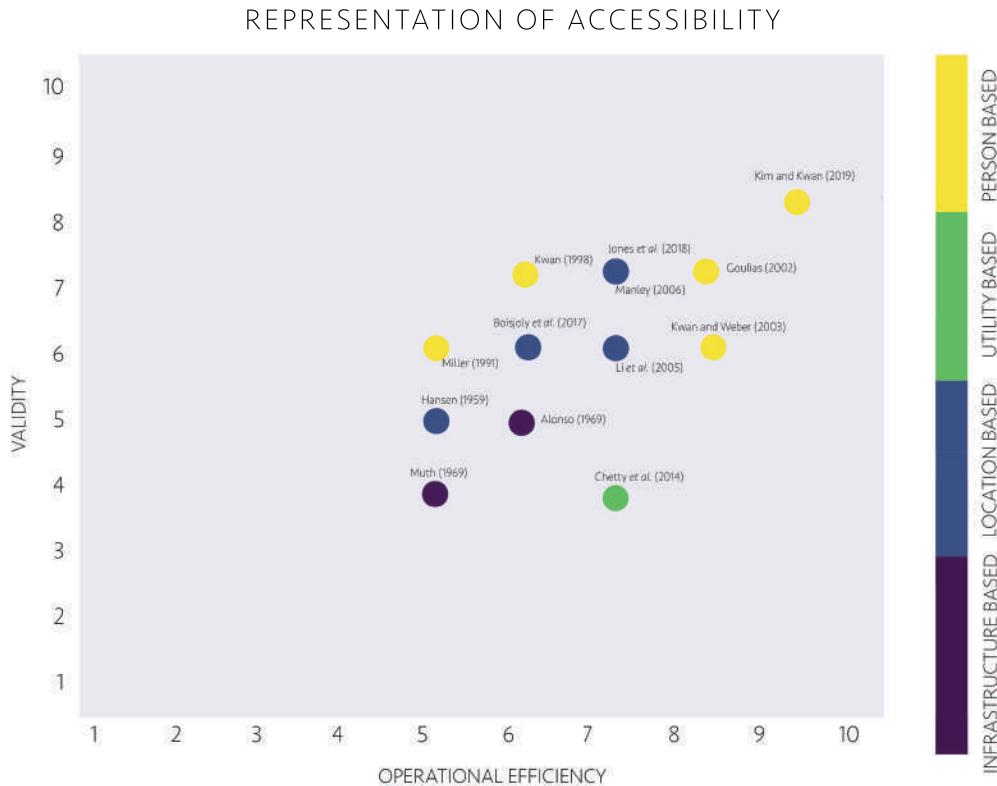


Figure 3 Graph showing representation scores for each key piece of literature. The node colour indicates the type of accessibility indicator the literature uses (infrastructure based, location based, utility based and person based).

2.3.1 Representing Space

Spatial accessibility can be analysed through a “zone-based, aggregate spatial framework (e.g. census tracts or block groups)” (Kwan *et al.* 2002: 130). More broadly, representing the spatial distribution of accessibility illustrates the latent structure of the location of individuals in relation to the location of employment opportunities. An example of expressing accessibility spatially is Chetty *et al.* (2014), who attribute accessibility to commuter zones, which are geographical aggregations representing local economies (USDA 2000). Chetty’s research uses large spatially aggregated datasets which rely on the assumption that processes determining the distribution of accessibility are only measurable at the regional level. Al-

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

though it may seem intuitive to represent access to employment by commuter zones, this is problematic as accessibility can be a function of household level processes, such as car ownership and income. Furthermore, the geography of accessibility in American cities is changing, therefore it is difficult to define the appropriate spatial aggregation at which spatial patterns of accessibility exist. The pivotal issue with aggregating accessibility to geographic units is the modifiable areal unit problem (MAUP) defined as the geographic pattern of accessibility that occurs only due to the areal unit that the data is collected (Openshaw and Taylor 1979). Within accessibility research the MAUP is of serious methodological concern, and presents a fundamental challenge to spatial aggregations of accessibility.

2.3.2 Representing Individuals

Answering the literature's lack of consensus on modifiable areal unit problem is the work of Mei-Po Kwan, who argues we should challenge the focus on place based measures of accessibility, and instead measure accessibility at the individual level. Kwan's research shifts our understanding of the MAUP to acknowledge there are multiple uncertainties introduced when measuring accessibility using the concept of UGCoP (The Uncertain Geographic Context Problem). For example, Weber and Kwan (2003) argue that spatially aggregated measures of accessibility are no longer able to cope with the complex relationships between "urban form, mobility and individual accessibility" (2003: 341). Although new data collection methods allow accessibility to be measured at the individual level, arguably, Weber and Kwan overemphasise the individual as the sole determinant of access to spatially distributed activity. Adopting this stance is Manley (2006) and Jones *et al.* (2018), who argue the social relationships regulating individual accessibility can only be observed "through the analysis of data at an aggregate level" (Manley 2006: 15). Responding to this, Kwan and Ren (2008) argue both geography and individual experience is represented by "space-time accessibility" (2008: 109).

2.3.3 Representing Space and Time

A space-time framework, originally developed by Hägerstrand (1970), attempts to measure the “behavioural possibilities of individuals” (Miller 1991: 287). A space-time framework suggests there are spatial and temporal limitations on individual accessibility, represented by the “space-time prism” (Miller 1991: 291). A space-time prism models individual activity participation at different locations throughout the day; an approach that values activity sequences rather than spatial aggregations of activity concentration. Recent representations of space-time accessibility include Kim and Kwan (2019), who illustrate that individuals living in the same residential location experience different exposure to congestion due to their “idiosyncratic activity-travel patterns” (2019: 15). The advent of “big spatiotemporal data” (Chen and Kwan 2018: 1), defined as data measured at an exceptionally granular level through space and time, allows researchers to represent space-time accessibility patterns more accurately. As this data becomes more available and computationally efficient to process, it is likely space-time representations will grow in importance in the accessibility literature.

2.4 Methodology

The methodologies for analysing accessibility are assessed in terms of their operational efficiency (model parsimony) and validity (the model’s incorporation of all significant predictors influencing accessibility) (figure 4).

2.4.1 Aggregated Modelling

During the 20th century, aggregated modelling measured accessibility as distance to the Central Business District (CBD) from the areal unit the resident occupies (Harris and Ullinan 1945, Alonso 1964, Muth 1969). The assumption underlying this measurement of accessibility is that the further away from the economic centre a resident lives, the less accessibility they have to spatially distributed activities. Although geographic proximity to spatially distributed activities is in no sense the sole determinant of accessibility, there is a strong correlation between access and centrality, in particular when measuring accessibility.

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

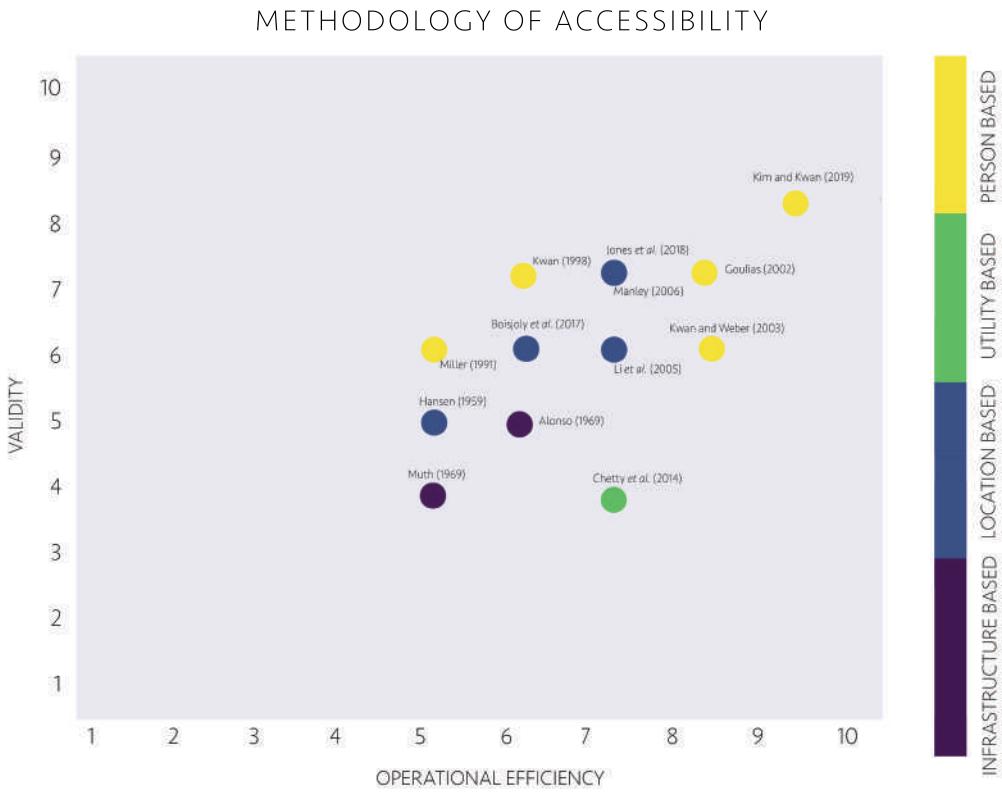


Figure 4 Graph showing methodology scores for each key piece of literature.

ility at small travel times. However, this simplification of accessibility results in a crude understanding of actual locational attractiveness, and there has been a dramatic shift in the explanatory capacity of the urban form to be an “organising principle” of accessibility (Kwan and Weber 2003: 343). For example, the construction of freeways through large American cities reduced the place based advantage of living in central metropolitan areas.

2.4.2 Multilevel Modelling

In transportation literature, multilevel analysis has heavily focussed on time-based accessibility to measure activity participation (Goulias 2002, Kwan *et al.* 2003), because multilevel analysis already lends itself well to explicitly model the spatial dependency of accessibility outcomes. For example, Hong and Goodchild (2014) use a multilevel model (MLM) to specify the spatial and temporal structure of accessibility to infer the relationship between land use patterns and transport emissions in urban and suburban locations. The multi-

Christina Last

level framework provides enough flexibility to analyse accessibility across space and through time in a two level model, with timestamps nested in spatial locations. Furthermore, multi-level modelling can account for both individual level accessibility within spatially aggregated units, as Weber and Kwan (2003) locate space-time accessibility values for individuals within neighbourhoods. Multilevel research offers a successful framework to model place based accessibility, but practitioners must specify the spatial or temporal structure of the data and cannot measure accessibility only at a single level.

2.4.3 Machine Learning

Although many geographers argue that the spatial level is where the fundamental relationships of accessibility are played out (Manley 2006, Jones *et al.* 2018), multilevel modelling cannot account for dynamics that are only measurable at the individual level. Complementing this pivot away from place based measures of accessibility is the development of georeferenced activity-travel data and efficient GIS-based methodologies to analyse big datasets (Chen *et al.* 2018). For example, Kim and Kwan (2019) collect real-time traffic congestion and activity-travel data of 250 individuals in Los Angeles to analyse exposure to congestion, and conclude that only individual activity trajectories can reveal the true variation in exposure to traffic congestion. An example of using big data to inform predictions is Li *et al.* (2019), who develop a machine learning algorithm to predict the location of individuals and their likely future activity spaces. Both studies move dramatically away from theory-driven approaches like multilevel modelling, and instead apply data-driven approaches to analyse individual accessibility. However, there is little consensus on the cleaning, analysis and reporting of results for big data, and the literature needs to develop a peer-reviewed framework to standardise analysis.

2.5 Applications

After discussing the issues with multiple representations and methodological concerns of analysing accessibility, I assess the social and economic applications of measuring accessibility. This is evaluated in terms of their operational efficiency (ease at which the measurement

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

of accessibility can be applied to urban and transportation planning) and validity (how meaningful the application is) (figure 5).

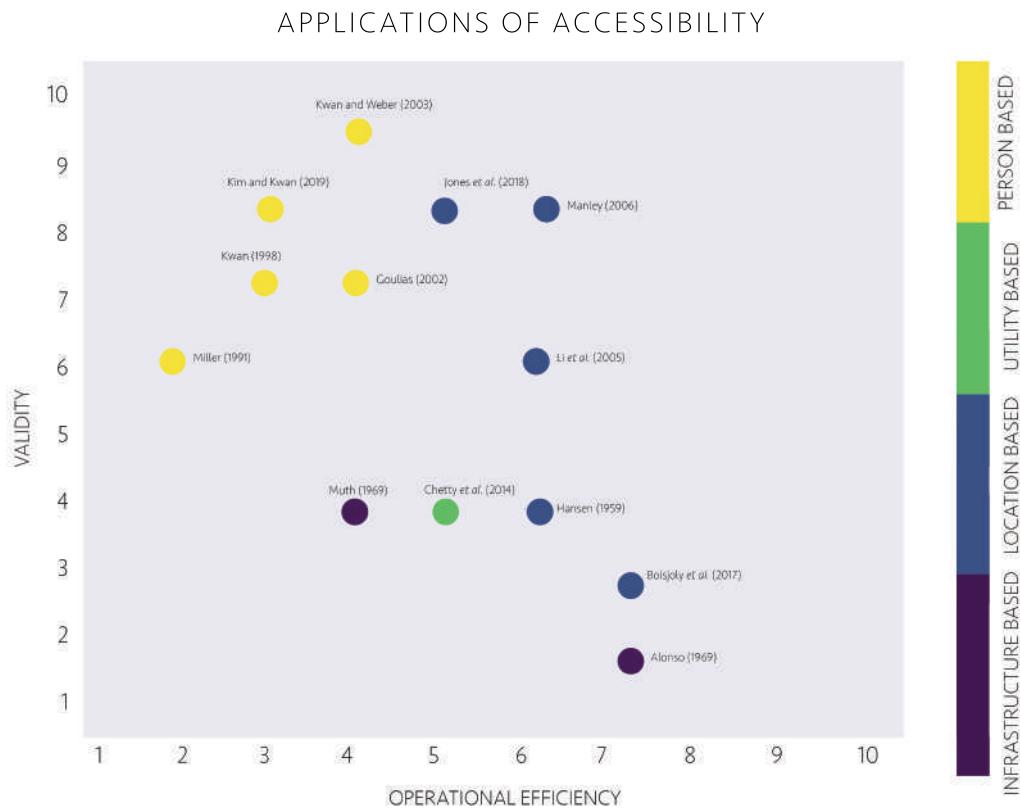


Figure 5 Graph showing the application scores for each key piece of literature.

2.5.1 Accessibility as an economic indicator

In 2012, although 23 percent of lower income residents lived in central urban locations, suburban poverty experienced a 139 percent increase between 2000-2012 (National Center for Law and Economic Justice 2014). These low income, suburban neighbourhoods face a compilation of challenges, such as high crime rate, poor health, low-performing schools and low job density, all of which are exacerbated by lacking transit options. Communities who experience low job accessibility are likely to earn a low wage, thus have reduced opportunities to achieve upward mobility (Kain 1968, Kasarda 1989, Wilson 1996). More recently, Chetty *et al.* (2014) performed a correlation analysis on spatially aggregated accessibility data,

Christina Last

showing that communities located in central city areas experience higher rates of upward mobility ($R^2 = 0.6$). Furthermore, many lower income households who relocate to suburban neighbourhoods do so to raise children, increasing the rate of suburban childhood poverty. The 2009 National Household Travel Survey show that households with children tend to travel over twice the distance than households without children, thus, there is strong evidence suggesting suburban, low income families spend a higher proportion of their income on transportation expenses (FHWA 2014). Households spend 33 and 16 percent of their average annual expenditure on housing and transportation, and as both costs continue to rise, lower income families are likely to face an increased cost burden for even basic travel expenses (Bureau of Labor Statistics 2017).

2.5.2 Accessibility as a social indicator

More recently, literature has focussed on the ethnic makeup of neighbourhoods and their location in American cities (Van Wee and Geurs 2011, Cort *et al.* 2014). National statistics indicate that African American and Hispanic households are disproportionately represented in low income, suburban neighbourhoods, and these cost-sensitive ethnicity groups experienced poverty rates of 27 and 26 percent respectively, when the overall poverty rate is much lower at 14 percent (DeNavas-Walt *et al.* 2012). Furthermore, Hispanics have the highest person and vehicle trips of all ethnic groups, and experience the lowest mobility and highest levels of poverty (FHWA 2014). Due to the residualisation of low income minority groups, the most recent National Household Travel Survey concludes it is essential for minorities have access to affordable and effective transportation options enabling them to access job opportunities. Nowhere is this more pertinent than in large American cities, as 44 percent of Hispanics live in the ten largest metropolises, and in Los Angeles the Hispanic population is 48 percent (FWHA 2014, U.S. Census Bureau 2012). Upward mobility is the ultimate goal of accessibility (Vickerman 1974), and therefore it is essential we consider the social and economic barriers to implement effective transportation policy.

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

2.6 Evaluating the Research Gap

After taking the average scores for the representation, methodology and application of accessibility for the key literature, I visualise the combined score on a 3D interactive plot² (figure 6). The plot represents the literature's average score for all three sections with the node size scaled by their standard deviation. Notably, there is a strong linear relationship between all three framework sections with person based and location based measures of accessibility performing highly. Location based measures of accessibility show a small standard deviation (suggesting high performance in all categories), which led me to adopt a location based measure of accessibility. However, what both person and location based measures fail to achieve is an accurate representation and useful application of accessibility. To narrow this research gap, I develop a simple methodology employing a novel machine learning technique to accurately represent accessibility patterning in Los Angeles.

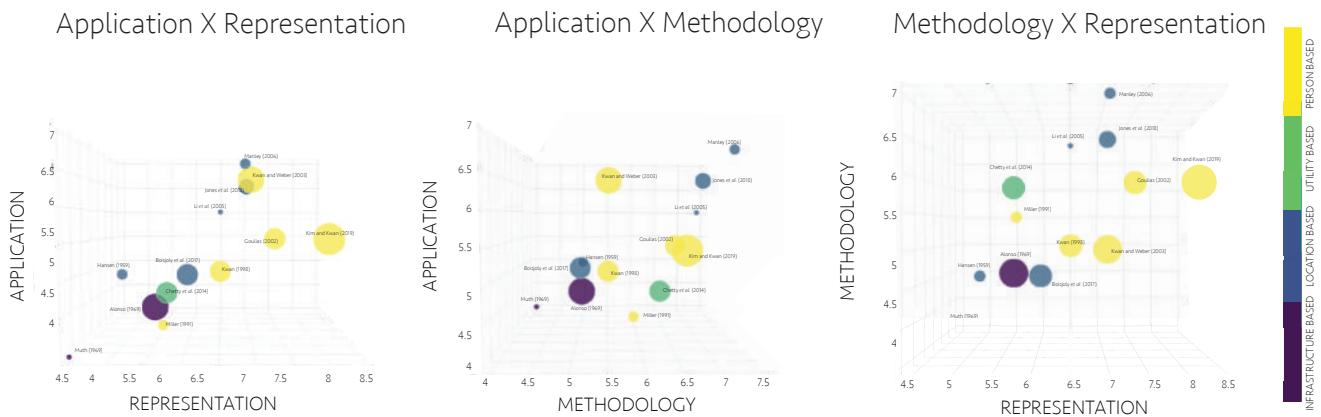


Figure 6 A 3D scatterplot showing the combined scores for the accessibility literature, showing three axis views of (a) application and representation, (b) application and methodology and (c) representation and methodology, with each literature node scaled by their standard deviation.

² Interactive version: <https://github.com/ChristinaLast/Equity-Accessibility-Los-Angeles>

2.7 Narrowing the Research Gap

The recent development of spatially granular accessibility data provides a rich evidence base to test the relationship between ethnicity and their household and transportation expenditure burdens. Additionally, there has been limited research comparing the predictive accuracy of multilevel modelling (MLM) and machine learning methods (Feng and Jones 2015), and none developing a clear methodology for analysis. Furthermore, there is no agreement on whether larger or smaller training data is more suitable for model training. Finally, there has been slow adoption of novel modelling approaches that account for the “spatial spillover effects” of low neighbourhood equity in the transportation literature.

To address this gap, I combine data from the Center for Neighborhood Technology (2017), the US Census Bureau (2010) and Chen *et al.’s* (2011) Opportunity-Based Accessibility indicators to analyse the variety of social, environmental and economic parameters influencing accessibility patterns. To improve the previous approaches to modelling accessibility, I build on Feng and Jones’ (2015) comparison of MLM and ANN by incorporating spatial spillover effects using an SLX model (Vega and Elhorst 2015). Introducing a spatial weights matrix into the multilevel model and the neural network evaluates the spatial spillover of low equity neighbourhoods while maintaining model parsimony. The results contribute important findings on the location of ethnic communities with low accessibility to the Metro 2028 Vision Plan and the Los Angeles Sustainability Plan (Petersen *et al.* 2018), both recent initiatives to improve access to high quality mobility options in Los Angeles County. Overall, this research aims to push our current ability to model location based accessibility with existing techniques, by adopting machine-learning modelling approaches that account for spatial effects.

2.8 Research Questions

Having motivated the research gap and modelling strategy, the following list presents the specific research foci I seek to answer:

1. To what extent do household and transportation costs as a percentage of total income matter in determining accessibility to employment opportunities?

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

2. Is there evidence of spatial spillovers that explain differences in accessibility to employment opportunities?
3. What is the relative effect of being a Hispanic or African American in determining the level of accessibility to employment opportunities?

3 Data and Methodology

This paper compares two advanced modelling approaches, artificial neural networks (ANN) and multilevel modelling (MLM) to model spatial patterns expenditure burdens faced by different ethnic groups and their effect on accessibility to employment in Los Angeles County. An artificial neural network is a modelling procedure that imitates the connections between biological neural networks, whereas a multilevel model is defined by the organisation of model parameters into different hierarchical structures. To account for the structural spillover effects of spatial units, I introduce a spatial weights matrix. I evaluate the model performance by mapping the residuals and the error values from the accessibility predictions.

3.1 Data overview and Setting

This study selects Los Angeles County as the study area there are a diverse range of transportation options and a mixture of urban and rural land use incorporating both affluent suburbs and inner city areas. The 2009 National Household Travel Survey shows that in Los Angeles, an individual's travel radius is affected by their level of income and ethnicity. For example, individuals earning over \$100,000 have a 23 mile travel radius, whereas for individuals below the poverty line this is only 11 miles (FWHA 2014). According to 2010 Census data, 53 percent of Hispanic and 25 percent of African American households are at or below the poverty line in Los Angeles County, compared to a county average of 15 percent. Compounding the effects of this limited travel radius, Los Angeles County also accounts for 481 million unlinked passenger trips, the third highest in the US, showing how restrictive travel choice can be for ethnic minorities in impoverished neighbourhoods.

To explore the relationship between equity and accessibility for both Hispanics and African Americans in Los Angeles County, this study uses equity and ethnicity data to predict local accessibility. The null model tests the hypothesis that there is no difference in local accessibility in Los Angeles (measured as the number of jobs accessible by private vehicle over 10 minutes of travel time on the road network), using a multilevel model framework (Table 1). Model 1 introduces household and transportation expenditure as a percentage of income, to predict accessibility in Los Angeles County. To test whether there are "spatial

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

spillover effects” of experiencing a similar level of equity burden as nearby neighbourhoods, I introduce a spatial lag of household and transportation expenditure in Model 2. In Model 3, I test if there is a relative effect of being in a high percentage African American or Hispanic neighbourhood on accessibility. To do this, I introduce the percentage of tract who are Hispanic and African American alongside household and transportation expenditure to predict local accessibility.

Model	Hypothesis	Equation	Variables Added
Null Model	There is no difference in accessibility between blocks, block groups and tracts.	$y_{ijk} = \beta x_{ijk} + \epsilon_{ijk}$	The number of jobs accessible by private vehicle over 10 minutes on the road network. (βx_{ijk})
Model 1	The percentage of income spent on household and transportation costs has an effect on accessibility to employment opportunities.	$y_{ijk} = \alpha_{jk} + \beta x_{ijk} + \epsilon_{ijk}$ Where $\alpha_{jk} = \theta z_{jk} + \eta_k + v$	Percentage of income spent on household and transportation costs. (α_{jk})
Model 2	The spillover effect of household and transportation costs from neighbouring geographic units has an effect on accessibility to employment opportunities.	$y_{ijk} = \alpha_{jk} + W_{jk}\alpha\theta + \beta x_{ijk} + \epsilon_{ijk}$ Where $W_{jk} = \frac{1}{d_{jk}^\gamma}$	The percentage of income spent on household and transportation costs. (α_{jk}) The spatially lagged value for the percentage of income spent on household and transportation costs. (W_{jk})
Model 3	There is a relative effect of being in high percentage African American or Hispanic neighbourhood on accessibility to employment opportunities.	$y_{ijk} = \alpha_{jk} + \eta_k + \tau + \beta x_{ijk} + \epsilon_{ijk}$ Where $\eta_k = \Gamma\xi_k + \nu$ $\tau_k = \Upsilon\omega_k + v$	Percentage of income spent on household and transportation costs (α_{jk}) The percentage of a census tract who are Hispanic. (η_k) The percentage of census tract who are African American. (τ_k)

Table 1 The hypothesis equation and description of the variables sequentially introduced into each model.

3.2 Variables

3.2.1 Response Variable

I make use of block level accessibility data produced from the Southern California Association of Governments (2001), who use a “second-by-second and parcel-by-parcel modeling and simulation” (Chen *et al.* 2011: 58). There is accessibility data for 16,470 blocks, indicating high accessibility (over 100,000 jobs) in Downtown Los Angeles, and under 62,000 in suburban locations, particularly in north and east Los Angeles County (Figure 7). This data is referred to as ‘true accessibility’ throughout the study, and is used to draw conclusions on the accuracy of the MLM and ANN predictions. To construct this accessibility index, Chen *et al.* (2011) develop a methodology using GIS to measure accessibility by cumulatively counting the instances of activities within travel time thresholds (10, 20, and 50 minutes) from each census block centroid. In this study I am interested in local accessibility, therefore I select total accessibility after 10 minutes of travel on the road network, and use the 11am-12pm low period to measure ambient accessibility. Over 25 percent of trips were completed in 7 minutes in the 2001 SCAG Regional Travel Survey, giving a good representation of local access (Chen *et al.* 2011: 63). I acknowledge as this data is modelled rather than directly measured, so there may be inconsistencies between the accessibility data and actual accessibility.

3.2.2 Equity Predictors

An important distinction between equity and equality is made by Van Wee and Geurs (2011: 351) who argue access to jobs can be unequal, however only the systematic exclusion of a certain population makes the distribution of accessibility inequitable. To measure equity, I include the percentage of income spent on Housing and Transportation costs (HTI) in the model specification. The HTI data is collected at the block group level and is constructed from data measuring “auto ownership, auto use and transit use” (Center for Neighbourhood Technology 2017: 6). Residents in Downtown Los Angeles spend around 10 percent of income on household and transportation, whereas areas in suburban Los Angeles spend between 60 to 80 percent of income on HTI costs (Figure 8). The distribution of both accessibility

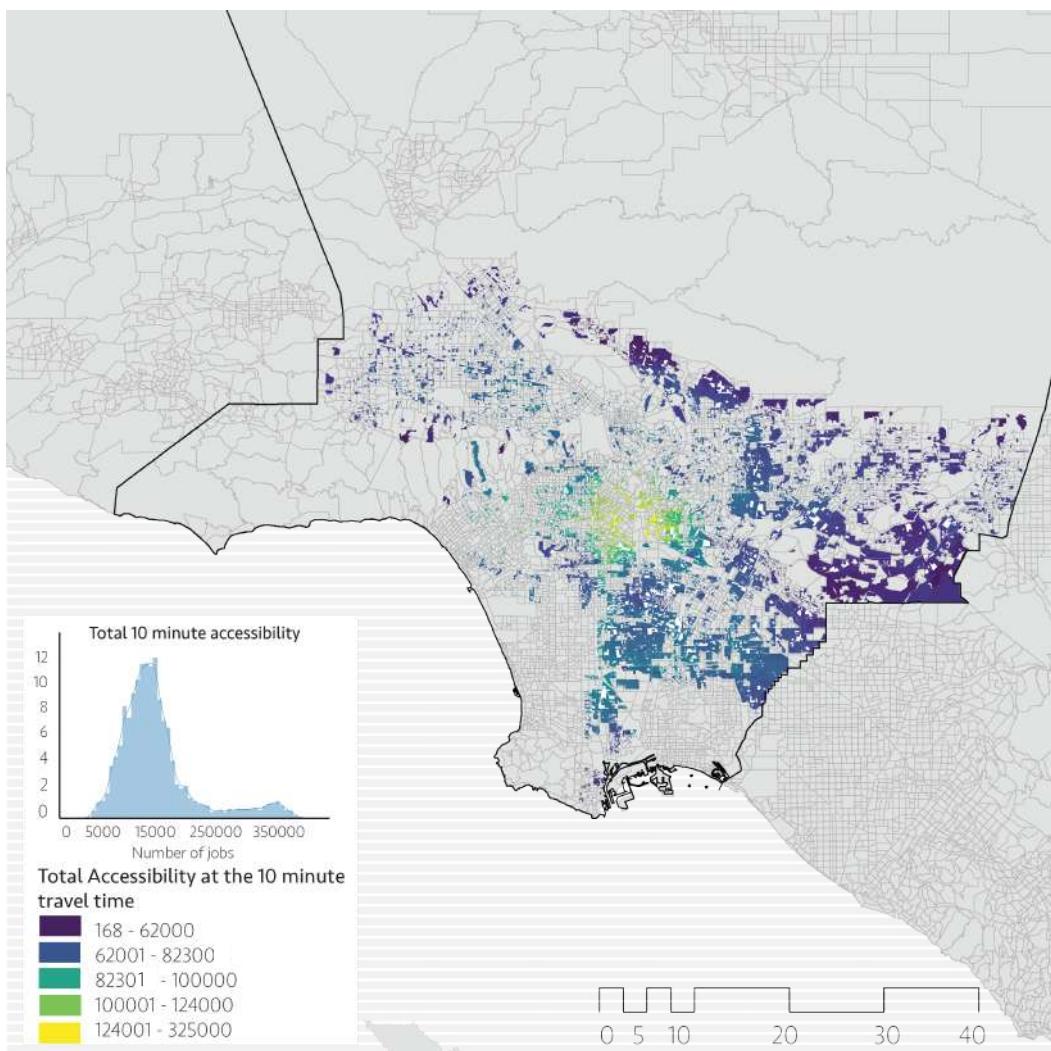


Figure 7 The distribution of total accessibility by private transportation after 10 minutes of travel time.

and HTI is similar, reflecting the significant influence transportation expenditure has on accessibility.

3.2.3 Ethnicity Predictors

Households that report a high percentage of income spent on household and transportation costs are disproportionately represented by African Americans and Hispanics (FWHA 2014). In the USA, these ethnic groups experience limited vehicle availability and less affordable transport options. To measure whether there is an inequitable distribution of access to jobs for different ethnicities in Los Angeles County, I introduce the percentage of the census tract

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

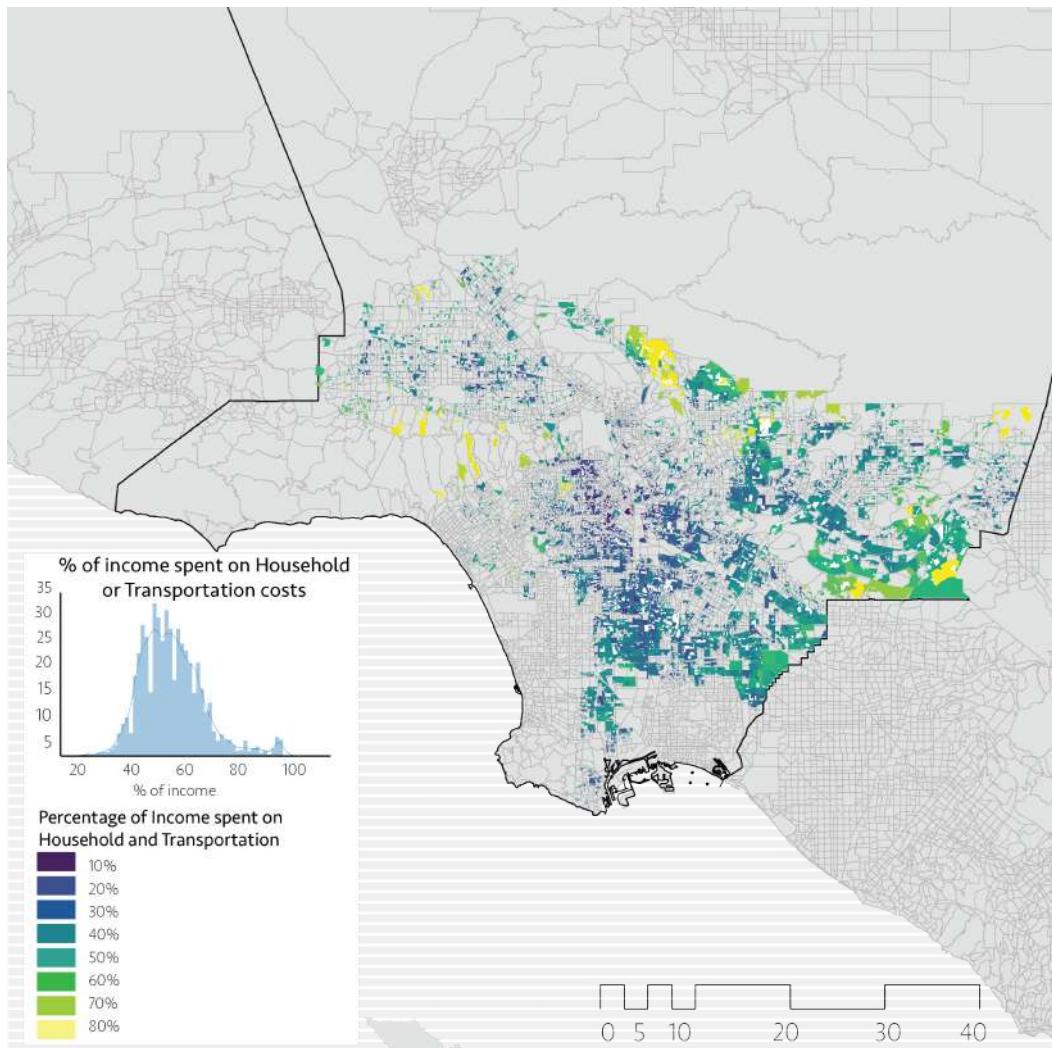


Figure 8 The distribution of household and transportation costs (% of income).

population who are Hispanic or African American. The majority of census tracts in south central Los Angeles are over 50 percent African American (Figure 9a). In contrast, figure 9b shows Hispanic populations are more spatially distributed, but there are many tracts in east Los Angeles with over 80 percent Hispanic population.

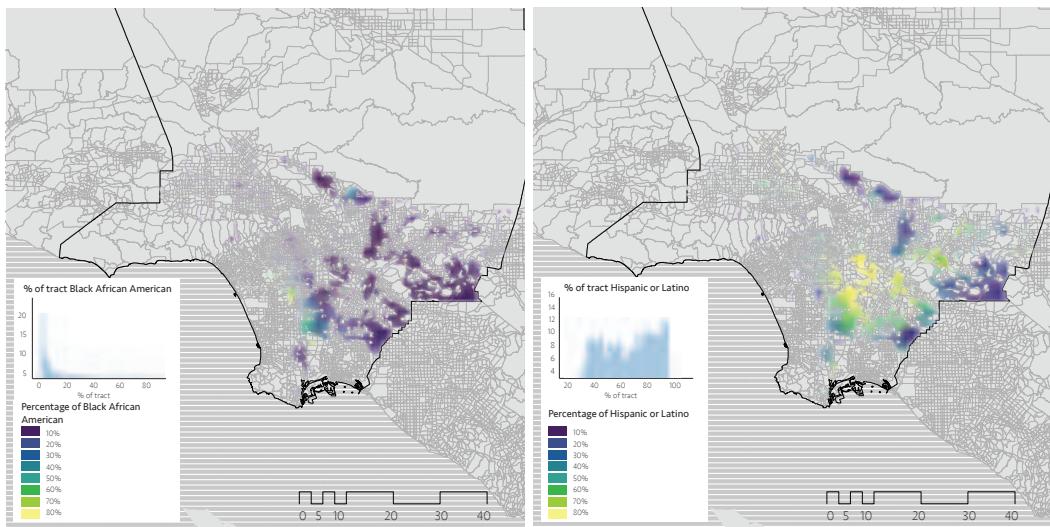


Figure 9 The distribution of (a) African American and (b) Hispanic population in Los Angeles County.

3.2.4 Spatial Weights Matrix

Dong *et al.* (2019) argue that “geography exists both ‘vertically’... and ‘horizontally’” in the sense that geographic spillover effects between locally clustered areal units exist (2019: 119). Previously, standard econometric techniques have difficulty parameterising the “spatial spillover effects” of a neighbouring geographic unit. Vega and Elhorst (2015) introduce the SLX model to incorporate spatial effects into advanced modelling procedures, allowing changes to explanatory variables to impact neighbouring geographic units. I incorporate the SLX model to measure the “spatial spillover effects” of experiencing a similar level of equity burden as near neighbourhoods. The logic underpinning this is the more a block spends on household and transport costs, the more likely neighbouring blocks will experience an increase in household and transportation costs, through rent increases or fare increases on transportation. The SLX is beneficial as it measures the magnitude and significance of spillovers in a flexible and parsimonious specification of the spatial weights.

3.3 Methodology Overview

3.3.1 Multilevel Model

As indicated in Chapter 2, multilevel modelling (MLM) is an invaluable approach to answering research questions with hierarchically organised data. More specifically, multilevel models analyse nested variability as a function of explanatory variables. Two important advantages of MLM is first, its flexibility in controlling for characteristics at multiple layers of analysis, and second, its ability to borrow strength from other data points.

To analyse accessibility in a multilevel model, I first visually describe accessibility and organise the data into a nested data structure (Figure 10). After replicating this structure in all model specifications (Technical Appendix, Section 8.1) I incorporate an additional parameter to answer each research question. Then I map the model residuals and model errors.

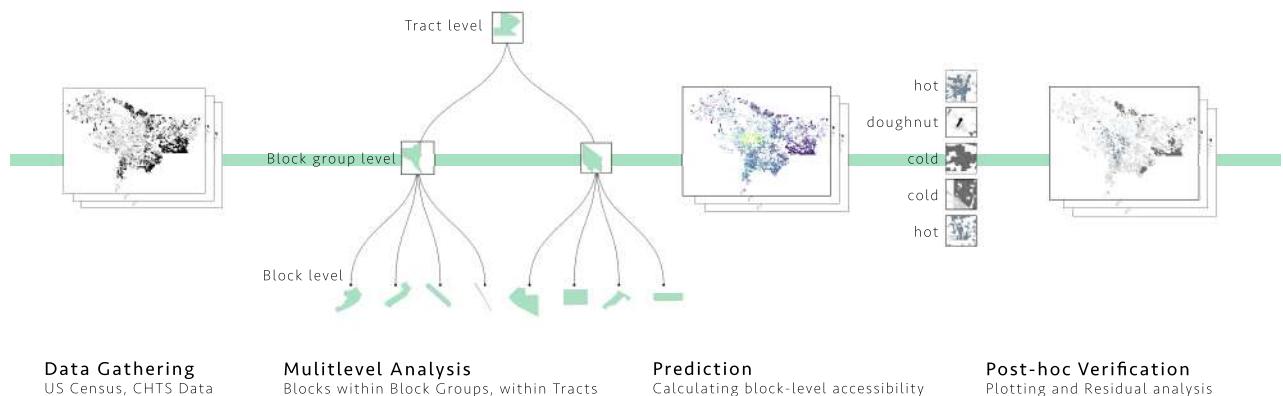


Figure 10 Figure showing the process of the multilevel analysis.

3.3.1.1 Model Specification

All multilevel models were specified in PyStan and fit using Hamiltonian Monte Carlo. To structure a multilevel framework for a nested three-level model, I specify the between-block equation, the between-block group equation, and the between-tract equation. The combined

Christina Last

three level model is represented at the block level by:

$$\hat{y} \sim N_N(\alpha + \beta x, \sigma_\epsilon^2) \quad (1)$$

At the block group level:

$$\hat{\alpha} \sim N_J(Z\theta + \eta, \sigma_\theta^2) \quad (2)$$

At the tract level:

$$\hat{\eta} \sim N_K(P\zeta, \sigma_\zeta^2) \quad (3)$$

And the error terms are given such that:

$$\hat{\epsilon} \sim N_N(0, \sigma_\epsilon^2);$$

$$\hat{\theta} \sim N_J(0, \sigma_\theta^2);$$

$$\hat{\zeta} \sim N_K(0, \sigma_\zeta^2);$$

The model structure includes y ; the estimated total accessibility after 10 minutes of travel time (TOTR10) at the block level N , J and K are block group and tract level indicators. η is the tract level influence on accessibility, and α the block group level influence. N is number of blocks in J block groups and K tracts and is constrained to be positive. β , Z and P correspond to the vectors of the regression coefficients I will estimate. The terms ϵ , θ and ζ are the residuals at the block, block group and tract levels, assumed to follow a normal distribution corresponding to the variance in σ_ϵ^2 , σ_θ^2 and σ_ζ^2 .

3.3.2 Artificial Neural Network

Artificial neural networks (ANN) are a form of artificial intelligence designed to simulate biological neurons to process information, such as data classification and pattern recognition. To analyse accessibility data using ANN, I build a sequential, feed-forward neural network

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

using two dense layers and visualise the residuals (Figure 11). To analyse the results, I map the model residuals and error to identify significantly high or low predictions of access to employment.

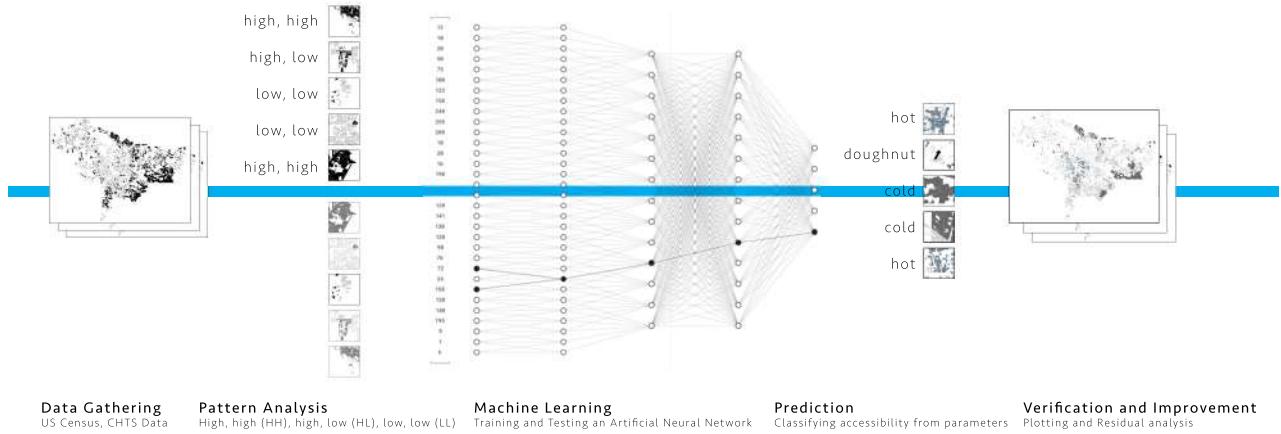


Figure 11 Figure showing the process of building and analysing the artificial neural network.

3.3.2.1 Model Specification

An artificial neural network is constructed by connecting processing nodes to create hidden network layers between an input and output layer. Each layer is made of neurons, and the strength of the connection between the neurons is a function of a set of weights. The weight w_{ij} depends only on information locally available: the state properties of the neurons i and j . For example, if neurons i and j are in the same state w_{ij} will increase, and if they are different w_{ij} will decrease (Coolen 1998). This definition can be written as:

$$S_i = S_j : \uparrow w_{ij}$$

$$S_i \neq S_j : \downarrow w_{ij}$$

$$w_{ij} \rightarrow w_{ij} + S_i S_j$$

3.4 Testing Predictions

I use Bayesian inference to quantify the uncertainty in my predictions of accessibility from the probability distribution of prior estimates. However, as it is impractical to calculate the probability distributions at each iteration of the model, I take advantage of the stochastic sampling method Markov Chain Monte Carlo (MCMC) using the probabilistic programming language Stan³. Due to the introduction of variance in the regression slopes for all predictor variables, MCMC estimates are beneficial as the predictions “yield exact inferences in situations where maximum likelihood only provide approximations” (Fielding and Goldstein 2006: 46).

To check for robustness in the multilevel model I evaluate a series of trace plots for each parameter introduced into the model. Uncertainty during Bayesian estimation is captured in the spread of the posteriors described in each traceplot. In addition to the traceplots, I evaluate the success of the multilevel model and the artificial neural network in terms of their goodness of fit (measured by R^2), predictive accuracy, measured as the Mean Absolute Error/Mean Absolute Percentage Error (MAE/MAPE) and explanatory capacity. To evaluate each model’s explanatory capacity, I analyse the residual plots and error plots for each modelling technique.

3.5 Software

All data cleaning, imputation, and visualisation was conducted in Python 3.7⁴. I use PyStan⁵ to build and run the multilevel model, and Tensorflow⁶ to build the neural network. PyStan provides a Python interface to Stan, a package for Bayesian inference and sampling strategy using Hamiltonian Monte Carlo estimation. Stan is more efficient in high dimensional problems where there are a large number of predictors to infer. This lends itself well to my study, where there is over 29,000 rows of data.

³ <https://github.com/stan-dev>

⁴ <https://www.python.org/downloads/release/python-370>

⁵ <https://pystan.readthedocs.io/en/latest/>

⁶ <https://github.com/tensorflow/tensorflow>

4 Results

This section compares the results of a multilevel model (MLM) and an artificial neural network (ANN) using the aforementioned research questions (Question 1, 2 and 3). I evaluate the biases in the MLM and ANN by mapping clusters of predicted accessibility that are significantly different from the 'true' accessibility values (Chen *et al.* 2009).

4.1 Descriptive Statistics

I describe the variables included in the study in Table 2. The full model for both the artificial neural network and multilevel model contains three predictors measured at the block group and tract level.

4.2 Performance Measures

I adopt Feng and Jones' (2015) analysis framework to compare the goodness of fit, predictive accuracy and the explanatory power of the multilevel model and artificial neural network (Table 3). After evaluating the model fit, I discuss the predictor variables most influential on accessibility, introducing their coefficients and associated variability.

4.2.1 Goodness of Fit

A comparison of goodness of fit is presented in Table 3. A higher R^2 value indicates the model is more confident that the predicted accessibility is similar to the 'true value' of accessibility. The goodness of fit for the MLM does not improve because most of the variation in accessibility can only be explained by block level latent effects, therefore, the introduction of higher level (block group and tract level predictor variables) does not significantly explain the variation in block level accessibility. Alternatively, The R^2 value for ANN decreases from Model 1 to Model 2, yet increases in Model 3 when ethnicity parameters are introduced. The ANN results may show a reduction in goodness of fit when additional parameters

Table 2 Descriptive statistics for all variables introduced into the models.

Variable	Type	Descriptive statistics	Unit of Measurement
Structural Variables			
Blocks		16470 blocks	A unique identifier for each of the 16470 blocks
Block groups		3221 block groups	A unique identifier for each of the 3221 block groups
Tracts		1306 tracts	A unique identifier for each of the 1306 tracts
Dependent Variables			
Total Accessibility	Continuous	Min = 169 Max = 324814 Average = 102862 SD = 57227	Accessibility indicators are constructed by Chen <i>et al.</i> (2016), and are calculated using an activity-based travel demand model by the Southern California Association of Governments.
Predictor Variables			
Household and Transportation Costs (as a percentage of income)	Continuous	Min = 17 Max = 100 Average = 57 SD = 14	The percentage of income spent on Household and Transport costs are constructed from a combination of “auto ownership, auto use and transit use” data (Center for Neighbourhood Technology 2017: 6)
Percentage of tract that are Hispanic	Continuous	Min = 0 Max = 100 Average = 56 SD = 28	The percentage of Hispanic is calculated from the total tract population who are Hispanic divided by the total tract population.
Percentage of tract who are African American	Continuous	Min = 0 Max = 83 Average = 9 SD = 15	The percentage of African American is calculated from the total tract population who are African American divided by the total tract population.

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

Table 3 Table comparing the goodness of fit and accuracy measurements for the MLM and ANN using Mean Absolute error (MAE), Mean Absolute Percentage Error (MAPE) and R^2 .

	MLM			ANN		
	MAE	MAPE	R^2	MAE	MAPE	R^2
Null Model	102861	99.9	0.1			
Model 1	5917.20	8.05	0.97	Model 1	33263.50	45.2
Model 2	100868.86	97.76	0.1	Model 2	33445.33	46.75
Model 3	79914.64	76.90	0.15	Model 3	31794.47	43.76

are added to the model because there are too few neurons (4,160 and 65 neurons in two densely connected layers) which can restrict the representation the network learns, causing underfitting.

4.2.2 Predictive Accuracy

Accuracy comparisons between the MLM and the ANN are summarised in Table 3. In one instance, the MLM shows high predictive accuracy in Model 1, as the mean absolute percentage error (MAPE) for MLM is reduced from 99 percent to 8 percent, however this increases in Model 2 when the SLX model is introduced (as the MAPE increases to 97 percent). In the other instance, the ANN improves gradually, reducing from 45 percent to 43 percent in Model 3. Other than Model 2, ANN remains the best performer in each scenario with an average error of 136,566 jobs. This demonstrates the capacity for ANN to extrapolate relationships from a small training dataset, even when access to jobs is non-normally distributed across Los Angeles County.

4.2.3 Explanatory Capacity

In the following sections I evaluate the explanatory capacity of each model specification, by visualising the residual errors from each model specification in comparison to the actual distribution of accessibility (TOTR10). I calculate the residual error by subtracting the accessibility prediction from the 'true value' of accessibility. To identify clusters of inaccurate predictions from the MLM and ANN I compute local error clusters from the predicted accessibility values, using a queen contiguity function. For any isolated predictions (where

queen contiguity cannot identify neighbouring blocks) I use the k-nearest neighbours algorithm to complete a matrix of predicted error values and their spatial lag. Using a local Moran's I clustering algorithm (Section 8.2, Technical Appendix), I obtain the local clusters of predicted accessibility that are significantly different from their 'true value', and visualise them using the classifications; hot spot, cold spot, doughnut, diamond and no significance. The maps simultaneously compare both errors from the MLM and ANN predictions with the 'true values' of accessibility.

4.3 Null Model

I specify the null model using a multilevel framework. In the multilevel model, the specification exposes all blocks to varying levels of accessibility. The average block has access to 24.77 jobs, however the between-block variance is 117,700. This suggests there is unaccounted variability between blocks in this model specification, thus confounding variables need to be modelled to explain block-level variation. The between block group variability is 1.49, substantially smaller than the block level variance, and variation explained at the tract level is almost 0 (Figure 12). Figure 13 shows the null model does not capture the spatial concentration of jobs in Downtown Los Angeles.

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

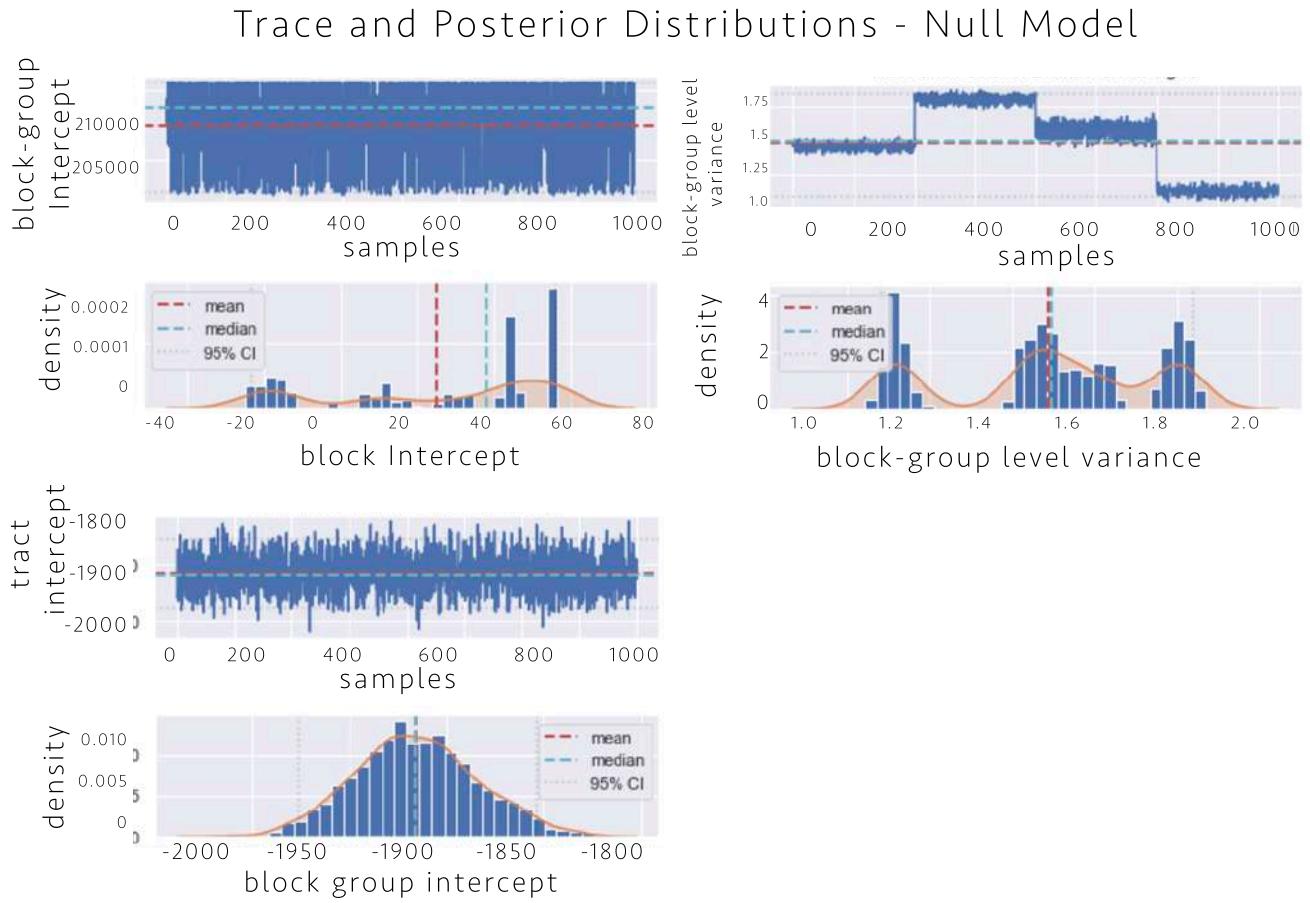
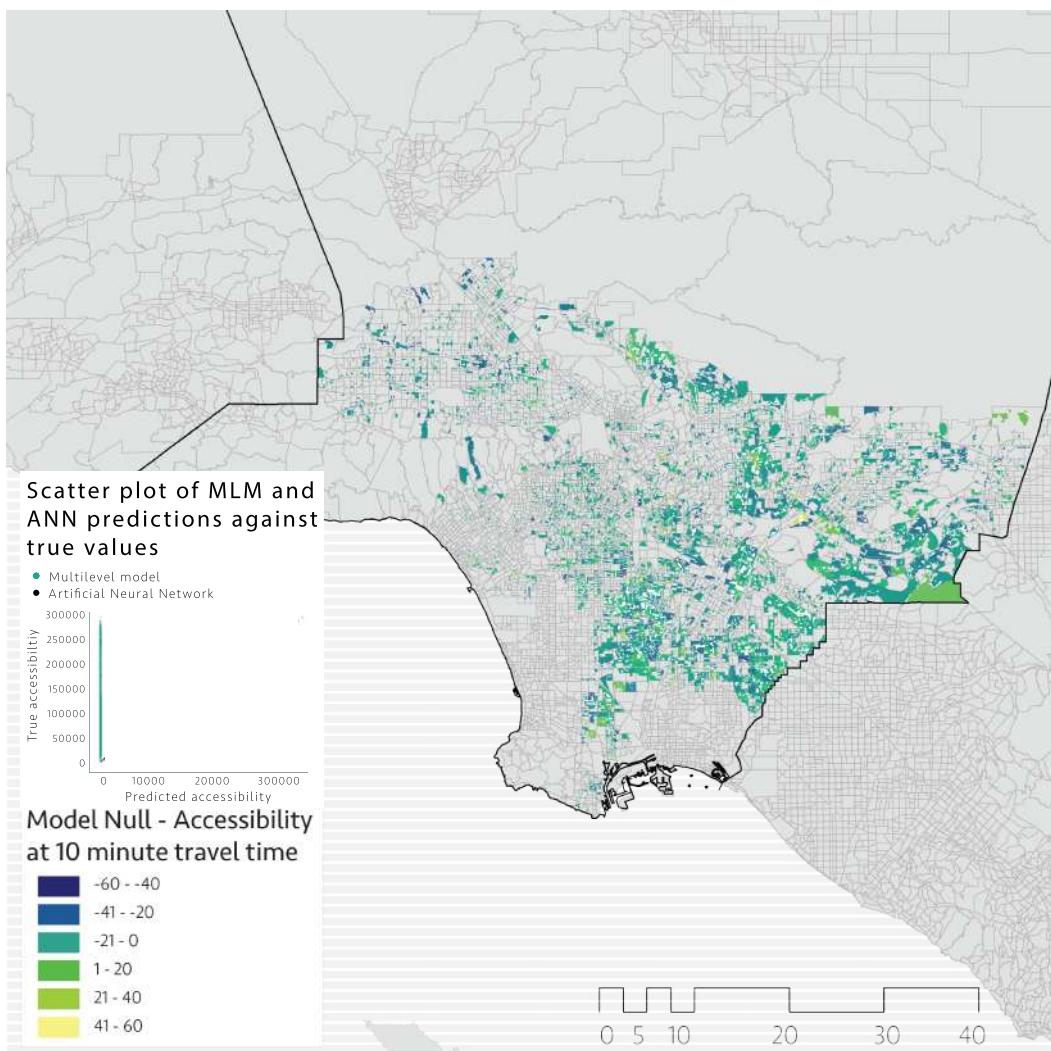


Figure 12 Trace and distribution plot for (a) block group intercepts; (b) tract intercepts, and (c) block-group level variance.



■ **Figure 13** Predictions for local Accessibility (TOTR10) from the Null Model.

4.4 To what extent do household and transportation costs as a percentage of total income matter in determining accessibility to employment opportunities?

Given my interest in exploring the relationship between equity and accessibility, Model 1 introduces the percentage of income spent on household and transportation costs (HTI) as block group level parameter. To check for uncertainty, I plot the potential regression line, as well as the trace and posterior distributions (Figure 14a). The plots of posterior distributions show the model estimates are close the modes of our distributions (Figure 14

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

b, c, d and e) and comfortably within the 95 percent confidence intervals. There is some bias between the peaks of the posteriors and the model estimates, which is a result of noise in the data.

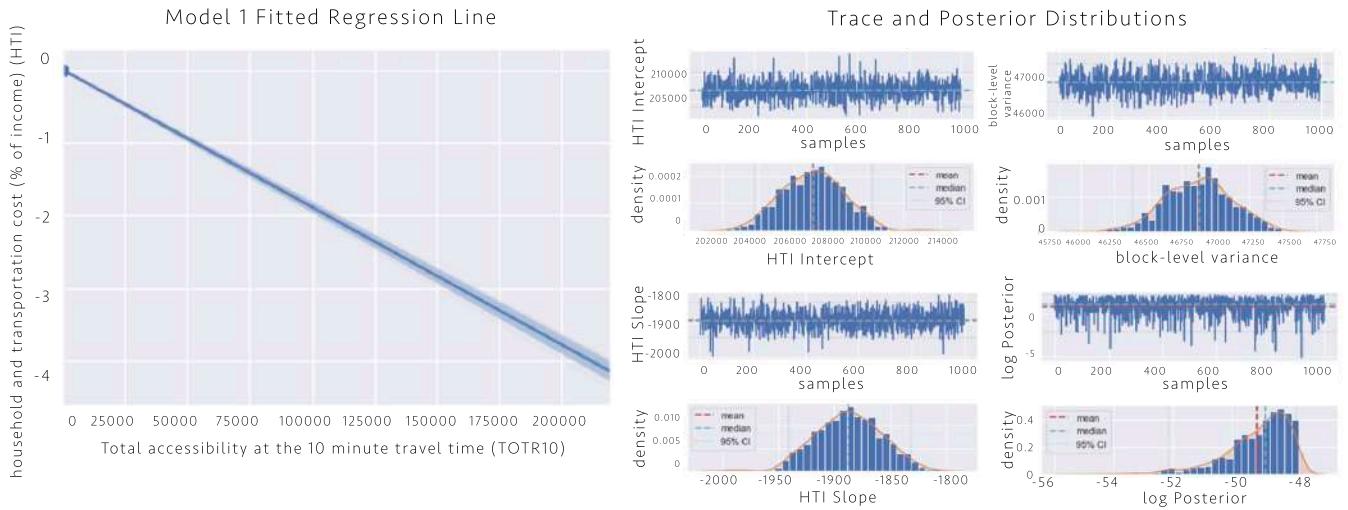


Figure 14 Regression plot for (a) Household and transport costs (HTI) and Accessibility (TOTR10), the trace and distribution plot for HTI intercept (b), for HTI slope (c), for the block group intercepts (d), the trace and posterior for the log (e).

In the MLM the coefficient for household and transportation costs is 11.4 percent of income, which can be interpreted as, for every 1 percent increase in household and transportation costs accessibility decreases by 11.4 jobs. The between-block variance shrunk from 117,700 to 8,890 jobs, suggesting the inclusion of HTI costs reduces block-level uncertainty. However, the block group variability only slightly increases to 2.66 and tract level variation increases to 1.86, which fails to sufficiently explain accessibility at both neighbourhood levels. Comparatively, the specification of the artificial neural network regresses HTI costs on accessibility. The average number of jobs accessible across all blocks is 103,178 jobs, however, due to the inability to partition variance at different levels, the ANN estimates block level variation in accessibility to be 35,691 (Table 6 , Model 1), larger than the multilevel model prediction (Table 4). Upon visual inspection, the multilevel model accurately captures the degradation of accessibility in suburban locations, as eastern Los Angeles County there are as few as 4,000 locally accessible jobs (Figure 15a). Although the neural network captures the spatial pattern of accessibility, explaining 39 percent of its spatial variation, the pre-

Christina Last

dictions become negative for suburban locations in north and east and south central Los Angeles (Figure 15b).

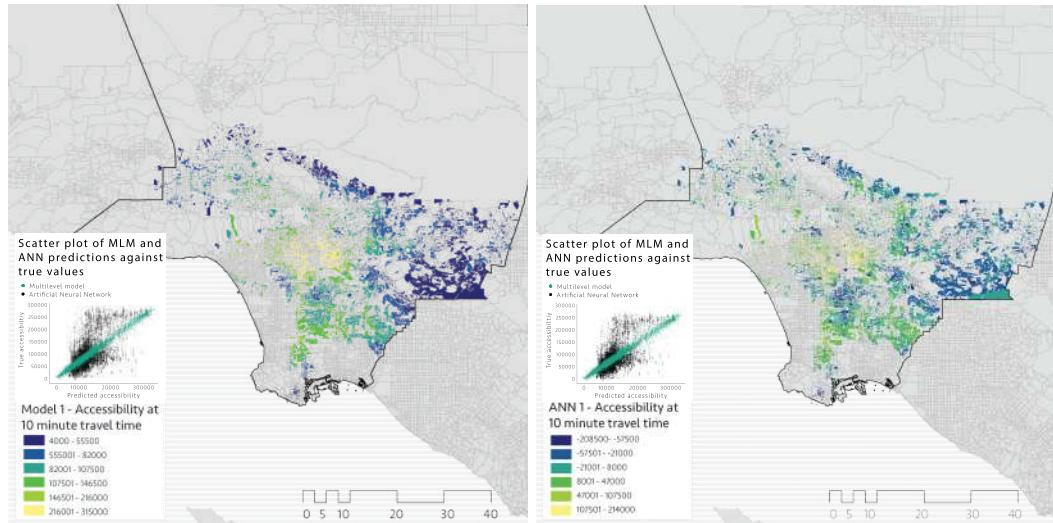


Figure 15 Model 1 predictions for MLM (a) and ANN (b) and a scatterplot to show the residual distribution.

Figure 16 overlays the errors for the multilevel model and the artificial neural network. It shows that both modelling procedures significantly under predict the accessibility in eastern Los Angeles County shown by the green error results (estimating under 95,000 jobs for the lowest 2.5 percent of blocks), and both MLM and ANN over predict some block level accessibility in Downtown Los Angeles shown by the yellow blocks, (estimating over 200,000 jobs) due to the right-skewed distribution of accessibility.

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

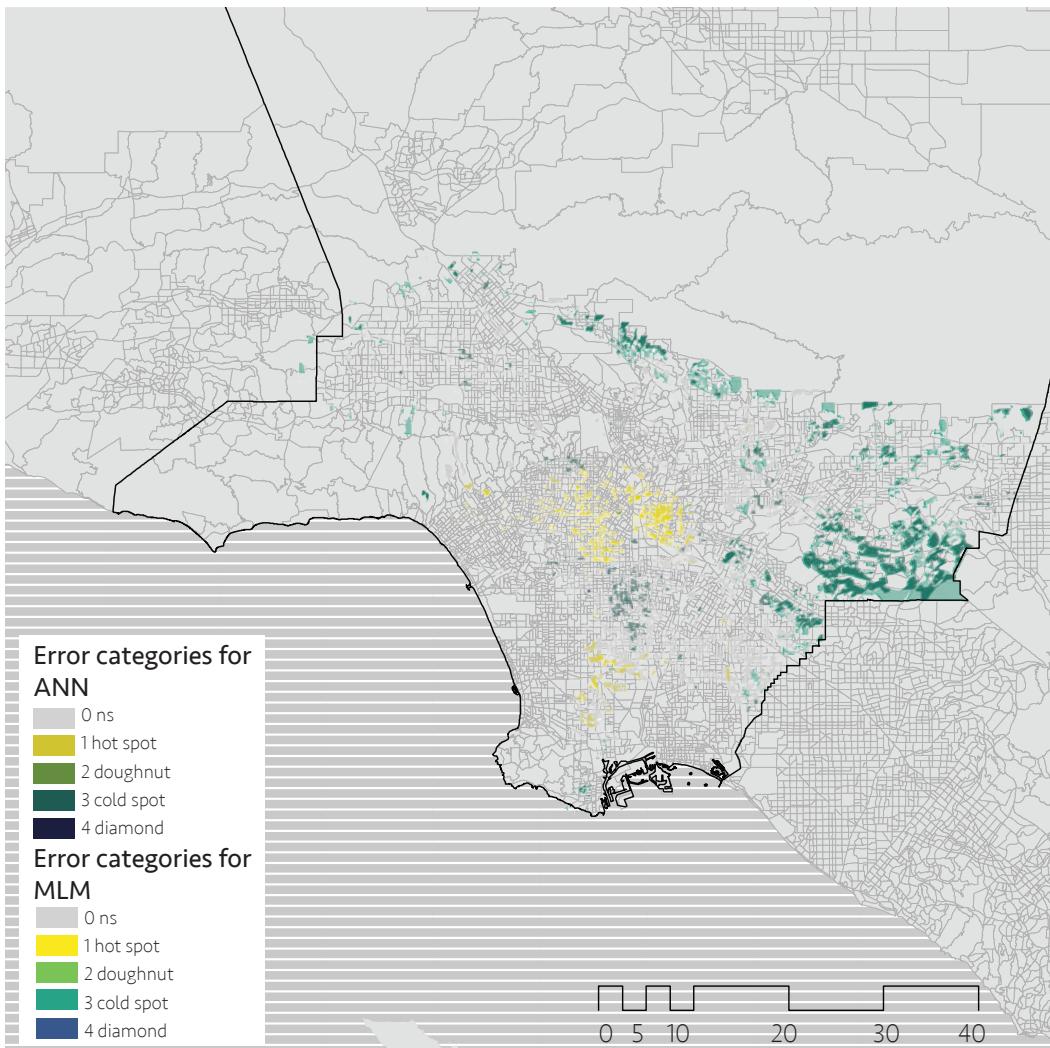


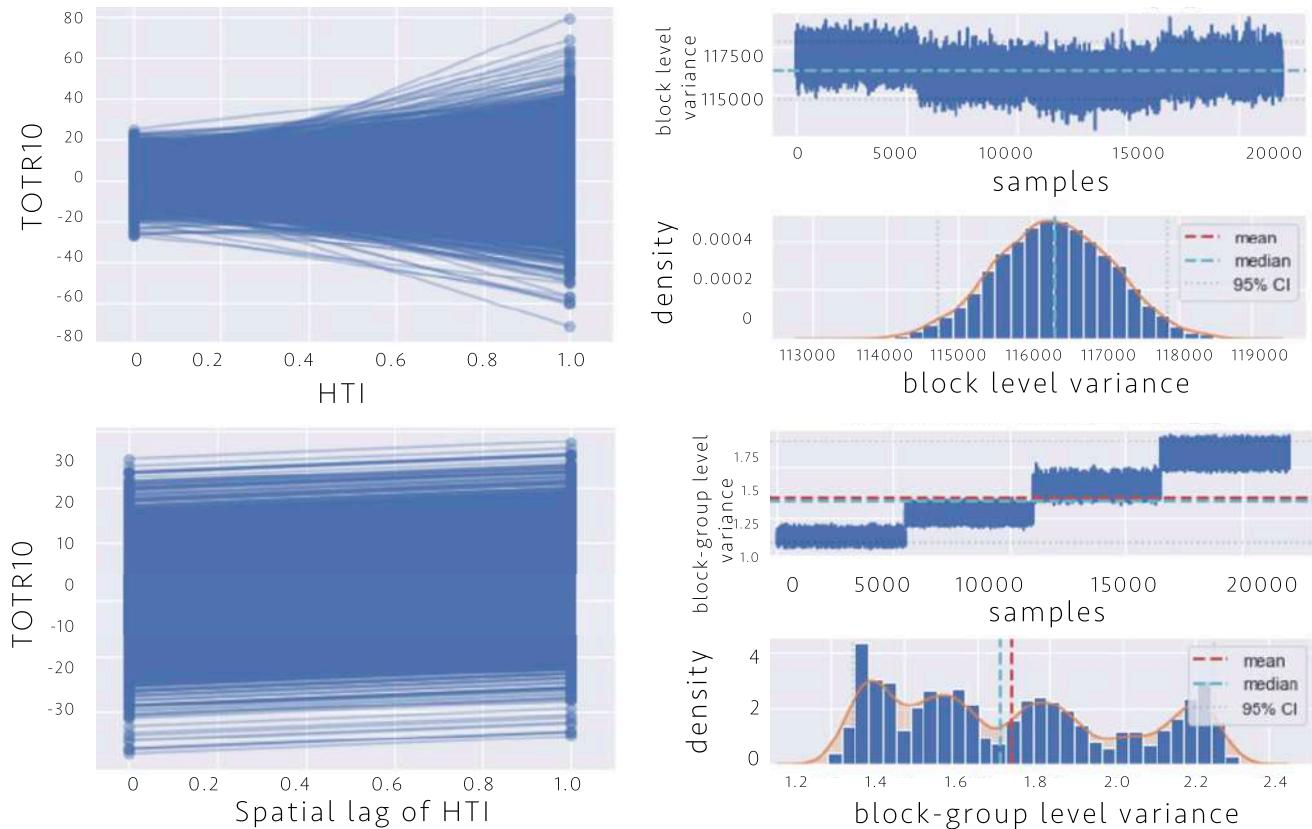
Figure 16 The error from Model 1 for the MLM and ANN.

4.5 Is there evidence of spatial spillovers that explain differences in accessibility to employment opportunities?

Model 2 incorporates the spatial lag of household and transportation costs to account for the spatial spillover effects of household and transportation costs. Put simply, the higher the proportion of income a census block spends on household and transportation, the more likely this will affect neighbouring census blocks, through rent increases or transportation fare increases. Analysing the uncertainty from figure 17c shows that the block-level variation is close to the mode of the distribution, and comfortably within the 95 percent confidence

intervals. However, figure 17d shows the Markov chains at the block group level struggle to converge, suggesting HTI costs and the spatial lag of HTI are perfectly collinear at the block group level, and therefore are not separately identifiable within the model predictions.

Trace and Posterior Distributions - Model 2



■ **Figure 17** Regression plots showing the relationship between (a) household and transportation costs (HTI) and Accessibility and (b) the spatial lag of HTI and Accessibility. The trace and distribution plot for (c) the block level variance and (d) block group level variance.

The coefficient for HTI costs decreases to 1.2, suggesting for every 1 percent increase in household and transport costs, job accessibility increases by 1.2 jobs. Additionally, the coefficient associated with the inclusion of spatially lagged HTI values is 3.1, indicating that for every 1 percent increase in the lag of HTI, accessibility increases by 3.1 jobs. Both results from the multilevel model indicate a weak positive relationship between household and transport costs and accessibility. Visualising the results for the MLM shows that the inclusion of a spatial lag removes the spatial pattern of accessibility seen in Model 2 (Figure

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

18a) predicts negative values as low as -23000 jobs in certain census blocks, and block-level variance which increases to 116,300 from 8,890 (Table 4, Model 2). The neural network predicts the average number of jobs accessible by block is 103,423, however the mean absolute error (MAE) is 35,000 jobs, suggesting high variability in block-level predictions. This uncertainty is further reflected in the ANN's interquartile range, which increases from 31294 to 31584 (Table 6, Model 2). Nonetheless, the ANN retains the correct spatial distribution of accessibility concentrating high values (around 100,000 to 200,000 jobs) in Downtown Los Angeles (Figure 18b), and explains 35 percent of the variation in accessibility using only HTI costs and the spatial lag of HTI as predictors.

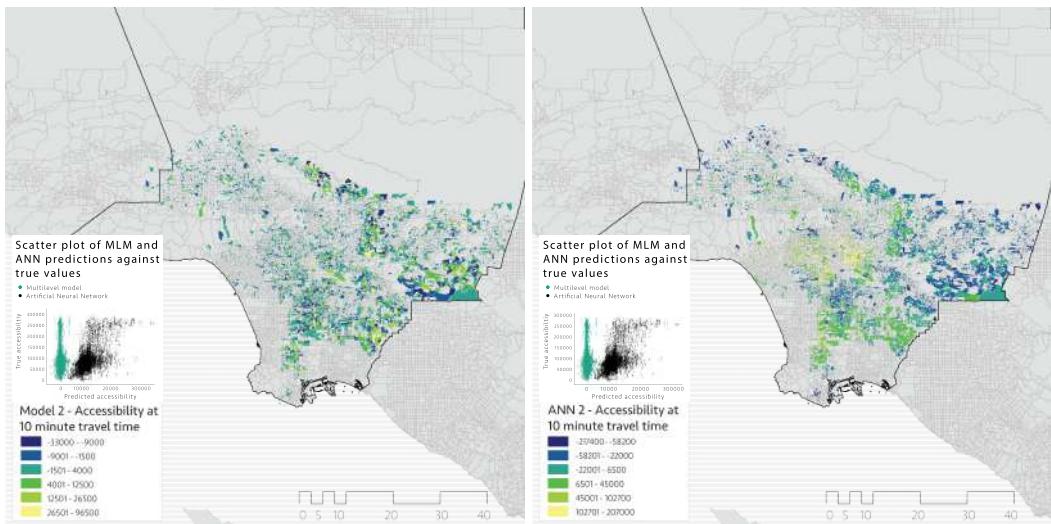
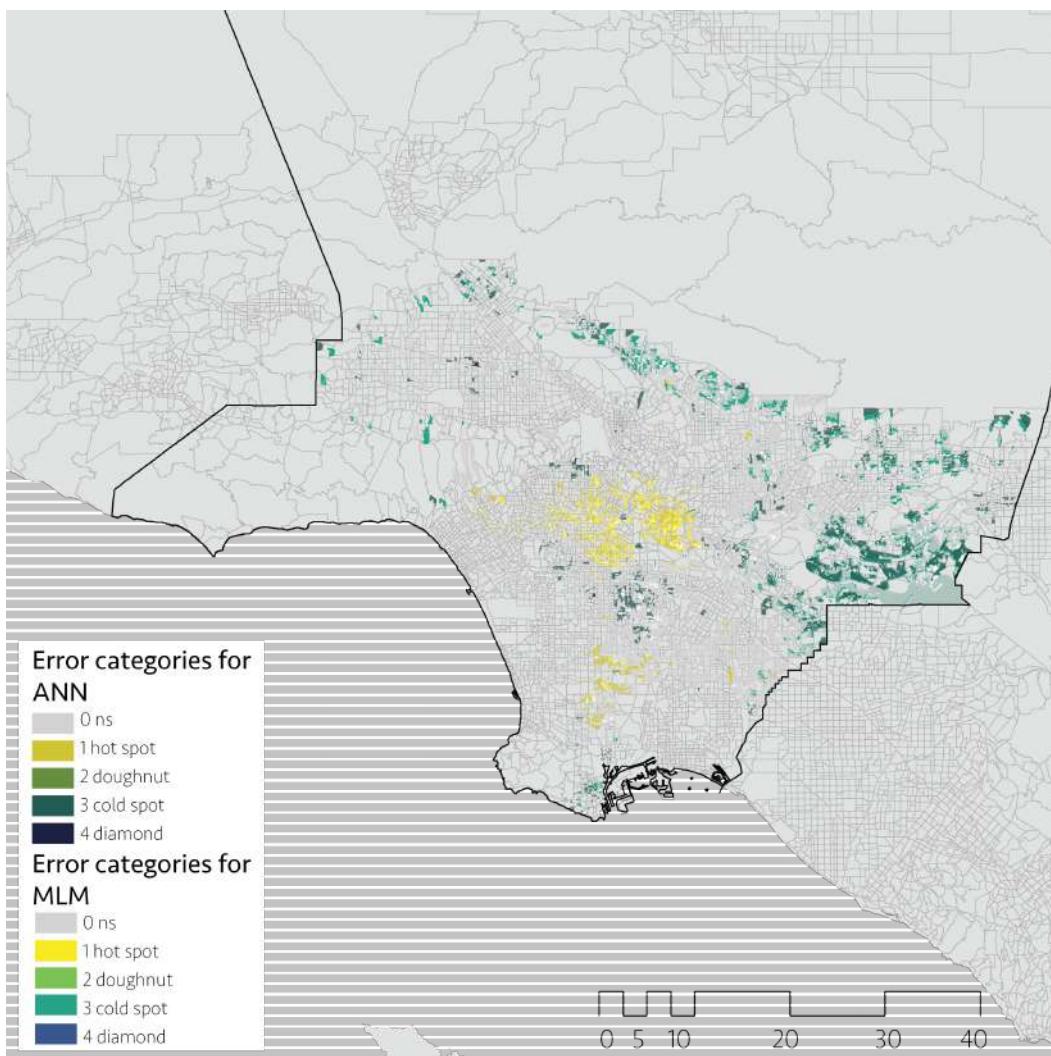


Figure 18 Model 2 predictions for the MLM (a) and ANN (b) and a scatterplot to show the residual distribution.

The corresponding errors in the MLM and ANN shows that both modelling techniques under predict levels of accessibility in eastern Los Angeles, due to the negative accessibility predictions, and both models over predict accessibility to employment opportunities in Downtown Los Angeles. Figure 19 illustrates this with a high concentration of 'hot spot' (yellow) blocks. Furthermore, the ANN under predicts accessibility in south central Los Angeles, shown by the dark green block errors. Overall, the error results are very similar to the predictions in Model 1 (where the only predictor introduced is household and transportation costs) suggesting the spatial lag of HTI has little influence on accessibility.



■ **Figure 19** The error from Model 2 for the MLM and ANN.

4.6 What is the relative effect of being African American or Hispanic in determining the level of accessibility to employment opportunities at the block level?

Model 3 incorporates the percentage of population that identify as African American or Hispanic, and is designed to determine the relative contribution of being in a high percentage Hispanic and African American census tract on accessibility. Therefore, Model 3 includes the percentage of individuals who identify as Hispanic or African American indicator variables for each tract, plus a tract level covariate. The posterior distributions (Figure 20d, e and f)

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

shows that both the block level and block group level variation is close to the modes of the distribution, however the tract level variation jumps during the 2nd Markov chain.

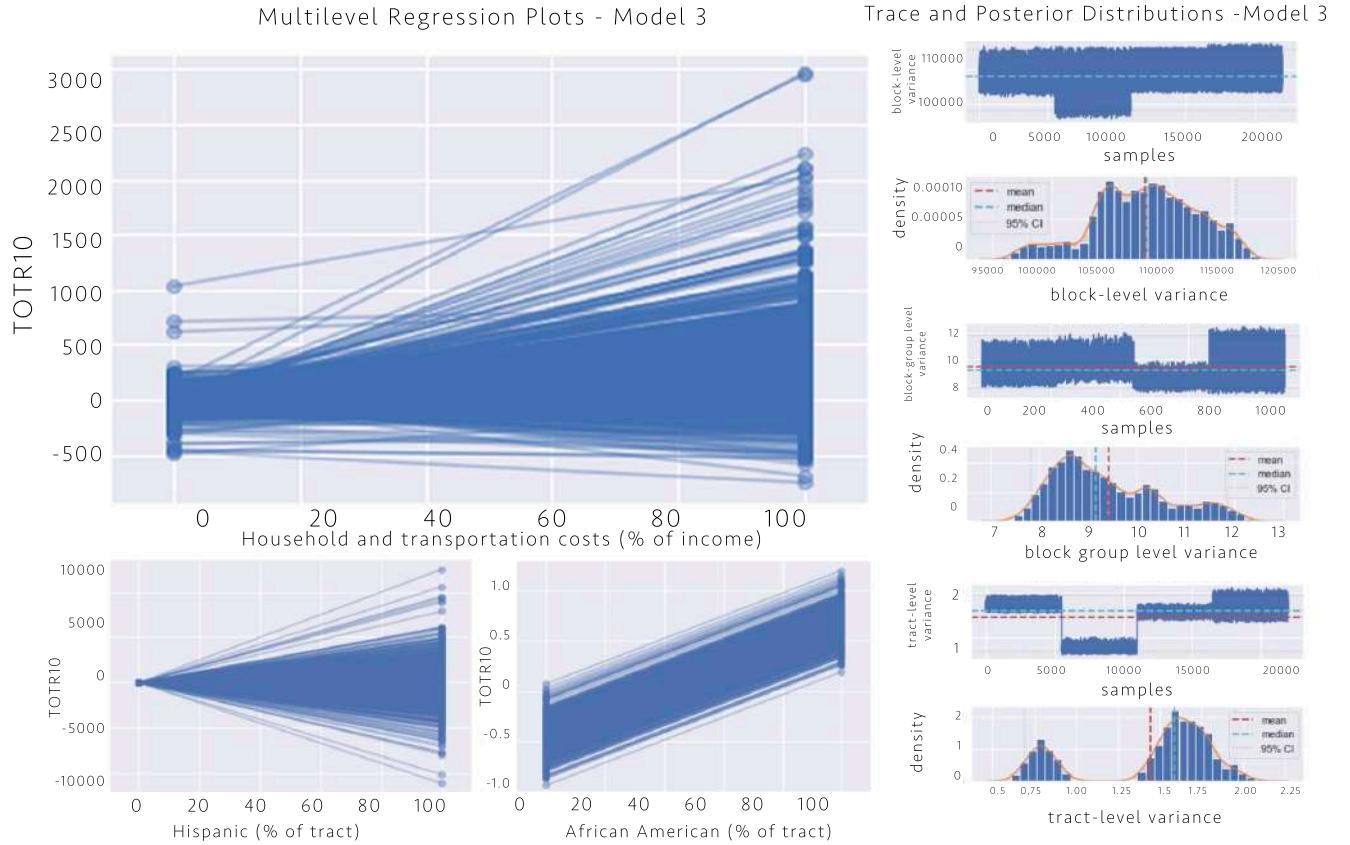


Figure 20 Figures (a, b, c) are the regression plots for (a) Household and transport costs (HTI) and Accessibility (TOTR10), (b) for Hispanic and Accessibility, and (c) African American and Accessibility. Figures (d, e, f) are the trace and distribution plot for (d) block level variance; (e) the block group variance, and (f) the tract level variance.

The average effect of household and transportation costs in the multilevel model is 190.2, suggesting that for every 1 percent increase in HTI costs, accessibility increases by 190 jobs. This is supported by the multilevel regression plot (Figure 20a) which shows that the HTI coefficient is positive for most block groups. The slope estimate for the Hispanic coefficient is approximately -2.65 percent (Figure 20b), suggesting that for every 1 percent increase in the percentage of Hispanics decreases accessibility by 2.65 jobs. Furthermore, the slope estimate for the African American coefficient is approximately 1.2 (Figure 20c), suggesting that for every 1 percent increase in the percentage of African Americans at the tract level, accessibility increases by 1.2 jobs. Although the coefficients for the effect

of African American and Hispanic communities are low, the standard deviations of the coefficients are 208 and 396 respectively, suggesting there are extreme ethnicity effects on local accessibility. Additionally, the interquartile range for Model 3 increases from 31,584 to 39,676 jobs (Table 4).

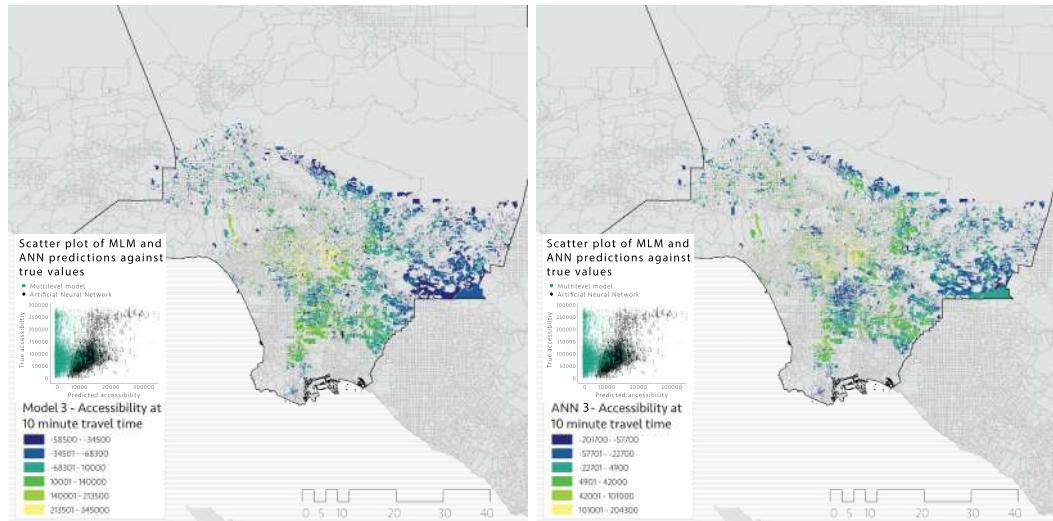


Figure 21 Model 3 predictions for the MLM (a) and ANN (b) and a scatterplot to show the residual distribution.

Including ethnicity as a tract level predictor in the multilevel model gives an indication of the racial differences in accessibility in Los Angeles County (Figure 21). There is a large range in accessibility for both model predictions, reflecting the high block level uncertainty of 108,000 in the MLM (reduced from 117,000 in Model 2) (Table 4, Model 3). However, the variation explained at the block group level increases from 1.69 to 9.42, suggesting the inclusion of ethnicity helps explain neighbourhood relationships of accessibility. The artificial neural network displays a similar spatial distribution of predicted accessibility, concentrating high values of accessibility in downtown locations (Figure 21b). On average, the ANN predicts 104,000 jobs are accessible from each block, and the variance decreases slightly to 34,000 (Table 6, Model 3). Overall, the neural network explains 42 percent of the variation in accessibility.

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

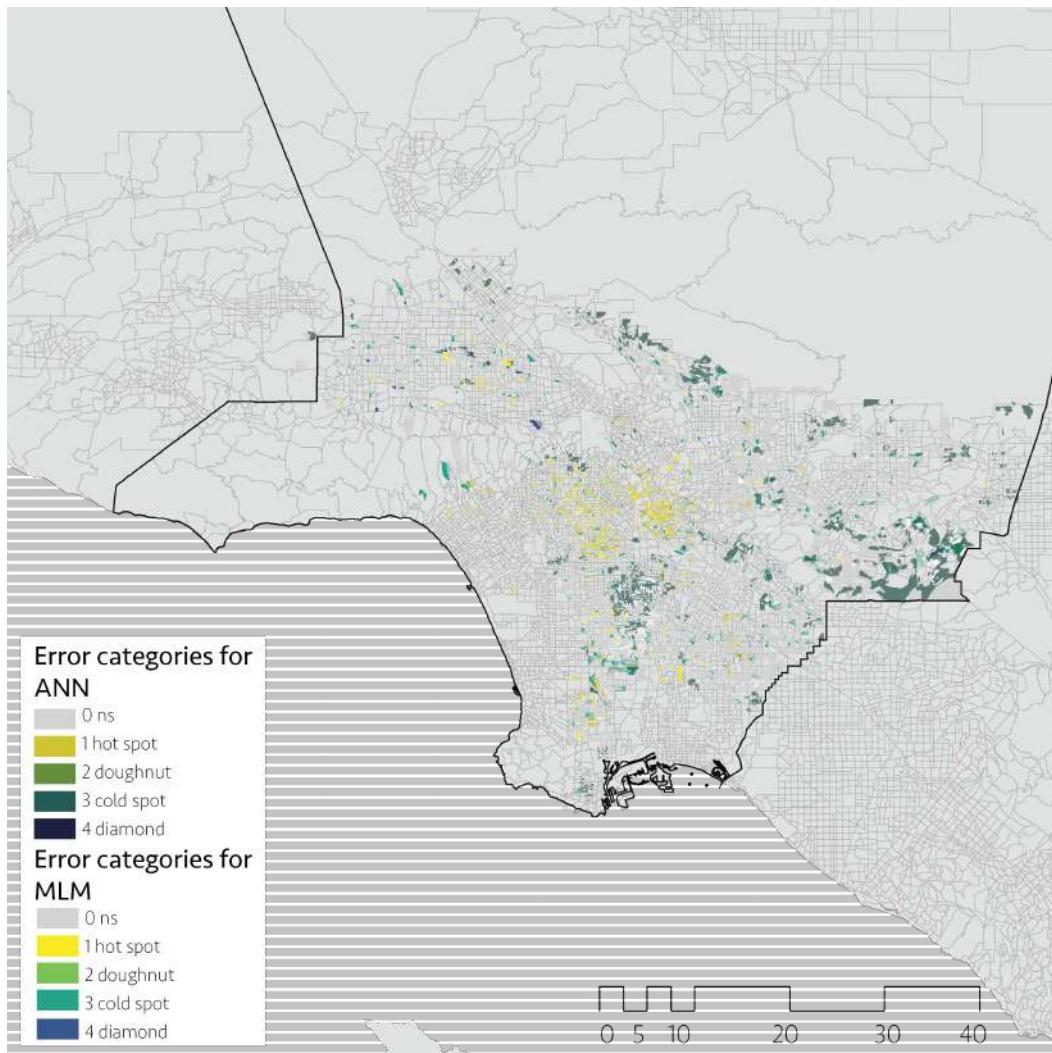


Figure 22 The error from Model 3 for the MLM and ANN.

Figure 22 shows a significant reduction in the multilevel model and artificial neural network errors in comparison to Figure 19, as there is a reduction in the number of blocks where prediction is significantly lower or higher than 'true' accessibility (represented as yellow hot spots and green cold spots respectively). In particular, the accuracy of MLM predictions significantly improves in eastern and northern Los Angeles County, suggesting the inclusion of ethnicity predicts low access in suburban locations well using a multilevel model. Contrary to this, the artificial neural network predictions still overpredicts accessibility in south central and eastern Los Angeles.

5 Discussion

This study outlines the importance of equity in determining accessibility to jobs in Los Angeles County. Van Wee and Geurs (2011) suggest that access to jobs can be unequal, however only the systematic exclusion of a certain population would make the distribution of accessibility inequitable. As mentioned, high household and transport costs can limit an individual's travel radius and thus their accessibility. Model 1 confirms that the percentage of income spent on household and transportation costs does predict the local accessibility, suggesting census blocks experiencing high household and transport costs cannot afford to live proximate to jobs, or cannot afford to travel long distances to job centres in LA County.

The results from Model 2 use Vega and Elhorst's (2015) SLX model to separate and measure internal and external neighbourhood effects of being in a low equity neighbourhood cluster (measured as the spatial lag of HTI costs). When incorporating the spatial weights matrix the artificial neural network absorbs any additional variance, the predictions from ANN remain robust and the estimated effects are similar to Model 1 (Table 6, Model 2). However, the multilevel model loses the ability to register any spatial patterning of accessibility, a likely symptom of perfectly collinear elements that are not separately identifiable within the model predictions. Due to the multilevel model's limited ability to incorporate a spatial weights matrix, the results from the spatially lagged models are difficult to compare.

Combining this knowledge with Model 3, I test whether high HTI costs exclude certain ethnicities from accessing employment opportunities. Incorporating tract level data on Hispanic and African Americans alongside household and transportation costs into the neural network best predicts accessibility in this study. The results display large disparities in accessibility for Hispanic population groups and do suggest Hispanics experience an 'ethnic penalty' in accessibility. Furthermore, low levels of car ownership and higher rates of suburbanisation are likely to be the causal mechanisms of low accessibility for Hispanic communities. Although Model 3 shows that Hispanic populations experience lower access to jobs, the results are not conclusive, as both African American and Hispanic populations live proximate to the downtown area (Figure 9). This is unexpected, as African Americans and Hispanics tend to experience higher poverty rates and a smaller travel radius, thus

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

limiting accessibility (FWHA 2014). Yet diversity-rich metropolitan areas like Los Angeles have centralised ethnic populations, improving their level of accessibility. However, these results, and their lack of significance could evidence selection bias, as the US Census may be unable to account for minority populations in informal housing situations. Furthermore, recent work by (Fowler *et al.* 2019) suggests block level ethnicity data can contain a high degree of uncertainty, and instead a measure of "individual context" (2019: 5) is required to avoid the contextual fallacy of aggregating data to spatial units. Therefore using both multilevel and machine learning techniques on individual level data could further investigate this relationship.

This study suggests that blocks that experience high levels of inequity are less likely to experience good access to local employment opportunities by private transportation. Overall, the strength of the multilevel model lies in its ability to draw correct inferences from regression coefficients at different hierarchical levels. In comparison to neural networks, multilevel modelling allows us to evaluate group effects by nesting blocks within block groups and tracts, which can increase certainty in significant results. In contrast, the artificial neural network cannot account for structural hierarchies in the data, therefore the estimates of accessibility do not appreciably change when coefficients of higher-level ethnicity predictor variables are introduced. However, due to high block level variability, the nested structure of the multilevel model fails to identify strong effects corresponding to the block group and tract levels. In this sense, the artificial neural network predicts accessibility better than the multilevel model. Chapter 6 gives the limitations of the study and suggests implications for future research.

6 Limitations and implications for future research

Firstly, to evaluate the success of each multilevel model it is essential the Markov Chain Monte Carlo estimators converge on the ‘true expectation’ of accessibility. This study was limited by the multilevel model, as a result of perfectly collinear elements in the model parameters the multilevel model would not converge in more complex model specifications. This limited the number of variables used to predict accessibility, as increasing the number of parameters made it difficult for the model to converge and removed the expected spatial pattern of accessibility. Potential reasons for the lack of convergence are that the burn-in period of 1000 samples was ineffective, and the model may require a longer spin-up time. Furthermore, using optimisation could be an alternative method of finding the high-probability region when evaluating a model with Hamiltonian Monte Carlo.

Secondly, due to the inability to run models with more predictor variables, the model fails to control for the effect of income. Incorporating income into the model would test the research question: To what extent does the level of income matter in determining accessibility to employment opportunities? Answering this would identify whether individuals living in census blocks with high household and transportation costs experience low access to jobs because they are located far from Downtown Los Angeles, or because low incomes minimise their travel radius.

Thirdly, although this study establishes a relationship between equity and accessibility by the transportation network, it does not address the inequalities of the labor market that prevent certain ethnic groups from obtaining employment. In the multilevel analysis, all models display block level variation that substantially exceeds any variation explained at the block group or the tract level, suggesting within-block differences explain most of the disparity in accessibility. The high block-level variation in accessibility indicates the importance of individual differences in determining level of access, such as car ownership and household income. This research confirms that neighbourhoods cannot fully explain the differences in accessibility, especially after controlling for neighbourhood demographic characteristics such as the percentage of ethnic minorities in a census tract.

Lastly, this study can be extended in a number of ways, such as using individual activity-

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

travel data to explore the relationship between place based characteristics and daily activity using an artificial neural network. Individual level data, such as the ethno-racial diversity metric (Fowler *et al.* 2019) could measure relationships between demographic characteristics such as age, gender, disability, and ethnicity on local accessibility. For example, the research could test whether female-led single-parent families, or adults with a disability are more likely to experience low accessibility (FHWA 2014). Alternatively, multilevel analysis could be used to explore the differential effects of household characteristics on accessibility in relation to neighbourhoods, using cross-level interactions between household and neighbourhood level characteristics. Such developments are likely to lead to both enhanced predictive accuracy and explanatory capacity, as this approach would be capable of exploring whether block level variation of accessibility is due to the household or the individual.

7 Conclusion

This study investigated the relationship between level of inequity and accessibility to employment opportunities for Hispanic and African Americans in Los Angeles County. The study structured a response with three research questions. The conclusions will address each research question in turn:

- To what extent do household and transportation costs as a percentage of total income matter in determining accessibility to employment opportunities?

The multilevel model and the neural network revealed a strong relationship between household and transportation costs and the level of accessibility to jobs. Overall, blocks that are more centrally located are more likely to spend a lower percentage of income on household and transportation costs, and live proximate to jobs. Operationally, the single predictor variable explains more of the variation in local accessibility in the multilevel model than the neural network. Although neighbourhood level covariance is low (2.66 at the block group level and 1.2 at the tract level), the multilevel model predicts accessibility more accurately, however it cannot infer neighbourhood level effects on accessibility.

- Is there evidence of spatial spillovers that explain differences in accessibility to employment opportunities?

Christina Last

When incorporating ‘spatial spillover effects’ of household and transportation costs, the multilevel model loses its ability to predict accessibility accurately, whereas the predictions from ANN remain robust. However, incorporating the spatial lag of HTI costs into the ANN fails to further explain variation in accessibility (ANN $R^2 = 0.35$) suggesting there are limited spillover effects of neighbouring census blocks experiencing a similar level of household and transportation costs.

- What is the relative effect of being a Hispanic or African American in determining the level of accessibility to employment opportunities?

The results suggest more suburban Hispanic neighbourhoods experience an ‘ethnic penalty’ in accessibility. However, by modelling the accessibility of African Americans and Hispanics, neither model was capable of using neighbourhood level ethnicity data to draw conclusions on level of accessibility. Despite the insignificance of ethnicity parameters, the ANN explains the largest amount of variation in accessibility ($R^2 = 0.42$) and shows the largest reduction in areas of overprediction and underprediction of accessibility compared to the ‘true value’.

As well as answering the research hypotheses, this study compared the residuals from a multilevel model and an artificial neural network to actual accessibility, drawing comparisons between their goodness of fit, their predictive accuracy and their explanatory capacity. Although the MLM is capable of partitioning variance at the neighbourhood level, the amount of variance in accessibility explained by ANN is larger and more consistent than the results from the MLM, and the results indicate ANN offers better predictive accuracy and higher explanatory capacity. One of the defining strengths of multilevel analysis is its capability of measuring neighbourhood effects, however, because of the variation in accessibility is observed at the block level, multilevel analysis cannot infer neighbourhood level relationships. Furthermore, this indicates that individual or household level determinants may have a more significant impact on accessibility in Los Angeles County.

This is the first study that demonstrates that artificial neural networks can be applied to a large dataset to measure accessibility, and shows that ANN is better at measuring access to employment than MLM when tract-level ethnicity variables are incorporated into the model

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

specification. Although a multilevel approach claims to be more capable of inferring the relationships that between structural census blocks, block groups and tracts, the artificial neural network achieved more robust predictions, and offers a promising entry point to further research on mechanisms that drive accessibility in Los Angeles County.

8 Technical Appendix

The objective of this study was to compare the ability of a multilevel model (MLM) and artificial neural network (ANN) to predict local accessibility (total accessibility by private transport in 10 minutes of travel time) in Los Angeles County. I developed a three-level multilevel model and an artificial neural network with two dense layers and compared their ability to model local accessibility using goodness of fit, predictive accuracy and explanatory capacity. Both approaches answer the research questions by sequentially controlling for housing and transportation costs (HTI), the spatial lag of HTI costs, ethnicity (African American and Hispanic, by introducing block group and tract level level characteristics.

Listings

1	Null Model example	61
2	Model 1 example	62
3	Model 2 example	63
4	Model 3 example	65
5	Artificial Neural Network (all models) example	67

8.1 Data Preparation

The original dataset contains 29,438 rows, of which 7,149 contain missing values. To handle missing data, I use an imputation strategy that predicts a missing value based on observation values for alternative predictors that correspond to the same census block. To complete a row-wise imputation I ran the following analysis:

- I split the data into a ‘no missing dataset’ containing the unique rows that have values for all predictor variables, and a ‘full dataset’ all rows including prediction variables with missing data.
- I then used the ‘no missing dataset’ to run a regression analysis on each block to identify the relationship between the predictor variables and the outcome variables (TOTR10).
- After generating the predictions the for the ‘no missing dataset’ I selected all the outcome

- variables (TOTR10) for the ‘full dataset’.
- I use the outcome variable (TOTR10) to predict the missing predictor variables using the relationships identified in stage 2. I assign these imputed values to the original dataframe.

8.2 Spatial Autocorrelation

Spatial autocorrelation is a method to identify spatial patterning of the values of sample points across space (Getis and Ord 1995). Correcting for autocorrelation improves Hamiltonian Monte Carlo estimation as this reduced the gap between conditional variances (variance of a random variable given the parameter values) and marginal variances (the expected value of the conditional variance) (Betancourt and Girolami 2015). Identifying spatial autocorrelation in predictor variables (in this case, household and transportation costs) helps to diagnose issues with model fitting. In this section, I describe both global and local measures of autocorrelation in local accessibility (TOTR10).

8.2.1 Global Binary Spatial Autocorrelation

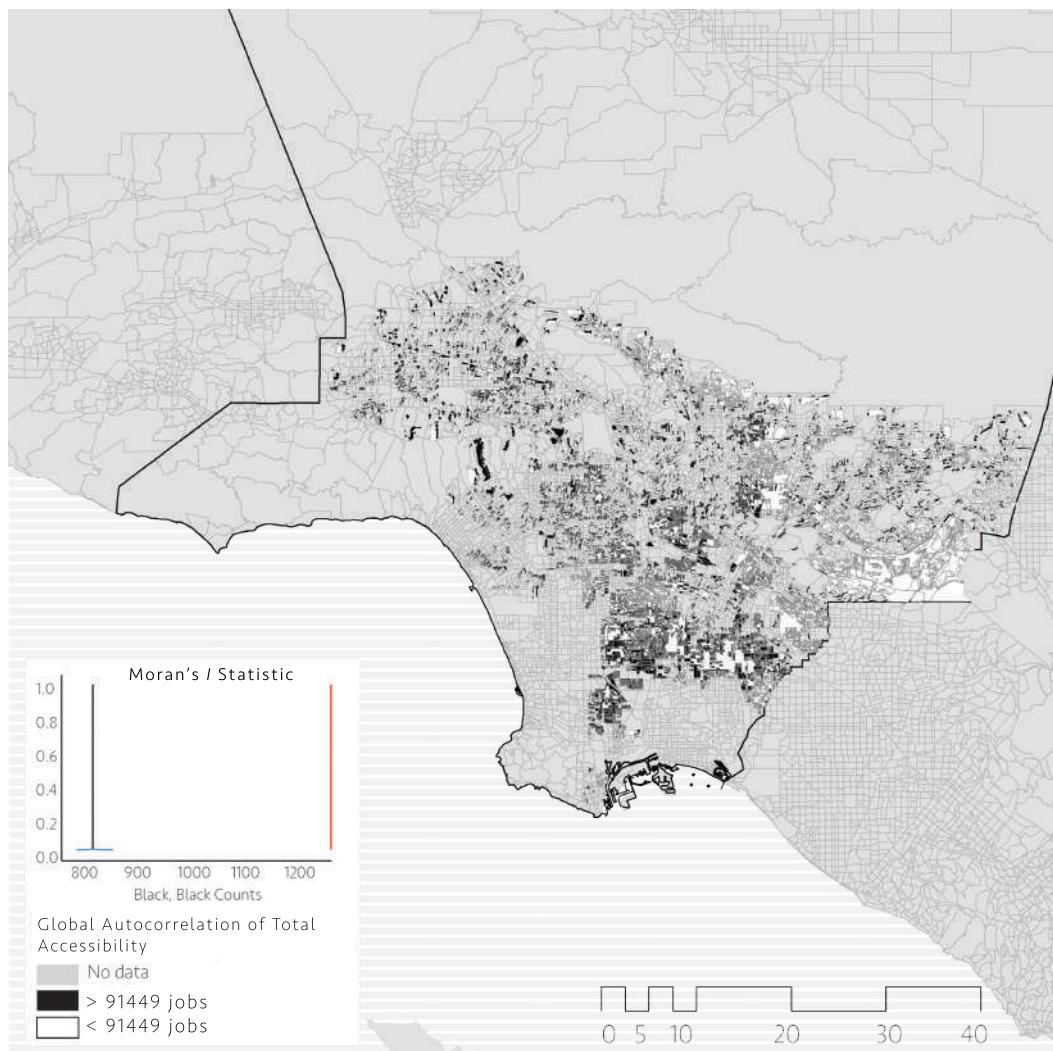
A way to formalise a binary test for spatial autocorrelation can be achieved through a binary join. This join exists for each neighbor block group pair of observations, reflected in a binary spatial weights object. Each unit takes one of two values, either higher or lower than the median percentage of income spent on household and transportation expenditure (HTI), therefore each given pair of neighbouring block groups three types of join can arise:

- High High (HH)
- Low Low (LL)
- High Low (or Low High) (HL)

Figure 23 identifies 8,235 High accessibility census blocks (values above the median block accessibility of 91,449 jobs) and 12,398 High, High accessibility value census block group pairs. Given the number of high value polygons in figure 23, the number of high polygons I expect to generate, if values were randomly assigned to census block group locations in Los Angeles, should not be significantly different from 12,398 if we are to assume no spatial

Christina Last

autocorrelation. Indeed, PySAL generates random spatial permutations of the observed accessibility values to generate a spatial autocorrelation join, formed under the null of complete spatial randomness (CSR). The average number of High High census block pairs results from the synthetic model prediction under the null hypothesis is 905, which is significantly less than my observed join count of 12,398. Therefore, we can reject the null of CSR at the binary level.

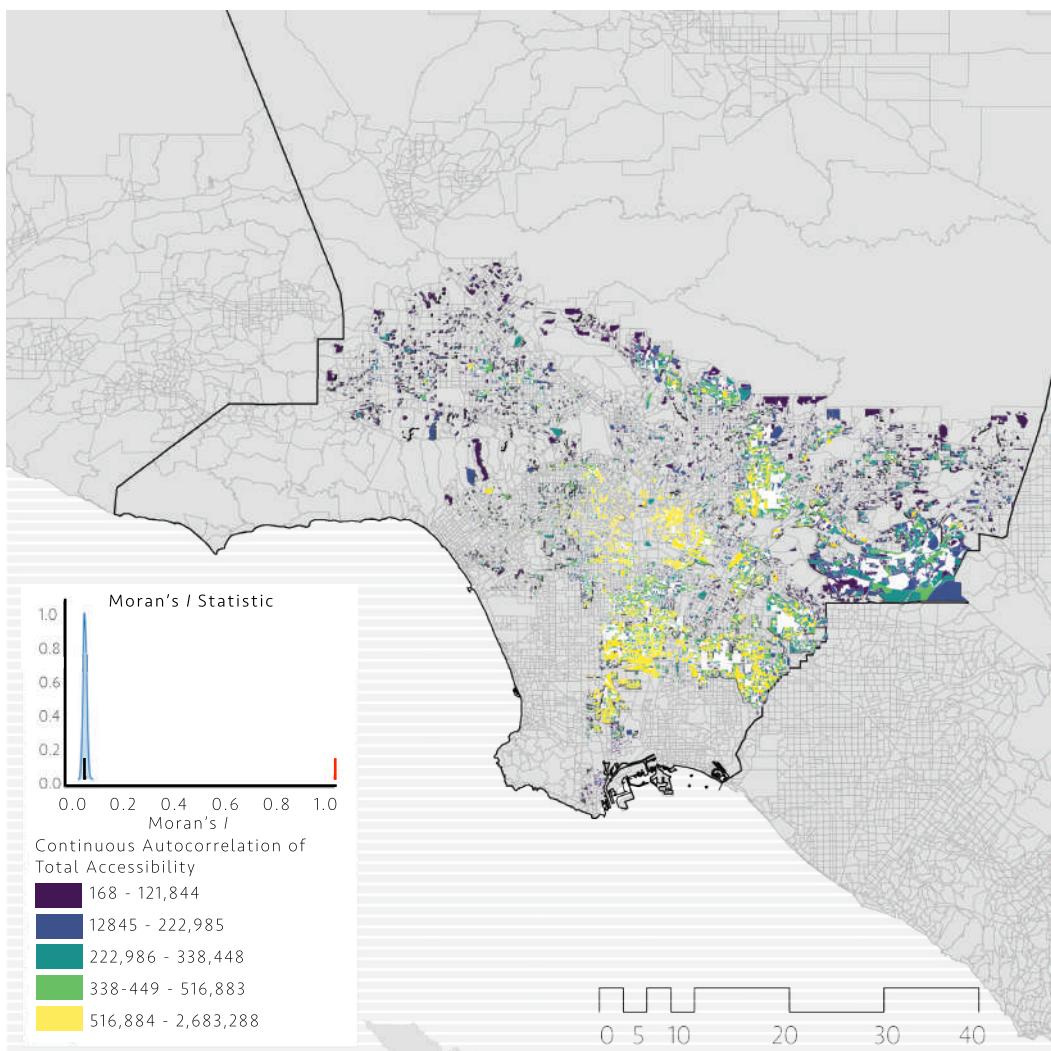


■ **Figure 23** The binary spatial autocorrelation of total accessibility.

8.2.2 Local Continuous Spatial Autocorrelation

Although the binary join accounts for the presence and absence of high accessibility (defined as accessibility scores above the median value), the creation of the artificial binary variable removes a large amount of information in our originally continuous attribute. Furthermore, when the data is structured as successive layers (as with multilevel modelling), correlations are locally determined. To describe the correlations between successive layers I compute a local autocorrelation statistic.

A popular approach to investigate local spatial autocorrelation is to compute the Moran's I statistic, and test the null hypothesis of no spatial autocorrelation or whether there is an observed departure in the residuals from their expected spatial distribution. Furthermore, by calculating Moran's I is a test for local autocorrelation for local accessibility (TOTR10), I am able to explore continuous spatial autocorrelation for the accessibility variable for each focal location (figure 24).



■ **Figure 24** A quintile plot of the local spatial autocorrelation of total accessibility.

I distinguish the specific type of local spatial association reflected in the four quadrants of the Moran Scatterplot and attribute these values to geographic locations in Los Angeles County (figure 25). These are classified as four types of accessibility (hot spot, cold spot, doughnut and no significance). Unsurprisingly, hot spots with significantly higher household and transport costs are located in Downtown Los Angeles County, and cold spots are those blocks located in suburban locations of Los Angeles County. The doughnut and diamond classification represents blocks with significantly higher accessibility than their neighbouring blocks.

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

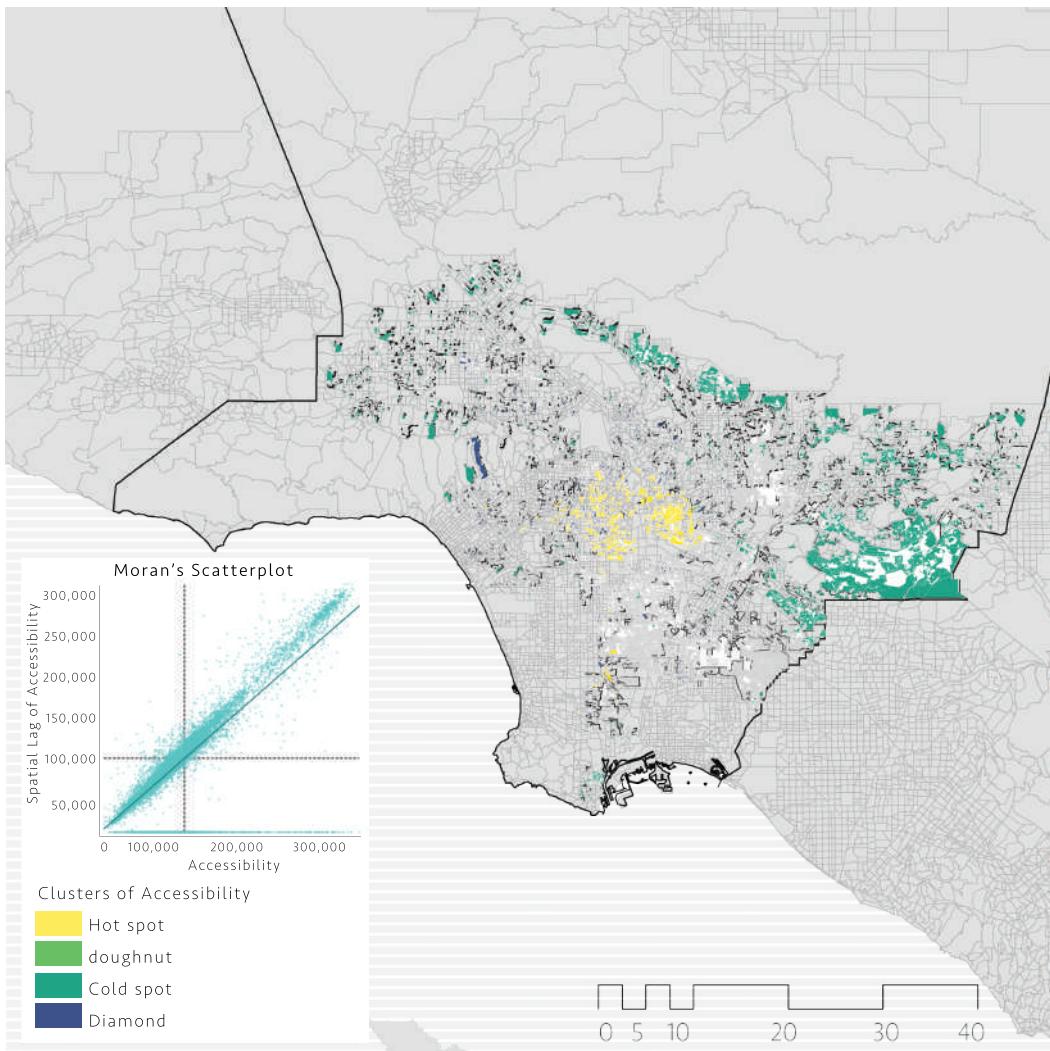


Figure 25 Specific types of local spatial association reflected in the four quadrants of the Moran Scatterplot of Accessibility.

8.3 Model Specifications

For each model, I specify 6000 iterations of sampling from each Markov chain in PyStan, and I use a warm up period of 1000 samples, which is the amount of samples that will be discarded from the beginning of each chain. This ‘burn in’ period is defined because the early samples will be drawn when the Markov chain will not have had the chance to reach equilibrium. Because the Markov sampling is stochastic, it is advantageous to have more chains, therefore I specify 4 chains from which the samples will be combined to form the posterior distribution. With 4 chains and 6000 samples, I retain 20,000 samples after

the burn in period. I evaluate the model using Hamiltonian Markov Chain Monte Carlo, a probabilistic model that uses a prior distribution to express dependence between probabilistic events, allows the model to fit even with high-dimensional, correlated data (Betancourt and Girolam 2013).

8.3.1 Specification of the Null Model

Listing 1 Null Model example

```
Null_Model_MLM = """  
  
data {  
    int<lower=0> N;  
    int<lower=0> P_N;  
    int<lower=0> J;  
    int<lower=0> P_J;  
    int<lower=0> Q;  
    int<lower=0> P_Q;  
    int<lower=1, upper=J> block_to_bg[N];  
    int<lower=1, upper=Q> bg_to_tract[J];  
    vector[N] y;  
}  
  
parameters {  
    vector[J] bg_intercept;  
    vector[Q] tract_intercept;  
    real hypermean;  
    real<lower=0> sigma_block;  
    real<lower=0> sigma_bg;  
    real<lower=0> sigma_tract;  
}  
  
transformed parameters {  
    vector[N] y_hat;  
    for (i in 1:N)  
        y_hat[i] = bg_intercept[block_to_bg[i]];
```

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

```
}

model {
    for (j in 1:J)
        bg_intercept[j] ~ normal(tract_intercept[bg_to_tract[j]] ,
        sigma_bg);
    y ~ normal(y_hat, sigma_block);
}
"""

```

8.3.2 Specification of Model 1

Building on the varying slope and intercept model, In Model 1 I introduce the percentage of income spent on household and transportation costs (HTI) for each block on the mean private accessibility at the 10 minute travel time, without introducing group-level predictors or any contextual effects.

■ Listing 2 Model 1 example

```
Model_1_MLM = """

data {
    int<lower=0> N;
    int<lower=0> P_N;
    int<lower=0> J;
    int<lower=0> P_J;
    int<lower=0> Q;
    int<lower=0> P_Q;
    int<lower=1, upper=J> block_to_bg[N];
    int<lower=1, upper=Q> bg_to_tract[J];
    vector[J] ht_ami;
    vector[N] y;
}
parameters {
    vector[J] bg_intercept;
    vector[Q] tract_intercept;
```

```

real hypermean;

vector[J] ht_ami_slope;

real<lower=0> sigma_block;
real<lower=0> sigma_bg;
real<lower=0> sigma_tract;

}

transformed parameters {
    vector[N] y_hat;
    for (i in 1:N)
        y_hat[i] = bg_intercept[block_to_bg[i]]
            + ht_ami[block_to_bg[i]] *
            ht_ami_slope[block_to_bg[i]];
}

model {
    for (j in 1:J)
        bg_intercept[j] ~ normal(tract_intercept[bg_to_tract[j]],
        sigma_bg);
    y ~ normal(y_hat, sigma_block);
}
"""

```

8.3.3 Specification of Model 2

In Model 2, I incorporate spatial lags of HTI to measure the “spatial spillover effects” of experiencing a similar level of equity burden as near neighbourhoods. More specifically, the more a block is spending on household and transport costs, the more likely the neighbouring blocks will have to increase the percentage of income spent on household and transportation. To incorporate the “spatial spillover effects” of HTI costs, I compute a spatial weights matrix with queen contiguity to formalize this notion of spatial similarity.

■ **Listing 3** Model 2 example

```
Model_2_MLM = """
```

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

```
data {
    int<lower=0> N;
    int<lower=0> P_N;
    int<lower=0> J;
    int<lower=0> P_J;
    int<lower=0> Q;
    int<lower=0> P_Q;
    int<lower=1, upper=J> block_to_bg[N];
    int<lower=1, upper=Q> bg_to_tract[J];
    vector[J] ht_ami;
    vector[J] ht_ami_lag;
    vector[N] y;
}

parameters {
    vector[J] bg_intercept;
    vector[Q] tract_intercept;
    real hypermean;
    vector[J] ht_ami_slope;
    vector[J] ht_ami_slope_lag;
    real<lower=0> sigma_block;
    real<lower=0> sigma_bg;
    real<lower=0> sigma_tract;
}

transformed parameters {
    vector[N] y_hat;
    for (i in 1:N)
        y_hat[i] = bg_intercept[block_to_bg[i]]
                    + ht_ami[block_to_bg[i]] * ht_ami_slope[block_to_bg[i]]
                    + ht_ami_lag[block_to_bg[i]] *
                    ht_ami_slope_lag[block_to_bg[i]];
}

model {
    for (j in 1:J)
```

```
bg_intercept[j] ~ normal(tract_intercept[bg_to_tract[j]] ,  
    sigma_bg);  
y ~ normal(y_hat , sigma_block);  
}  
"""
```

8.3.4 Specification for Model 3

In Model 3, I introduce tract level effects (the percentage of the tract of Hispanic or African American ethnicity) described by their mean, the standard deviation and errors.

■ **Listing 4** Model 3 example

```
Model_3_MLM = """  
  
data {  
    int<lower=0> N;  
    int<lower=0> P_N;  
    int<lower=0> J;  
    int<lower=0> P_J;  
    int<lower=0> Q;  
    int<lower=0> P_Q;  
    int<lower=1, upper=J> block_to_bg[N];  
    int<lower=1, upper=Q> bg_to_tract[J];  
    vector[J] ht_ami;  
    vector[Q] afam;  
    vector[Q] hispan;  
    vector[N] y;  
}  
  
parameters {  
    vector[J] bg_intercept;  
    vector[Q] tract_intercept;  
    real hypermean;  
    vector[J] ht_ami_slope;  
    vector[Q] afam_slope;
```

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

```
vector[Q] hispan_slope;
real<lower=0> sigma_block;
real<lower=0> sigma_bg;
real<lower=0> sigma_tract;
}

transformed parameters {
vector[N] y_hat;
for (i in 1:N)
y_hat[i] = bg_intercept[block_to_bg[i]]
+ ht_ami[block_to_bg[i]] *
ht_ami_slope[block_to_bg[i]];

}

model {
for (q in 1:Q)
tract_intercept[q] ~ normal(hypermean, sigma_tract);
for (j in 1:J)
bg_intercept[j] ~ normal(tract_intercept[bg_to_tract[j]]
+ afam[bg_to_tract[j]] *
afam_slope[bg_to_tract[j]]
+ hispan[bg_to_tract[j]] *
hispan_slope[bg_to_tract[j]],
sigma_bg);
y ~ normal(y_hat, sigma_block);
}
"""

```

8.3.5 Specification of Neural Network

The artificial neural network predicts the level of 10 minute accessibility to employment (TOTR10) using a basic regression framework. I incorporate parameters such as the percentage of income spent on household and transportation costs (HTI), the spatial lag of

Christina Last

HTI, the percentage of census tract who are African American and Hispanic sequentially, to respond to the three research questions. To incorporate each predictor variable I select 10 data points as the training data, and use small network with few hidden layers to avoid overfitting. I use a back-propagation training algorithm to calculate the errors between the predicted accessibility and the target accessibility using the predictor variables, and this adjusts the connected weights between neurons in adjacent layers in the network. This is done using a sequential neural network with two densely connected hidden layers, and an output layer that returns a single, continuous value of accessibility (TOTR10). The model is trained for 1000 epochs, and the training and validation accuracy is recorded. I evaluate its success using the Mean Squared Error (MSE) (a common loss function used for regression problems), and the Mean Absolute Error (MAE).

■ Listing 5 Artificial Neural Network (all models) example

```
def build_model():

    model = keras.Sequential([
        layers.Dense(64, activation=tf.nn.relu, input_shape=[len(train_dataset.keys())]),
        layers.Dense(64, activation=tf.nn.relu),
        layers.Dense(1)
    ])

    optimizer = tf.keras.optimizers.RMSprop(0.001)

    model.compile(loss='mse',
                  optimizer=optimizer,
                  metrics=['mae', 'mse'])

    return model
```

8.4 Results

8.4.1 Multilevel Modelling

Table 4 Results from all Multilevel models

<i>Variable</i>	μ	<i>s.e.</i>	σ	2.5%	25%	50%	75%	97.5%	N_e	\hat{R}
Null Model	μ	24.77	20.12	29.11	-23.31	-7.70	38.10	51.54	56.0	2.09
	σ_{block}	117691	30.74	630.85	11647	117263	117674	118103	118942	421.32
	$\mu_{bgintercept}$	0.16	8.60	12.23	-16.70	-10.25	-0.01	10.60	17.32	2.04
	$\mu_{tractintercept}$	0.15	9.32	13.24	-17.90	-11.16	-0.02	11.53	18.51	1.03
	\hat{y}	0.15	8.54	12.15	-16.52	-10.21	-0.01	10.53	17.23	2.04
Model 1	μ	4.22	2.845	4.10	-1.97	1.08	4.86	8.16	9.46	2.07
	σ_{block}	8873	1.08	53.75	8873	8909	8977	2473	1.0	
	σ_{bg}	2.66	0.95	1.43	0.81	1.38	2.38	3.48	5.66	2.27
	σ_{tract}	1.86	0.45	1.0	0.66	1.04	1.66	2.48	4.03	5.00
	μ_{HTI}	2297	1.71	96.77	2106	2231	2296	2361	2486	3221
	$\mu_{bgintercept}$	4.21	2.95	5.51	-5.66	0.42	3.22	8.24	15.60	3.49
	$\mu_{tractintercept}$	4.22	2.90	4.61	-3.77	0.50	3.48	8.23	12.16	2.52
Model 2	\hat{y}	102861	64.09	3607	95788	100426	102862	105299	109936	3216
	μ	-16.57	21.38	30.49	-47.38	-39.95	-26.38	3.72	41.78	2.03
	σ_{block}	116299	302.65	799.80	114721	115746	116298	116895	117845	6.98
	σ_{bg}	1.70	0.22	0.32	1.26	1.50	1.67	1.97	2.25	2.10
	μ_{HTI}	1.16	17.38	26.39	-38.16	17.92	2.56	23.05	46.93	2.55
	μ_{bgHTI}	3.07	17.49	26.68	-38.16	-17.92	2.56	23.05	42.93	3221
	$\mu_{bgintercept}$	0.15	8.42	12.14	-17.85	-9.95	0.21	10.34	17.93	2.15
	$\mu_{tractintercept}$	0.14	9.10	13.06	-18.92	-10.77	0.15	11.15	19.04	2.12
Model 3	\hat{y}	1812	3452	5284	-6123	-2447	1596	5598	10875	2.58
	μ	-0.45	0.36	0.54	-1.68	-0.80	-0.20	-0.06	0.06	2.21
	σ_{block}	107762	1524	4348	98234	104843	107890	110928	115366	8.14
	σ_{bg}	9.42	0.38	1.13	7.84	8.55	9.15	10.20	11.88	8.68
	σ_{tract}	1.45	0.27	0.40	0.69	1.15	1.59	1.72	1.95	2.16
	μ_{HTI}	190	213	413	-391	-70.92	93.14	359.17	1144.85	5.19
	$\mu_{AfricanAmerican}$	1.12	108.54	208.61	-400.25	-104.64	1.0	109.87	396.50	4.94
	$\mu_{Hispanic}$	-2.65	196.79	389.92	-762.55	-189.81	-2.34	182.04	751.57	5.44
	$\mu_{bgintercept}$	1.22	45.41	84.47	-155.95	-48.94	0.39	50.57	162.77	4.97
	$\mu_{tractintercept}$	-0.45	0.26	1.58	-3.83	-1.39	-0.37	0.54	2.63	53.90
	\hat{y}	14888	12481	23631	-15972	-1365	8280	25634	69369	4.90



■ **Figure 26** The significance, visualisation and covariance of the multilevel model predictions.

8.4.2 Artificial Neural Network



■ **Figure 27** The significance, visualisation and error history of the Artificial Neural Network predictions.

 **Table 5** Results from all Artificial Neural Networks

<i>Model</i>	μ	<i>s.e.</i>	σ	2.5%	25%	50%	75%	97.5%
Model 1	102800	270	34619	58440	81681	95702	112975	198532
Model 2	103424	271	34733	56187	81501	96205	113085	199726
Model 3	104016	287	36793	54941	78597	97886	118273	200270

 **Table 6** Performance measures for the cross validation data and the training data using an Artificial Neural Network

<i>Model</i>	<i>ValidationLoss</i>	<i>ValidationMAE</i>	<i>ValidationMAE</i> ²	<i>Loss</i>	<i>MAE</i>	<i>MAE</i> ²
Model 1	2384244007	35691	2384244007	2298795973	34652	2298795973
Model 2	2283495639	35023	2283495639	2229693956	34229	2229693956
Model 3	2215789768	34080	2215789768	2145429819	33204	2145429819

References

- Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent.* Cambridge, Massachusetts: Harvard University Press.
- Betancourt M. and Girolami, M. (2015) Hamiltonian Monte Carlo for Hierarchical Models. *In:* K. U. Satyanshu, U. Singh, K. D. Dipak, A. Loganathan eds. *Current trends in Bayesian methodology with applications*, Florida, USA: CRC Press, 2015, 116-138.
- Bogue, D. J. (1949) *The structure of the metropolitan community: A study of dominance and subdominance*, Horace H. Rackham School of Graduate Studies: University of Michigan.
- Brueckner, J. K., and Rosenthal, S. S. (2009) Gentrification and neighborhood housing cycles: Will America's future downtowns be rich? *The Review of Economics and Statistics*, 91(4), 725-743.
- Bureau of Labor Statistics (2017) *Consumer Expenditures-2016*, Washington D.C.: U.S. Department of Labor. (Available Online: www.bls.gov). [Accessed 23rd February 2019]
- Carter, S. (2019) Commutes in this city are the longest in the US—and it's not NYC or San Francisco. *CNBC*, (Available Online: www.cnbc.com) [Accessed 7th February 2019].
- Center for Neighbourhood Technology (2017) *H + T Index Methods*. (Available Online: htaindex.cnt.org)
- Chen, Y., Ravulaparthy, S., Deutsch, K., Dalal, P., Yoon, S. Y., Lei, T. Goulias, K. G. Pendyala, R. M., Bhat, C. R., and Hu H-H. (2011) Development of indicators of opportunity-based accessibility. *Transportation Research Record*, 2255(1), 58-68.
- Chen, B. Y. and Kwan, M. P. (2018) Guest Editorial: Spatiotemporal Big Data Analytics for Transportation Applications, *Transportmetrica A: Transport Science*, 1-4.
- Chen, B. Y., Wang, Y., Wang, D., Li, Q., Lam, W. H. and Shaw, S. L. (2018) Understanding the impacts of human mobility on accessibility using massive mobile phone tracking data. *Annals of the American Association of Geographers*, 108(4), 1115-1133.
- Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014) Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal*

Christina Last

- of Economics*, 129(4), 1553-1623.
- Coolen, A. C. C. (1998) A beginner's guide to the mathematics of neural networks. In: L. J. Landau eds. *Concepts for Neural Networks: A Survey*, London, UK: Springer, 1998, 13-70.
- Cort, D. A., Lin, K. H., and Stevenson, G. (2014) Residential hierarchy in Los Angeles: An examination of ethnic and documentation status differences. *Social science research*, 45, 170-183.
- de Brujin, J. A., de Moel, H., Jongman, B., Wagemaker, J. and Aerts, J. C. J. H. (2018) TAGGS: Grouping tweets to improve global geoparsing for disaster response. *Journal of Geovisualization and Spatial Analysis*, 2(1), 2.
- DeNavas-Walt, C., Proctor, B. and Smith, J. C (2013) Income, Poverty, and Health Insurance Coverage in the United States: 2012 Current Population Reports. *United States Census Bureau*. (Available from: www.census.gov) [Accessed 1st February 2019]
- Dong, G., Wolf, L., Alexiou, A. and Arribas-Bel, D. (2019) Inferring neighbourhood quality with property transaction records by using a locally adaptive spatial multi-level model. *Computers, Environment and Urban Systems*, 73, 118-125.
- Feng, Y. and Jones, K. (2015) Comparing multilevel modelling and artificial neural networks in house price prediction. In: Y. Leung, D. Guo and C. Chen eds. *2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM) 2015*, Fuzhou, 108-114.
- Fowler, C. S., Frey, N., Folch, D. C., Nagle, N. and Spielman, S. (2019) Who are the People in my Neighborhood?: The "Contextual Fallacy" of Measuring Individual Context with Census Geographies. *Geographical Analysis*, 0, 1-14.
- FWHA (2014) Mobility Challenges for Households in Poverty. *2009 National Household Travel Survey*, 1-4.
- FWHA (2016) Hispanics and Transportation. *2009 National Household Travel Survey*, 1-4.
- Geurs, K. T. and Wee, B. (2004) Accessibility evaluation of land-use and transport strategies: review and research directions, *Journal of Transport Geography*, 12(2) 127-140.
- Glaeser, E. L., Kahn, M. and Rappaport J. (2008) Why do the poor live in cities? The role

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

- of public transportation. *Journal of Urban Economics*, 63(1), 1–24.
- Goulias, K. G. (2002) Multilevel analysis of daily time use and time allocation to activity types accounting for complex covariance structures using correlated random effects. *Transportation*, 29(1), 31-48.
- Hansen, W. G., (1959) *Accessibility and Residential Growth*, Cambridge: Massachusetts: MIT Press, 1959, 1-46.
- Harris, C. D., and E. L. Ullman (1945) The Nature of Cities. *The Annals of the American Academy of Political and Social Sciences*, 242, 7-17.
- Hong, J. and Goodchild, A. (2014). Land use policies and transport emissions: Modeling the impact of trip speed, vehicle characteristics and residential location. *Transportation Research Part D: Transport and Environment*, 26, 47-51.
- Jones, K., Manley, D., Johnston, R. and Owen, D. (2018) Modelling residential segregation as unevenness and clustering: A multilevel modelling approach incorporating spatial dependence and tackling the MAUP. *Environment and Planning B: Urban Analytics and City Science*, 45(6), 1122-1141.
- Kain, J. F. (1968) Housing segregation, negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics*, 82(2), 175-197.
- Kasarda, J. D. (1989) Urban industrial transition and the underclass. *The Annals of the American Academy of Political and Social Science*, 501(1), 26-47.
- Kneebone, E and Garr, E. (2010) The Suburbanization of Poverty: Trends in Metropolitan America, 2000 to 2008. *Metropolitan Policy Program*, Washington, DC.: The Brookings Institution. (Available from: www.brookings.edu) [Accessed 21st March 2019]
- Kim, J. and Kwan, M-P. (2019) Beyond Commuting: Ignoring Individuals' Activity-Travel Patterns May Lead to Inaccurate Assessments of Their Exposure to Traffic Congestion. *International Journal of Environmental Research and Public Health*, 16(1), 89.
- Kwan, M-P. (1998) Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework. *Geographical Analysis*, 30(3), 191-216.
- Kwan, M-P., Murray, A., O'Kelly, M. (2003) Recent advances in accessibility research:

Christina Last

- Representation, methodology and applications, *Journal of Geographic Systems*, 5, 129-138.
- Kwan, M-P. and Ren, F. (2009) Analysis of Human Space-Time Behavior: Geovisualisation and Geocomputational Approaches. In: K. S. Hornsby and M. Yuan eds. *Understanding Dynamics of Geographic Domains*, Florida: CRC Press, 2009, 93-113.
- Kwan, M-P. and Weber, J. (2003) Individual accessibility revisited: implications for geographical analysis in the twenty-first century. *Geographical Analysis*, 35(4), 341-353.
- Li, M., Lu, F., Zhang, H. and Chen, J. (2019) Predicting future locations of moving objects with deep fuzzy-LSTM networks. *Transportmetrica A: Transport Science*, 1-18.
- Litman, T. (2003) Measuring Transportation: Traffic, Mobility and Accessibility, *Institute of Transportation Engineers. ITE Journal*, 73(10), 28-32.
- Manley, D. J. (2006) *The modifiable areal unit phenomenon: An investigation into the scale effect using UK census data*. Thesis(PhD). University of St Andrews.
- Miller, E. J. (2018) Accessibility: measurement and application in transportation planning, *Transport Reviews*, 38(5), 551-555.
- Miller, H. J. (1991) Modelling accessibility using space-time prism concepts within geographical information systems, *International Journal of Geographical Information System*, 5(3), 287-301.
- Muth, R. F. (1969) *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. Chicago: University of Chicago Press, 1969.
- National Center for Law and Economic Justice (2014) *Poverty in the United States: A Snapshot*. (Available at: www.nclcj.org) [Accessed 3rd February 2019].
- Openshaw, S. and P. J. Taylor (1979) A million or so correlation coefficients, three experiments on the modifiable areal unit problem. In: N. Wrigley ed. *Statistical Applications in the Spatial Sciences*. London: Pion, 1979, 127-144.
- Ord, J.K. and Getis, A. (1995) Local Spatial Autocorrelation Statistics: Distributional Issues and an Application, *Geographical Analysis*, 27(4).
- Petersen, M., Bardacke, T., Reyes, S., Obergfell, J., Firestone, H., Samulon, M. and Cole, R. (2018) *pLAN: Transforming Los Angeles* (Available at: plan.lamayor.org). [Accessed

Predicting accessibility in Los Angeles County: a comparative study using MLM and ANN

13th March 2019]

- Terrazas, A. (2018) Lower-Income Renters Pay Almost 2/3 of Their Income on Even Modest Apartments, *Zillow Research*. (Available at: www.zillow.com) [Accessed 12th March 2019]
- U.S. Census Bureau (2012) *Hispanic or Latino Origin by Race American Community Survey 5-year estimates*. (Available at: censusreporter.org). [Accessed 22nd January 2019]
- United States Department of Agriculture (2000) *Commuting Zones and Labor Market Areas*. (Available at: www.ers.usda.gov). [Accessed 12th February 2019]
- Van Wee, B. and Geurs, K. (2011) Discussing equity and social exclusion in accessibility evaluations. *European Journal of Transport and Infrastructure Research*, 11(4).
- Vega and Elhorst (2015) The SLX Model, *Journal of Regional Science*, 55(3) 339-363.
- Vickerman, R. W. (1974) Accessibility, attraction, and potential: a review of some concepts and their use in determining mobility. *Environment and Planning A*, 6(6), 675-691.
- Wang, C., Quddus, M., Enoch, M., Ryler, T. and Davidson, L. (2014) Multilevel modelling of Demand Responsive Transport (DRT) trips in Greater Manchester based on area-wide socio-economic data, *Transportation*, 41, 589–610.
- Weber, J. and Kwan, M-P. (2003) Evaluating the Effects of Geographic Contexts on Individual Accessibility: A Multilevel Approach, *Urban Geography*, 24(8), 647-671.
- Wilson, D. (1996) Metaphors, growth coalition discourses and black poverty neighborhoods in a US city. *Antipode*, 28(1), 72-96.