

## **Exploring spatial non-stationarity in the social determinants of time to work in a metropolitan area**

**Paulo Rui Anciaes**

Centre for Transport Studies, University College London

**Giles Atkinson**

Dept. Geography and Environment, London School of Economics and Political Science

### **Abstract**

Social differences in time to work are a growing concern of transport planning, as they can increase the risk of social exclusion of some groups. Recent work has found that those differences tend to vary across the urban space. Geographic Weighted Regression models this variability by estimating relationships in each place, weighting data by distance to that place. This paper builds on this work by assuming that households and policy-makers make decisions over space by comparing places according to their function within the metropolitan space, regardless of their geographic proximity. The determinants of time to work then vary in the space defined by variables identifying that function. This procedure is applied to the case of the Lisbon Metropolitan Area at two moments in time. Higher times to work are associated with less-qualified populations, with the strength of the relationships generally increasing with centrality in relation to employment locations and dependence on jobs in the main metropolitan centre. The priority to investments in road transport in the period concerned extended the disadvantages to second-ring suburban areas.

## **1. Introduction**

Political interest in social differences in time to work has increased in recent years, following calls from researchers and activists for integrating social aspects in the formulation of transport policies (Hine 2003; Lucas and Stanley 2009; Lucas and Jones 2012). The perception of a social dimension in the distribution of accessibility also has an effect on the socio-political feasibility of those policies, often leading to protest by political parties and local communities. However, in face of multiple claims for accessibility in different neighbourhoods, the definition of policy priorities is not straightforward. Further research and political discussion is needed in order to clarify which areas are at disadvantage. The accessibility needs of the populations living in different parts of large urban regions tend to be heterogeneous, and therefore the identification of policy-relevant disadvantages should be based on comparable sets of neighbourhoods.

Empirical studies have explored the variability of accessibility differences across space, using methods such as Geographically Weighted Regression (GWR). The contribution of this paper to this field is to consider that relationships between time to work and socio-economic factors vary not in geographic space but in the space defined by the geographical attributes that shape the perceptions of households and policy-makers. These attributes define the function of each place within the metropolitan space, that is, its role as origin and destination of daily trips.

The analysis is applied to the case of the Lisbon Metropolitan Area. This area is characterized by a high degree of centralization in the main city (Lisbon) and by the heterogeneity in the social structure and land use of the different sub-regions. Given these characteristics, the variables chosen to identify the function of each place are centrality (measured by a gravity-type indicator of potential job accessibility) and dependence in relation to the main centre (measured by the proportion of employed population working in

that centre). The analysis focuses on the 1990s, a period of unprecedented changes in the motorway network and fast urban growth and fragmentation. The paper shows that the social determinants of time to work vary across functional space and that the transport changes in the period concern extended the areas where disadvantages are visible towards outer suburbs. These results raise questions regarding policy priorities and the overlap between the spatial scales at which inequalities appear and the scales of application of policy instruments.

The paper proceeds as follows. The next section reviews existing knowledge on disadvantages in time to work and their spatial variability, and describes how this paper adds to this knowledge. The following section presents the model formulations used to explain the relationships between time to work and socio-economic variables. The results of the global (unweighted) regression model and of the model weighted in geographic space (GWR) are presented next. These results provide the background for the main section of the paper, which presents the results of the models weighted in functional space. The two last sections discuss and summarize the policy and research implications of the study.

## **2. Time to work and spatial non-stationarity**

It is now generally accepted that transport planning is an important part of social policy, as there is evidence that the lack of accessibility may reinforce the risk of social exclusion of some groups, such as the elderly, young workers, and low-income or low-qualified individuals (Preston and Rajé 2007; Lucas 2012; Delbosc 2012, Reardon and Abdallah 2013, Stanley and Stanley 2014). Equity is increasingly considered as one of the criteria for evaluating the performance of a transport system, which should fulfil the needs of all individuals regardless of the area of the city where they live or their socio-economic characteristics.

A variable of special concern is time to work, due to its links with employment prospects and wages. The 'spatial mismatch' hypothesis (Kain 1968) states that the employment outcomes of underprivileged groups are in part explained by the distance to the areas where job opportunities are located. Evidence in a large number of urban areas found inequalities in the distribution of job accessibility, even after accounting for occupational differences (Cervero, Rood, and Appleyard 1999; Stoll, Holzer, and Ihlanfeldt 2000). However, it is difficult to pinpoint general mechanisms of causality in the emergence of those inequalities (Gobillon, Selod, and Zenou 2007). The reason for this complexity is that accessibility differences are not a purely geographic issue of separation between jobs and people. They depend on economic factors, such as differences in average wages or in the competition between workers at each employment centre (Van Wee, Hagoort, and Annema 2001; Wang 2003; Bunel and Tovar 2013). They also depend on the availability and efficiency of the transport system linking jobs and people in different parts of the urban area (Ong and Miller 2005).

Social differences in time to work are often associated with different levels of access to different modes of transport, and in particular, on levels of car ownership and use. This may lead to a mismatch between the type of accessibility provided by the transport system at each place and the modes of transport that the population consider as feasible options (Kwok and Yeh 2004; Kawabata 2009). This hypothesis is especially relevant in the cases of low-income households and ethnic minorities, who are often dependent on public transport (Shen 1998; Hess 2005; Grengs 2012). This phenomenon tends to be more visible in cities with employment dispersion, low population densities and orientation towards car travel (Kawabata and Shen 2006). Although these are traditional characteristics of North American cities, they increasingly apply in the European context, due to tendencies for suburbanization

and employment decentralization (Dujardin, Selod, and Thomas 2008; Korsu and Wenglenski 2010; Matas, Raymond, and Roig 2010).

In the existing literature, the social distribution of time to work is usually studied by searching for patterns occurring in a pre-determined scale: that of a city, metropolitan area or region. However, the borders of these areas do not necessarily correspond to discontinuities in the processes affecting the spatial distribution of population, jobs and transport infrastructure. The analysis may hide relevant detail if these processes are non-stationary in space, that is, if they vary among the different parts of the study area. In fact, the relevance of socio-economic factors as determinants of time to work depends on the contextual factors that influence individuals' preferences and choices in the housing and transport markets and the transport and land use policies applied in each area. This level of detail can only be captured by models where parameters are allowed to vary with location.

The spatial variability in the relationships between accessibility measures and socio-economic factors has attracted growing interest, following advances in modelling procedures and in the development of specific software. Spatial variability can be addressed by estimating separate statistical associations between social variables and accessibility indicators on different partitions of the data. These partitions can be based on a broad distinction between inner/outer/fringe/regional areas (Delbosc and Currie 2011) or by cluster analysis (Boschmann and Kwan 2010). The associations can also be related across more than one scale. For example, multilevel models estimate relationships at the level of individuals and places aggregated at different levels. These models have been used in the analysis of inequalities in distance between home and work (Shuttleworth and Gould 2010) and in space-time accessibility (Kwan and Weber 2008). These methods assume, however, that the processes analysed are discontinuous at the borders between areal units, a hypothesis that collides with the fact that most spatial processes are unbounded (Leung 1987).

An alternative is to assume that model parameters follow trends across space (spatial expansion models). This approach was used by Roorda et al. (2010) and Morency et al. (2011) to show that the elderly and low-income groups tend to make fewer trips and travel shorter distances, but that mobility limitations are found only in suburban areas. The hypothesis also applies to measures of accessibility to specific destinations such as food shops (Páez et al. 2010). Although this method is less rigid than estimations on data partitions, it still presupposes a pre-determined, explicit, and regular structure for the variability of the processes modelled.

Geographically weighted regression (GWR) (Brunsdon, Fotheringham, and Charlton 1996; Fotheringham, Brunsdon, and Charlton 2002) provides a solution to model the cases where there are no clear assumptions about the scale at which non-stationary processes operate or the trends they follow over space. This method assumes that there is a continuous surface of regression parameters across the study area. A separate regression is then estimated at each place, using the data for all other places weighted by a decreasing function of distance. The method is flexible in the definition of hypothesis about the type of variability, as it relies on user-defined options regarding the number of data points included in each regression and the specification of the weight function. Lloyd and Shuttleworth (2005) have used this method to show the non-stationary character of differences in commuting times within a broad region. GWR has also been used to explore the spatial variability in accessibility to health services (Comber, Brunsdon, and Radburn 2011) and green space (Maroko et al. 2009).

The hypothesis of this paper is that the spatial variability of the social determinants of time to work can be assessed not only in terms of geographic space but also in terms of the space defined by attributes defining the function of each place, as these are the attributes limiting the options and shaping the perceptions of the actors behind the economic or political processes leading to social differences in time to work. The geographic scale of these

processes is linked to the set of alternative places considered by households, businesses and policy-makers when deciding on the location of residences, jobs and transport infrastructure.

The assumption implicit in the GWR method is that the degree to which a place is an alternative for another place depends on the geographic locations of both. As weights are a decreasing function of distance, in practice, local regressions measures the relationships occurring in the areas around each place. However, urban space is not usually perceived as isotropic and continuous, especially in large and heterogeneous metropolitan areas. In fact, actors make decisions over space by considering a set of places located in different parts of the metropolitan area, but having a similar function. This function may be assessed by the degree of centrality of each place, with different sets formed by centres with different sizes and peripheral places. It may also be assessed by its degree of dependence to the main metropolitan centre, with sets formed by the main centre, surrounding commuting rings, secondary centres and semi-rural areas.

Time to work will then vary within the spaces defined by those two attributes, depending on factors such as the spatial mismatch between residences and jobs, the normal speed on the available private and public transport links, the choices of the population over travel modes, and the effects of congestion. Places of residence and work and transport facilities may then be sorted according to the socio-economic characteristics of the population, leading to different times to work. The social determinants of time to work will vary in each subset of the study area because the function of those subsets determines the degree to which that sorting occurs. The approach of this paper is to assume that the relationships between time to work and socio-economic variables are specific to all places with similar degrees of centrality and dependence in relation to the main centre. A separate regression is then estimated for each place, using the data for all places weighted by their distances measured not in geographic space but in the space defined by indicators of those two attributes.

### 3. Variables and models

The models explain the spatial distribution of time to work as a function of socio-economic variables in 1991 and in 2001, using the census enumeration district as unit of analysis.

The estimation of average time to work in each district is based on data from the 1991 and 2001 census commuting flows databases, which are available at the level of administrative areas. These data were disaggregated into flows from enumeration districts to a large set of employment centres, using ancillary datasets such as business data, mobility surveys, land use maps, and location of major employers in both years of analysis. The flows were then split by travel mode (private transport, public transport and walking) and time of day (peak and off-peak). A Geographic Information System was used to estimate the optimal routes and corresponding travel times from districts to employment centres by travel mode and period of day. This modelling considered all walking and waiting sections of the journey to work and the effects of congestion on crossing times in all links of the private and public transport networks. Travel times were finally averaged in each district according to commuting flows.

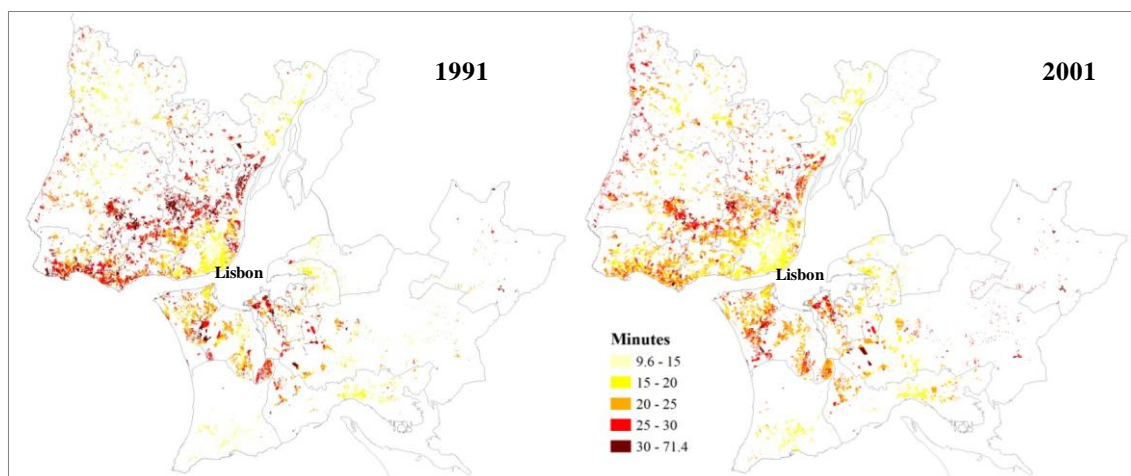
The distribution of times to work in 1991 and 2001 is depicted in **Figure 1**. Times are highly variable and generally higher in the municipalities around Lisbon. However, there are significant differences in values found in different parts of each municipality and within the different commuting corridors to Lisbon. There are also clusters of high times to work in some parts of the outer municipalities. Times to work decreased slightly from 1991 to 2001 in most suburban areas, but showed a tendency to increase in the Southwest part of the North Bank and in the Eastern part of the South Bank.

The explanatory variables are a series of vectors derived from a factor analysis to the pooled dataset of 1991 and 2001 census data. The factors are standardized, with zero mean and unit variance. Although there is an element of subjectivity to the interpretation of the factors, this approach was chosen because it minimizes methodological issues related to



variable selection and multicollinearity. This is because the method synthesizes the social and geographical structure of each neighbourhood using a small number of largely uncorrelated factors.

**Figure 1: Estimated time to work in the Lisbon Metropolitan Area**



**Table 1: Correlations between factors and original census variables**

	Qualifications	Age	Urbanization
Men (% of adults)	-0.19	-0.27	-0.30
Elderly	0.02	0.90	-0.16
No/Lowest qualification	-0.67	0.34	-0.45
Graduates	0.80	0.13	0.34
Employment/Population	0.20	-0.47	0.39
Employment in services	0.51	0.28	0.48
Family: 1-2 members	-0.03	0.86	0.19
Family with children (<15yrs old)	-0.03	-0.84	-0.02
Households per dwelling	0.07	0.30	0.11
Informal dwellings	-0.02	-0.06	0.00
Dwellings with basic facilities	0.17	-0.02	0.19
Large dwellings (>5 rooms)	0.83	-0.06	-0.17
Owned dwellings	0.46	-0.49	-0.03
Dwellings per building	0.10	-0.25	0.73
Building: more than 40 yrs old	-0.04	0.73	-0.08
Building: Exclusively residential	-0.04	-0.12	-0.70

**Note:** All original variables are expressed as ratios. Households per dwelling and dwellings per building are percentages of maximum values in each year.

**Table 1** presents the correlations between the socio-economic factors and the original census variables. The most relevant factor for the analysis is labelled *Qualifications* and is positively correlated with the percentage of graduates, employment in services, and large and owned-occupied dwellings, and negatively correlated with the percentage of individuals with no qualifications or the lowest qualifications. Two other factors are treated as control variables, as they are general indicators of the geographic and socio-economic characteristics of each area. *Age* is related to the age of buildings and people and to the proportion of 1-2 member families with no children. *Urbanization* is positively related to number of dwellings per building and negatively related to the proportion of non exclusively-residential buildings.

The models are formed by a set of regressions, centred at each enumeration district. In the regression for district  $i$ ,  $y_i$  is the dependent variable,  $x_{ik}$  are the explanatory variables and  $b_0$  and  $b_k$  are respectively the intercept and the regression parameters. These parameters are a function of  $S_i$ , the coordinates in space. In the case of GWR, the coordinates are defined in geographic space. In the case of models weighted in functional space, they are defined in the space defined by variables defining centrality and dependence on Lisbon.

$$y_i = b_0(S_i) + \sum_k b_k(S_i)x_{i,k} + \varepsilon_i \quad (1)$$

Centrality is specified by a gravity measure (Hansen 1959). In each district, the measure is the sum of jobs in all employment sites, weighted by a negative exponential function of the distance to those sites, measured on the road transport network. The parameter that defines the steepness of the decay of centrality with distance was obtained by modelling commuting flows as a function of distance, using a simple trip-distribution gravity model (Ortúzar and Willumsen 2006, ch.5). Dependence on Lisbon is measured by the percentage of workers in each district commuting to Lisbon municipality, estimated from census data. It should be noted that the two variables are largely independent from each other, thus providing two distinctive axes along which the function of each place can be assessed.

Three different models were tested. Two of them use a unidimensional space formed by the values of each of the two variables described above. A third model uses the bi-dimensional space formed by the standardized values of the two variables. These values are obtained by ranking observations.

The local regressions are linear, as it was found in preliminary analysis that non-linear transformations of dependent or independent variables tend to produce poorer models. The specification of the error terms  $\varepsilon_i$  follows the assumptions of the standard linear regression models: independence and identical distribution of the residuals, which have zero mean and constant variance.

The estimation of the regression for each district includes data for all districts weighted by a decreasing function of distance. The vector of parameters for the regression at district  $i$  is estimated as  $b=[XTW(i)XJ^{-1}XTW(i)Y]$ , where  $W(i)$  is a diagonal matrix whose elements in the main diagonal give the weights assigned to each district in the data set. (See Fotheringham, Brunsdon, and Charlton, ch.2.7). The distance between districts in the models weighted in functional space is the difference in the values of centrality and dependence on Lisbon. For example, in the first case, the regression for a given district includes all districts with the same level of centrality. Districts with different levels of centrality are only partly included, weighted by the difference in centrality between the two districts. Districts with very different centrality levels have a weight close to zero and are virtually excluded from the estimation. In the case of the model in bi-dimensional space, the distances are the Euclidian distances in the space formed by the standardized values of the two variables. It is assumed that weights decrease according to a Gaussian function. The weight of district  $j$  in the regression for district  $i$  is given below, where  $d_{ij}$  is the distance between the two districts.

$$w_{ij}=\exp[-(d_{ij}/b)^2] \quad (2)$$

This function depends on a positive parameter  $b$  called the bandwidth. The lower the bandwidth, the steeper is the decay of the weights with distance. This parameter can be assigned a-priori or estimated by minimizing indices of the models' goodness of fit. In the analysis that follows the value of the bandwidth is the one that minimizes the Akaike Information Criteria (AIC), an index that takes into account the estimated standard deviation of the error term and the number of degrees of freedom in the model, which depend in turn on the value of the bandwidth. Details of this procedure can be found in Hurvich, Simonoff, and Tsai (1998). Because it incorporates the number of degrees of freedom of the model, the AIC index is also the most accurate method to compare the goodness of fit of the global model and of different weighted models using the same data.

It is important to note the advantages of this formulation when comparing with alternatives. Spatial variability could be inferred by modelling regression parameters as functions of centrality and dependence on Lisbon, using interaction terms or more complex formulations. However, this method imposes a rigid structure to the variability of the relationships. Besides being more flexible, regressions weighted in functional space also correct for spatial autocorrelation. This issue tends to affect accessibility studies as both census and commuting variables usually assume similar values in nearby places, leading to biased and unreliable parameter estimates. The use of GWR mitigates this problem (Fotheringham, Brunsdon, and Charlton, pt.5) but it may introduce other distortions. As GWR models are estimated in the areas around each place, their explanatory power depends on the availability of data at a detailed level. In the case of Lisbon, the variations in time to work within small areas would not take into account differences in the populations' daily destinations, as this information is only available at a level higher than the enumeration district. This problem is partially solved in the models weighted in functional space, as regressions include places located in different parts of the metropolitan area, thus accounting for variations in destinations.

#### 4. Global model and GWR

**Table 2** presents the results of the global (unweighted) regression model. In both years, time to work is related negatively with the three factors. In other words, time to work is in average lower in districts inhabited by populations with higher qualifications and in older and more urbanized parts of the metropolitan area. *Age* is the most important factor in 1991 but in 2001 *Qualifications* is slightly more important. The goodness of fit of the model is modest, as it explains only 16% and 18% of the dependent variable in 1991 and 2001 respectively.

The small explanatory power of the global model suggests that this model is hiding significant variations in the regression parameters across the study area. **Figure 2** sheds some light on this hypothesis, by synthesizing the results of the GWR model. The figure plots the four quartiles of distribution of the *Qualifications*, *Age* and *Urbanization* parameters. The three parameters are more variable in 1991 than in 2001. Despite the fact that the global model yields negative parameters for the three variables, almost 25% of the districts show positive values for at least one parameter in 1991. The number of districts with positive values is lower in 2001. The average of the estimates is slightly higher for all three parameters in 2001, especially in the case of the *Age* variable. Overall, the GWR is a better model than the global model, as shown by a lower AIC in both years.

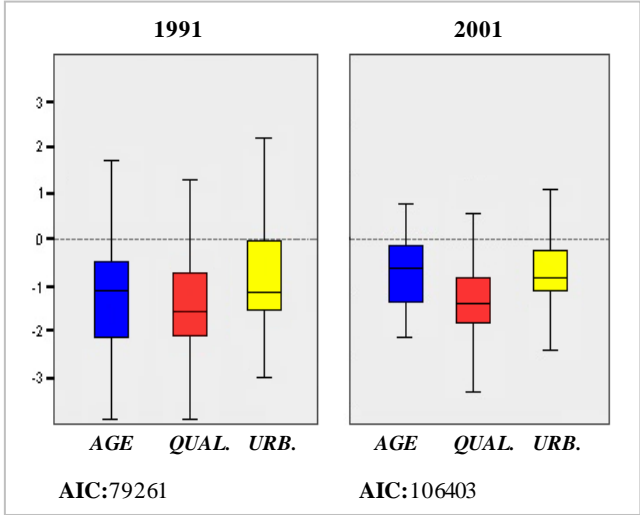
The spatial distribution of the *Qualifications* parameter is given in **Figure 3**. The positive parameters in 1991 tend to occur in the western part of the North Bank of the River Tejo. However, by 2001, most of this region shows negative parameters. The areas showing the highest negative values remain relatively stable from 1991 to 2001, including Lisbon and surrounding municipalities, the Northwest corridor of access to Lisbon, and parts of the South Bank.

**Table 2: Global regression between time to work and socio-economic variables**

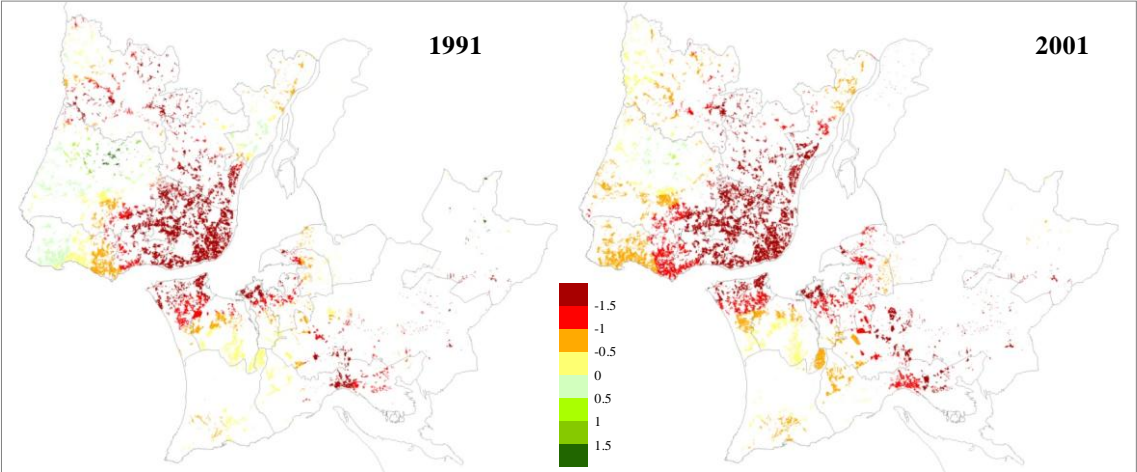
Variable	1991		2001	
	Coefficient	Standard deviation	Coefficient	Standard deviation
Constant	23.256	(0.054)***	22.364	(0.034)***
Age	-2.049	(0.044)***	-1.233	(0.029)***
Qualifications	-1.060	(0.045)***	-1.327	(0.026)***
Urbanization	-0.367	(0.051)***	-0.815	(0.030)***
Observations	14755		20847	
R <sup>2</sup>	0.16		0.18	
AIC	92001		118527	

\*\*\* significant at the 1% level, \*\* significant at the 5% level

**Figure 2: GWR model - parameter variability**



**Figure 3: GWR model: Distribution of the Qualifications parameter**



## 5. Models weighted in functional space

This section presents the results of the models weighted in functional space, comparing their parameter distribution and goodness of fit with the global and GWR models. The focus is first on the models weighted in the unidimensional spaces based on centrality and dependence on Lisbon, analysing parameter variability and the trends followed by parameters for different values of the coordinates. We then look at the model weighted in the dimensional space based on the two variables, analysing parameter variability and the spatial distribution of parameters.

**Figure 4** presents the statistical distribution of the parameter estimates of the two models weighted in unidimensional space. Higher times to work tend to be negatively associated with the three parameters. This means that when comparing with other districts with similar levels of centrality or dependence on Lisbon, populations who are more qualified or live in older and more urbanized districts tend to have shorter times to work. Although in 1991 some districts show positive values for all least one parameter in the model weighted on centrality levels, in 2001 only a small minority of districts show positive values, and only for the *Age* and *Urbanization* parameters. The parameters in the model weighted on levels of dependency to Lisbon are negative in both years. In both models, all parameters have lower variability in 2001. The average of parameter estimates is higher in 2001 in the case of the *Age* and *Urbanization* parameters, but remained roughly the same in the case of the *Qualifications* parameter.

**Figure 4: Models weighted in unidimensional functional space: parameter variability**

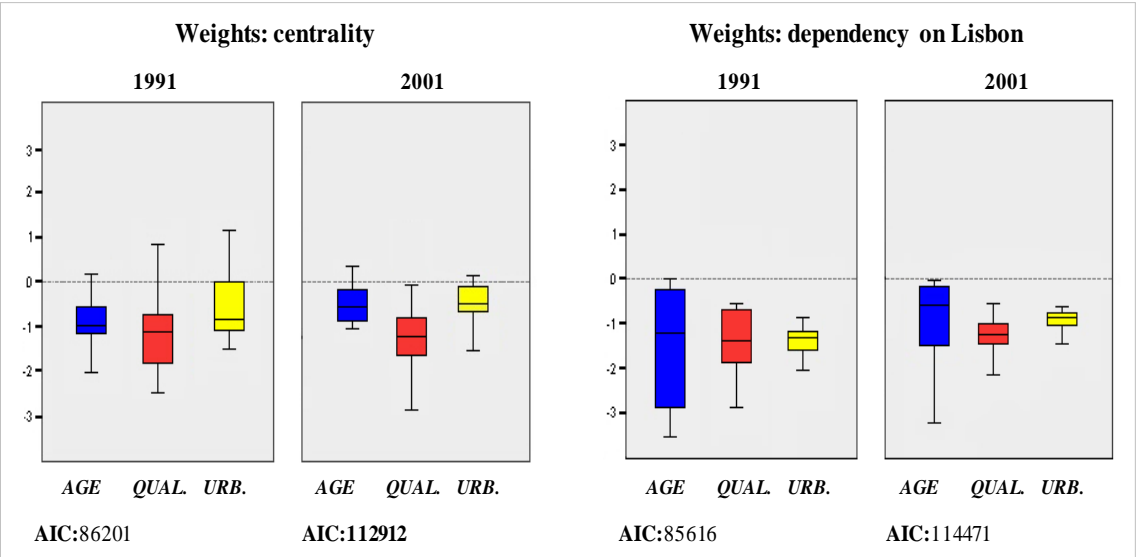


Figure 5 plots the values of the three main regression coefficients against the indicators of centrality. The *Qualifications* parameter is lowest in the third quartile of distribution of centrality. This interval corresponds to secondary centres and areas near main centres. As we move from these areas to the main centres themselves, or to less central areas, the negative influence of the *Qualifications* variable becomes less marked. However, the increase in the *Qualifications* parameter as we move towards less central areas has become less steep from 1991 to 2001. This suggests that the advantages of more qualified populations in average time to work have extended to less central regions. In particular, the positive parameters occurring in the second quartile of centrality in 1991 changed sign in 2001. The fluctuations of the *Age* and *Urbanization* parameters have also decreased considerably in the period of concern. In 2001, the influence of these factors on time to work was negative and relatively stable across the metropolitan area. The advantage of populations in older and more urbanized areas has therefore decreased in most regions.

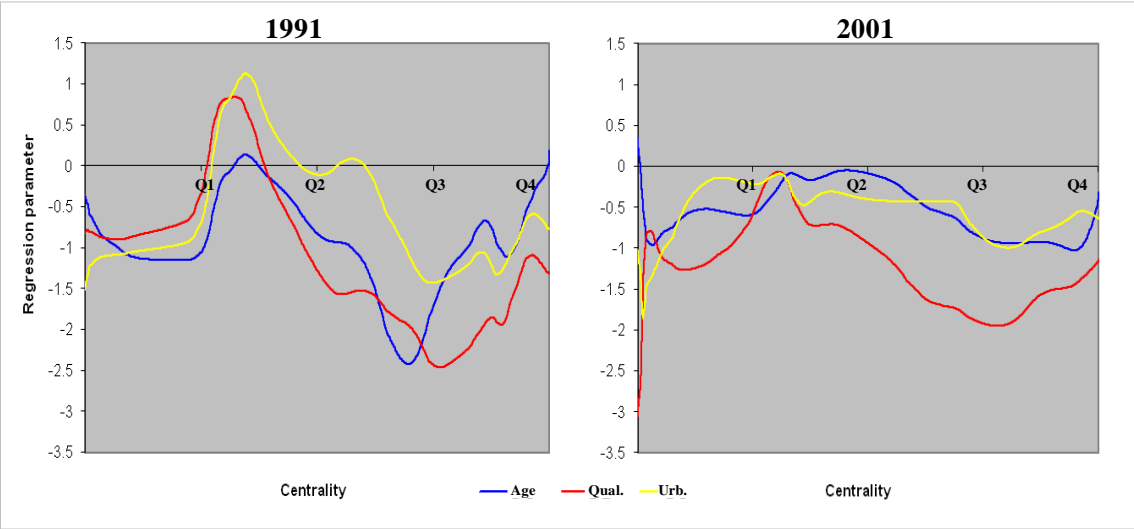
It is relevant to notice that positive parameters in 1991 are mainly limited to the second quartile of centrality. These are areas relatively far from main employment centres but not at the periphery of the metropolitan area. The changes occurring in the parameters of these areas



from 1991 to 2001 may be related to the construction of new transport infrastructure and the urbanization of these areas, which allowed for a widening of the commuting possibilities. The higher proportion of commuters to other municipalities amplifies the differences in time to work that are due to different levels of car usage, comparing with the original case when most workers travel to nearby destinations. The changes in the regression parameters in these areas may then reflect changes in daily destinations and the fact that less-qualified and older populations have lower rates of car usage.

**Figure 6** plots regression coefficients against levels of dependence on Lisbon. The profile of the *Qualifications* variable is very similar in both years, decreasing with levels of dependence on Lisbon. In other words, the highest the proportion of workers commuting to Lisbon, the highest the disadvantage of low-qualified populations in terms of time to work. The *Age* parameter follows a non-linear trend, increasing in the first quartile of dependence to Lisbon, approaching zero in the second quartile, decreasing in the third quartile and increasing in the fourth quartile. The advantage of people in older areas is therefore the highest in the third quartile. However, from 1991 to 2001, this advantage has decreased substantially. The *Urbanization* parameter increases slightly with dependence on Lisbon.

**Figure 5: Models weighted in the space defined by centrality space: trends**



**Figure 6: Models weighted in the space defined by dependence on Lisbon: trends**

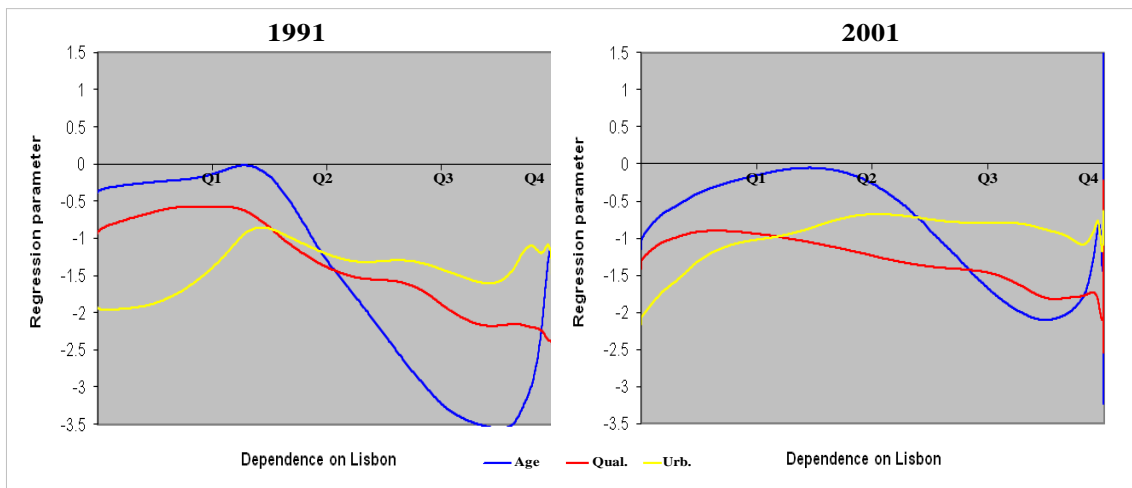
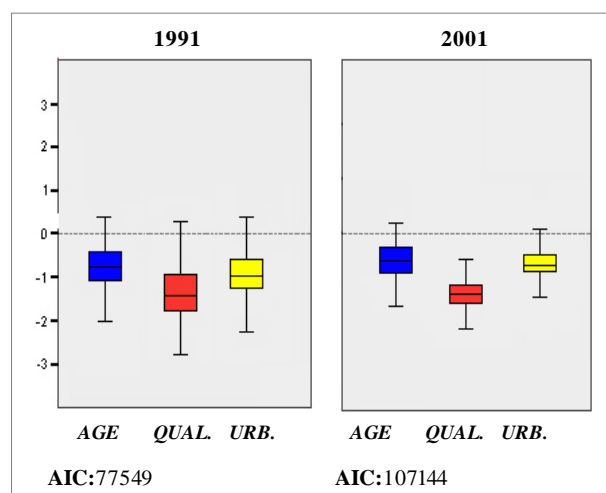
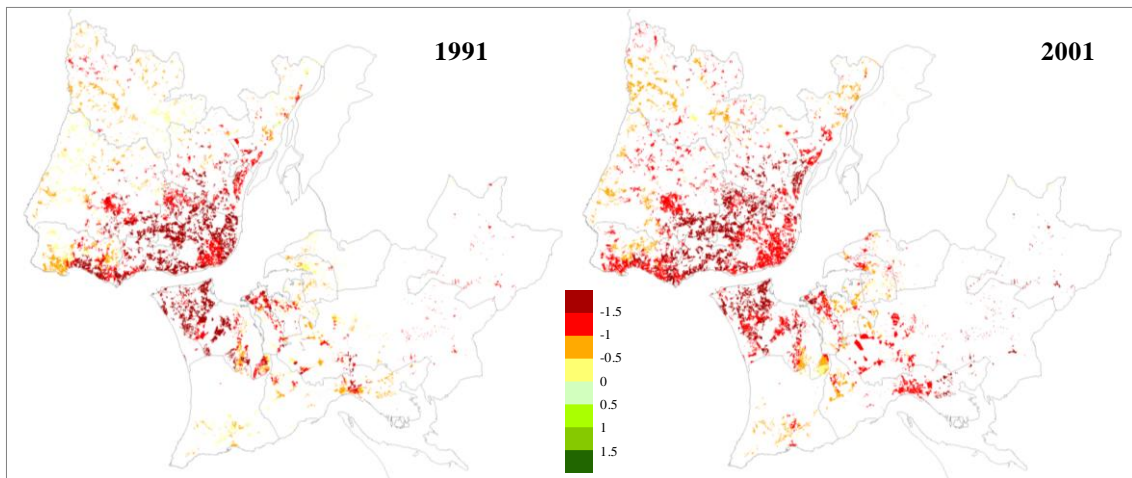


Figure 7 shows the statistical variability of the regression coefficients in the model weighted in the bidimensional space formed by degree of centrality and dependence on Lisbon. All parameters are less variable than those in the case of the GWR and of the models weighted in unidimensional space. Positive values appear only in a small minority of districts. However, contrary to the GWR case, the *Qualifications* parameter is always negative in 2001. Similarly to all previous models, variability in the three parameters decreased from 1991 to 2001.

**Figure 7: Model weighted in bidimensional functional space: parameter variability**



**Figure 8 Model weighted in bidimensional functional space: Distribution of the *Qualifications* parameter**



The maps in **Figure 8** show the distribution of the *Qualifications* parameter for the bi-dimensional model. The highest negative values for this parameter occur in the first suburban ring around Lisbon. However, the negative influence of qualification levels has extended towards the areas beyond that ring from 1991 to 2001. In addition, the parameter decreased in some of the most peripheral regions in the metropolitan area, including the Northwest municipality and the outer municipalities in the South Bank. There are areas in the lowest quartile of centrality but where commuting to Lisbon has increased considerably in the period concerned. The two most important geographic foci of disadvantage in 2001 are therefore the areas surrounding Lisbon and some of the peripheral areas.

The analysis of the values of the AIC index suggests that all three models weighted in functional space have a better goodness of fit than the global model. This confirms that relationships estimated by the latter cancel out the roles that socio-economic factors play in explaining time to work in different subsets of the metropolitan space. In 1991, the model weighted in bi-dimensional functional space has also a smaller AIC index than the GWR model. However, in 2001 the GWR model has a smaller AIC, thus showing a better goodness of fit.

The hypothesis of spatial non-stationarity was tested for the three parameters in all models. The test made use of Monte Carlo simulations to compare the position of the standard error of the coefficients estimated in all local regressions within a rank list of values obtained re-estimating the model with a large number of different random rearrangements of data in space (Brunsdon, Fotheringham, and Charlton 1996). The null hypothesis that the process is stationary was rejected in all cases at the 1% level, confirming that the parameter variability obtained by the models is the result of non-stationary processes.

## **6. Discussion**

This section reflects on the results of the quantitative analysis, focusing on three aspects: how the results answer the main research question posed in the introduction; the implications for transport policy; and the adequacy of the methods used to study the spatial variability in the social determinants of time to work.

The analysis confirmed the relevance of understanding spatial variability in the social determinants of time to work in terms of attributes other than geographic location. The hypothesis of variability in the space defined by centrality and dependence on the main centre was confirmed for all parameters. The models weighted in functional space also showed a better goodness of fit than the GWR model in one of the years analysed. More importantly, the charts and maps of models estimates revealed patterns that are not apparent when using the GWR approach. The results suggest that the sorting of social groups to areas allowing different times to work operates mainly at the scale of all suburban areas around the main centre, while there is also evidence of spatial sorting in some of the most peripheral areas.

The results have implications for transport planning in the Lisbon Metropolitan Area, which also apply to other urban areas with the same degree of internal heterogeneity in terms of socio-economic and land use attributes. The finding of multiple patterns of disadvantage

for less-qualified populations in different parts of the metropolitan area raises questions about which of these disadvantages deserves policy priority. The methods for dealing with these disadvantages are different, as the nature of the unfulfilled mobility needs depend on the context in which those needs arise. In suburban areas, the improvement of times to work for the populations at disadvantage would involve the provision of public transport or policies tackling congestion. In semi-rural regions, the inequalities are more purely geographical, as the populations at disadvantage tend to be the ones who live in villages away from the local centres of employment (the municipal capitals) and from the places with direct access to the main metropolitan centres. Given the dispersion of the population and the lack of economic viability for the extension of bus routes, the improvement of accessibility in these areas would involve either the construction of more roads or the provision of subsidies to individuals with special accessibility needs.

The results also shed light about the degree to which the scale at which inequalities appear overlaps the scale of application of transport policies. With the exception of the patterns occurring within the borders of the Lisbon municipality, the spatial distribution of the regression parameters suggests that the scale of the socio-economic and political processes that lead to inequalities does not correspond to the jurisdictions of the institutions with effective power and responsibility over transport issues in the metropolitan area, that is, the central and the municipal governments. The correction of inequalities would then require the coordination between different municipalities, especially in the suburban areas around Lisbon, and the increase in the means available to supra-municipal institutions, such as the recently created Lisbon Metropolitan Transport Authority.

However, the policy instruments to address accessibility inequalities may not always be available at every spatial scale, especially in the cases where these inequalities are linked to insufficiencies in public transport. In the Lisbon Metropolitan Area, policy intervention on the

public transport system is limited to the definition of broad strategies implemented by the state-owned companies managing the public transport infrastructure and to the financing of private companies providing bus services. The effectiveness of these strategies for dealing with transport-based inequalities is restricted by the fact that these inequalities extend over the different transport corridors but most bus companies operate in only one or two municipalities in each corridor.

It should be noticed that the local regressions in each model represent equilibrium relationships between time to work and the location of social groups at given moments in time, but do not explain the processes leading to these relationships. The analysis in this paper provides guidelines for the identification and interpretation of these relationships, but does not imply that they are product of an adjustment process of individuals to conditions existing in the different neighbourhoods. The hypothesis suggested by the **Figure 8** is that inequalities in time to work in outer suburban areas in the North Bank emerged in the period 1991-2001, following the expansion of the motorway network to this area, while public transport options remained relatively unchanged. Improvements in the private transport network may therefore trigger an adjustment of the social structure of the different neighbourhoods, with more qualified populations moving to the areas with easier road access to the main metropolitan centres. The validation on this hypothesis requires, however, a more detailed analysis of individual choices and housing market dynamics than the one provided in this paper.

The implications of this paper to transport policy must also take into account some caveats of the models used. It was assumed that spatial non-stationarity was a feature of the dataset. However, some of the variation captured by the parameter estimates may derive from the insufficiency of the set of independent variables in explaining differences in the dependent variable. The use of more detailed socio-economic data (ideally at the individual or household level) or the inclusion of further variables could bring further insights into the non-stationarity

character of some of the parameters. This is especially the case of data on household income, a variable that is not available to researchers at the level of the enumeration district.

Geographically Weighted Regression is a developing field in Quantitative Geography and its potential and limitations are still being explored. Simulation studies have found that the method may produce spurious sign reversals and variability and correlations between parameter estimates (Páez, Farber, and Wheeler 2011). However, the issue is less likely to affect analyses using large samples such as in the one in this paper.

Finally, although in the present study the models' residuals did not show any relevant spatial pattern, in general it is difficult to disentangle the effects of spatial non-stationarity, spatial autocorrelation and the errors implicit in the definition of dependent and independent variables. In this regard, the use of alternative definitions for the axis along which spatial variability is measured provides an advantage over the measurement of variability only on geographic space, as there is a smaller probability that the variability arose from chance or from insufficiencies of the model in all the different alternatives.

## **7. Conclusions**

This paper provided evidence that the socio-economic distribution of time to work is the result of a spatially non-stationary process and that instead of a single relationship defining inequality at the level of the metropolitan area, there are multiple relationships at a variety of scales. The use of functional space to model these relationships proved to be a useful method, as it yielded models with better goodness of fit than the global regression in both years of analysis, and better than the GWR model in one of the years analysed.

The analysis also shed light on the spatial dimension of the processes leading to the sorting of socio-economic groups to different times to work. The study found that from 1991 to 2001, a period of fast expansion of the road transport network, the regions where less-qualified

populations are at disadvantage extended from the first to the second suburban ring outside Lisbon. These findings raise several questions for metropolitan transport policy.

More generally, the paper adds to the idea that there are multiple perspectives to the assessment of transport-related social inequalities, and that these perspectives can be implicit in the statistical methods used. The assessment of social and political claims for accessibility should then be made with reference to a set of different perspectives.

### **Acknowledgements**

Thanks go to Steve Gibbons, Helena Titheridge and Andrew Lovett for their comments and suggestions on an earlier version of this paper. Helpful technical advice was provided by Tomoki Nakaya, developer of the GWR4 software. Financial support from the Portuguese Foundation for Science and Technology is acknowledged.

### **References**

- Boschmann, E. E., M-P. Kwan. 2010 Metropolitan area job accessibility and the working poor: Exploring local spatial variations of geographic context. *Urban Geography* 31(4):498-522.
- Brunsdon, C., A. S. Fotheringham, and M. Charlton. 1996 Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28(4):281-298.
- Bunel, M., and E. Tovar. 2013. Key issues in local job accessibility measurement: Different models mean different results. *Urban Studies*. Prepublished August 8, 2013, doi:10.1177/0042098013495573.
- Cervero, R., T. Rood, and B. Appleyard. 1999. Tracking accessibility: employment and housing opportunities in the San Francisco Bay Area. *Environment and Planning A* 31(7):1259-1278.



- Comber, A. J., C. Brunson, and R. Radburn. 2011. A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. *International Journal of Health Geographics* 10:44. Available at <http://www.ij-healthgeographics.com/content/10/1/44/>
- Delbosc, A. 2012. The role of well-being in transport policy. *Transport Policy* 23:25-33.
- Delbosc, A., G. Currie. 2011. The spatial context of transport disadvantage, social exclusion and well-being. *Journal of Transport Geography* 19(6):1130-1137.
- Dujardin, C., H. Selod, and I. Thomas. 2008. Residential Segregation and Unemployment: The Case of Brussels. *Urban Studies* 45(1):89-113.
- Fotheringham, A. S., C. Brunson, and M. Charlton. 2002. *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester: Wiley.
- Gobillon, L., H. Selod, and Y. Zenou. 2007. The mechanisms of spatial mismatch. *Urban Studies* 44(12):2401-2427.
- Grengs, J. 2012. Equity and the social distribution of job accessibility in Detroit. *Environment and Planning B* 39(5):785-800.
- Hansen, W. G. 1959. How accessibility shapes land use. *Journal of the American Institute of Planners* 25(2):73-76.
- Hess, D B. 2005. Access to employment for adults in poverty in the Buffalo-Niagara Region. *Urban Studies* 42(7):1177-1200.
- Hine, J. (Ed.) 2003. Social exclusion and transport systems (Special Issue). *Transport Policy* 10(4).
- Hurvich, C M., J. S. Simonoff, and C-L Tsai. 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society Series B* 60(2):271-293.
- Kain, J F. 1968. Housing segregation, negro employment and metropolitan decentralization. *Quarterly Journal of Economics* 83(2):175-197.
- Kawabata, M. 2009. Spatiotemporal dimensions of modal accessibility disparity in Boston and San Francisco. *Environment and Planning A* 41(1):183-198.
- Kawabata, M., and Q. Shen. 2006. Job accessibility as an indicator of auto-oriented urban structure: A comparison of Boston and Los Angeles with Tokyo. *Environment and Planning B* 33(1):115-130.
- Korsu, E., and S. Wenglenski. 2010. Job accessibility, residential segregation and risk of long-term unemployment in the Paris Region. *Urban Studies* 47(11):2279-2324.

- Kwan, M-P., and J. Weber. 2008. Scale and accessibility: implications for the analysis of land use-travel interaction. *Applied Geography* 28(2):110-123.
- Kwok, R. C. W., and A. G. O. Yeh. 2004. The use of modal accessibility gap as an indicator for sustainable transport development. *Environment and Planning A* 36(5):921-936.
- Leung, Y. 1987. On the imprecision of boundaries. *Geographical Analysis* 19(2):125-151.
- Lloyd, C., and I. Shuttleworth. 2005. Analysing commuting using local regression techniques: scale, sensitivity, and geographical patterning. *Environment and Planning A* 37(1):81-103.
- Lucas, K. 2012. Transport and social exclusion: Where are we now? *Transport Policy* 20:105-113.
- Lucas, K., and P. Jones. 2012. (Eds.) Social impacts and equity issues in transport (Special Issue). *Journal of Transport Geography* 21.
- Lucas, K., and J. Stanley. 2009. (Eds.) International perspectives on transport and social exclusion (Special Issue). *Transport Policy* 16(3).
- Maroko, A R., J. A. Maantay, N. L. Sohler, K. L. Grady, and P. S. Arno. 2009. The complexities of measuring access to parks and physical activity sites in New York City: a quantitative and qualitative approach. *International Journal of Health Geographics* 8:34. Available at <http://link.springer.com/article/10.1186/1476-072X-8-34>
- Matas, A., J-L. Raymond, and J-L. Roig. 2009. Car ownership and access to jobs in Spain. *Transportation Research A* 43(6):607-617.
- Morency, C., A. Páez, M. J. Roorda, R. Mercado, and S. Farber. 2011. Distance travelled in three Canadian cities: Spatial analysis from the perspective of vulnerable population segments. *Journal of Transport Geography* 19(1):39-50.
- Ong, P. M., and D. Miller. 2005. Spatial and transportation mismatch in Los Angeles. *Journal of Planning Education and Research* 25(1):43-56.
- Ortúzar, J de D., and L. G. Willumsen .2006. *Modelling transport (3rd edition)*. , Chichester: Wiley.
- Páez, A., S. Farber, and D. Wheeler .2011. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A* 43(12): 2992-3010.
- Páez, A., R. Mercado, S. Farber, C. Morency, and M. Roorda. 2010. Relative accessibility deprivation indicators for urban settings: definitions and application to food deserts in Montreal. *Urban Studies* 47(7):1415-1438.
- Preston, J., and F. Rajé. 2007. Accessibility, mobility and transport-related social exclusion. *Journal of Transport Geography* 15(3):151-160.

- Reardon, L., and S. Abdallah. 2013. Well-being and transport: Taking stock and looking forward. *Transport Reviews* 33(6):634-657.
- Roorda, M J., A. Páez, C. Morency, R. Mercado, and S. Farber. 2010. Trip generation of vulnerable populations in three Canadian cities: a spatial ordered probit approach. *Transportation* 37(3):525-548.
- Shen, Q. 1998. Location characteristics of inner-city neighborhoods and employment accessibility of low wage workers. *Environment and Planning B* 25(3):345-365.
- Shuttleworth, I., and M. Gould. 2010. Distance between home and work: a multilevel analysis of individual workers, neighbourhoods, and employment sites in Northern Ireland. *Environment and Planning A* 42(5):1221-1238.
- Stanley, J., and J. Stanley. 2014. Social exclusion and travel. In *Handbook of sustainable travel*, edited by T. Gärling, D. Ettema and M. Friman. Dordrecht: Springer, p.165-184.
- Stoll, M. A., H. J. Holzer, and K. R. Ihlanfeldt. 2000. Within cities and suburbs: Racial residential concentration and the spatial distribution of employment opportunities across sub-metropolitan areas. *Journal of Policy Analysis and Management* 19(2):207-231.
- Van Wee, B., M. Hagoort, and J. A. Annema. 2001. Accessibility measures with competition. *Journal of Transport Geography* 9(3):199-208.
- Wang, J. 2003. Job proximity and accessibility for workers of various wage groups. *Urban Geography* 24(3):253-271.