The definitive, peer-reviewed and edited version of this article is published and can be cited as

Duran-Fernandez, R. and G. Santos (2014), 'Gravity, Distance, and Traffic Flows in Mexico', *Research in Transportation Economics*, Vol. 46, pp. 30-35. DOI: 10.1016/j.retrec.2014.09.003

Gravity, Distance, and Traffic Flows in Mexico

Roberto Duran-Fernandez[1*] and Georgina Santos[2]

[1]Transport Studies Unit, University of Oxford, UK
Tel: +52 (55) 5249 5060
E-mail address: r.duran.fernandez@gmail.com
[*]Corresponding author

[2]School of Planning and Geography, Cardiff University, UK, and Transport Studies Unit, University of Oxford, UK
Tel: +44 (0) 29 208 74462
E-mail address: SantosG@Cardiff.ac.uk

ABSTRACT

This paper presents an econometric analysis that compares the performance of different measures of distance in a gravity model using state data for Mexico. The estimation shows that at this geographic scale, the definition of distance does not affect the explanatory power of the model significantly. However, time-based definitions of distance have a marginal improvement on the model fit in comparison to length-based measures. When geographic specific fixed effects are unknown, the model shows that distance measured as road network distance is a better predictor. The paper concludes that time-based definitions of distance present several advantages in comparison to traditional length-based definitions. However, at large geographic scales, where relative distances between every geographic unit are long, the use of length-based distance instead of time-based distance to approximate travel costs generates similar results.

Key words: gravity model, distance measures, time-based distance, length-based distance, travel costs, Mexico

JEL codes: O18

1 INTRODUCTION

For a long time distance has been recognised as an important variable for explaining social phenomena. Tolber's first law of geography (TFL) captures the importance of distance in social science: *everything is related to everything else but near things are more related than distant things*. Distance had a straightforward meaning 200 years ago when von Thünen proposed one of the first models to explain how the economy organizes itself in space. However, even in a world shrunk by transportation and communication technologies distance is still meaningful. The *Death of Distance* argument may be fallacious because it assumes that communication and transport technologies are pure substitutes (Miller 2004) when in fact the raise of demand for telecommunication services has typically been preceded by an increase in travel demand at all geographic scales (Couclelis 2000).

TFL is still an important concept. It proposes nearness as a key determinant of social interactions among individuals. Nearness is usually defined as the shortest path between any two agents on the surface of the Earth. However, this definition is not useful for phenomena that do not follow pure Euclidean relations. Miller (2004) suggests that social interactions that do not appear to be consistent with TFL may be conditioned by geographic factors such as terrain, land cover, infrastructure, and traffic congestion. To model these attributes, the concept of distance has to be generalised to least-cost paths through geographic space (Angel and Hyman 1976).

In the literature, empirical measures of nearness are generally based on the straight-line segment connecting two locations; however, this is not the only measure that satisfies the metric space conditions (Miller 2004). Among the available measurements of distance, we can find the minimum distance or time through the transport network, and pecuniary costs.

This paper, presents an analysis that compares the performance of different measures of distance in a gravity model using state data in Mexico. The results show that at this geographical scale, time-based measures report marginal improvements in the explanatory power of the tested models. However, the different definitions of distance do not affect the overall accuracy of the model significantly. When geographic specific fixed effects are unknown, the distance measured through the transport network is a better predictor. Finally, the analysis suggests that at a large geographical scale, length, and time based measures' of distances generate similar results.

The paper is organised as follows: Section 2 presents the model; Section 3 describes the data. Sections 4 and 5 show the results of the estimation and discuss the forecasting power of the model, respectively. Section 6 compares row normalised weight matrices using different definitions of distance. Finally, Section 7 contains final remarks.

2 THE MODEL

The gravity equation is an analytic tool widely used for modelling bilateral flows between different geographic entities. The model resembles Newton's gravitation law and has been applied to the analysis of a large number of socio-economic interactions such as migration, international trade and price convergence. The gravity model is particularly important for transport modelling since it is the basis for the trip distribution estimation in the four-step transport forecasting model.

Equation 1 presents the general gravity model, where $T_{i,j}$ is the flow from origin i to destination j, K is a constant, M_i and M_j are the relevant economic variables of the two locations (also known as economic masses for its resemblance with mass in Newton's Gravity law), and $f(\cdot)$ is a function that depends negatively on transport costs $d_{i,j}$ between the two locations. In the original formulation $d_{i,j}$ is raised to a power θ, where $\theta < 0$.

Equation 1 $$T_{i,j} = K\ M_i M_j f(d_{i,j} \mid \theta)$$

In transport planning, the gravity equation can be derived as the solution of the doubly constrained entropic and gravity type model (Wilson 1970). Under this interpretation, the gravity equation is a solution to the Hitchcock problem. This problem considers the cost minimisation of commodities distribution given certain destinations and production sites. The gravity equation can be interpreted as a representation of supply and demand forces where distance acts as a break that imposes a lower trade flow in equilibrium. The model emerges as an *ad hoc* formulation of supply and demand forces; however, more recently some authors have derived it formally from micro economic foundations (Head 2003).

The estimation presented in Equation 1 is straightforward using different econometric techniques. It can be transformed into a linear model taking logarithms on both sides of the equation. However, Anderson and van Wincoop (2003) argue that this specification is not correct, since it does not take into account multilateral resistance terms. The solution proposed by the authors is to consider explicitly importer and exporter fixed effects. Another modification is the inclusion of a *remoteness* index that measures the average distance of a region from all trading partners. The final specification is presented in Equation 2, where fe and r represent a fixed effect and the remoteness index, respectively, and ε is a stochastic variable.

Equation 2 $$\ln(T_{i,j}) = \ln(M_i) + \ln(M_j) - \theta \ln(d_{i,j}) + \beta_1 fe_i + \beta_2 fe_j + \beta_3 r_{i,j} + K + \varepsilon_{i,j}$$

The empirical literature has used different socio-economic definitions as economic variables, *M*, such as gross production, gross value added, population, and workforce, among others. On the other hand, costs are usually measured as the physical distance between locations *i* and *j*. Distance is usually measured through the *great circle* formula, which assesses the minimum distance between any two points on the Earth surface assuming that its shape is a perfect sphere. A clear limitation of the great circle approach is that it does not necessarily reflect the real freight routes used in trade. It can also lead to biases when average speed is not uniform across all routes.

The empirical literature from international trade presents several examples for the estimation of gravity models. In general, the objective of these studies is the analysis of the determinants of trade, so that distance plays only a secondary role as a control in the estimation. In general, the mean elasticity of distance with respect to trade flows has been estimated at 0.9 with a 90 percent of estimates lying between -0.28 and -1.55. These results have been estimated using meta-analysis techniques on an exhaustive survey of existing literature (Disdier and Head 2006). McCallum (1995) estimates this elasticity for Canada at -1.52 and Wolf (2000) estimates it for the US at -0.77, using provincial and state trade flows, respectively. For Mexico, these calculations can only be made using data on international trade flows. The estimated elasticities lie between -0.9 and -1.4 (Lopez-Cordova 2002, Montenegro and Soloaga 2006, and Soloaga *et. al.* 1996).

3 DATA

In this section, we present the results of the estimation of an inter-state gravity model using data from the Origin and Destination of Passengers and Freight Survey in Mexico (Instituto Nacional de Estadística, Geografía e Informática, INEGI 1999). The dataset is a module of the National Economic Census of the National Institute for Statistics, Geography, and Informatics. The database contains for each of the 31 states and the Federal District the lorry freight flow for their main 11 national destinations. The rest of the freight flow is aggregated into a single category labelled *others*. On average, this flow represents only 14 percent of state freight. For each state, the freight label *others* is equally allocated among the rest of the states. Freight flows S are used to build an interstate origin-destination (OD) matrix. For each possible origin and destination combination $\{i, j\}$ we define total freight flow T as $T_{i,j}= S_{i,j} + S_{j,i}$. The OD matrix is symmetric under this definition, therefore the analysis uses only the entries of the upper triangle to avoid including repeated observations.

In order to assess the robustness of the model, different estimations were carried out using each of the following economic variables: total population, workforce (total, industrial, services, and both), and gross state product GSP (total, industrial, services, and both). The estimation of the model shows that the estimated elasticity of distance with respect to freight flow is remarkably stable independently of the chosen economic variables.

We consider five different measures of distance. The first three ones (A, B, and C) reflect the optimal road network path between any two points in the country. The last two measures (D and E) are estimated using the great circle formula.

A. **Network Time (Unrestricted)**: Transit time, expressed in hours, along the route that minimises time. Minimum time for interstate routes is equal to the average minimum time of intercity routes.
B. **Network Time (Restricted)**: Transit time, expressed in hours, through the route that minimises time, with a mandatory rest time after 11 continuous hours of service. Minimum time for interstate routes is equal to the average minimum time of intercity routes.
C. **Network length**: Length in kilometres associated to the route that minimises time. Minimum time for interstate routes is equal to the average minimum time of intercity routes.

D. **Great circle length (City Average)**: Average distance in kilometres between each metropolitan area in a given state. As explained above, this measure is not in any way related to any network links.

E. **Great circle length (Geographic Centroid***)*: Distance in kilometres between the geographic centroid of each state. Again, this measure is not related to the road network.

Network distances are estimated using a GIS model of the National Road Network based on INEGI's Topographic Digital Dataset. The construction of the model is presented in Duran-Fernandez and Santos (2014). We use the Network Analyst utility of ArcMap 9.1 to estimate the distance between each of the 69 standard metropolitan areas in Mexico as defined in Duran-Fernandez (2007). The algorithm used for this purpose searches the minimum cost between any two points, taking as an impedance variable the transit time (in hours) for each section on the network. The programme also assesses the length (in kilometres) for each optimal route. We estimate an additional measure based on time. We assume a mandatory rest-period of 13 hours after 11 continuous hours of travel. Finally, the distance between states i and j is estimated as the weighted average of the distances between all the cities in those states. This exercise is performed for the estimated time and length measures and uses metropolitan population as weight.

We estimate two additional measures using the great circle approach. First, we calculate the distance between every standard metropolitan area. The distance between any two states is equal to the weighted average distance of their cities. The average also uses metropolitan population as a weight. The second measure is the great circle distance between the geographic centroid of each state. This is the standard approach used in trade literature. Table 1 presents the correlation matrix between the different definitions of distance.

Table 1 Correlation matrix of distance measures

	A	B	C	D	E
A. Network time (U)	1.000	0.976	0.88	0.84	0.874
B. Network time (R)	0.976	1.000	0.871	0.838	0.877
C. Network distance	0.880	0.871	1.000	0.962	0.956
D. Great circle length (CA)	0.840	0.838	0.962	1.000	0.977
E. Great circle length (GC)	0.874	0.877	0.956	0.977	1.000

U: Unrestricted; R: Restricted

The remoteness index is calculated following Equation 3. N_j is the population of state j, $d_{i,j}$ is the distance between state i and state j, and A is a standardisation. A different index is estimated for each of the five distance measures.

Equation 3

$$r_i = \frac{1}{A} \sum_j \frac{N_j}{d_{i,j}}$$

The model is estimated using ordinary least squares (OLS) with robust standard errors. Endogeneity between the economic variables and trade flows has been a concern in the literature. In order to avoid this problem, the model is estimated using 2 stage-least squares (2SLS) using the state's population in 1940as instrumental variable.

4 ESTIMATION

The model as presented in Equation 2 is estimated using different economic variables (population, workface, GSP). The estimated elasticity of traffic flow with respect to distance is remarkably stable across the different definitions for economic variables, both under the OLS and the 2SLS estimation. The results suggest that state fixed effects work as a control for any omitted variable, generating an unbiased estimation of the parameters of the model. Due to the similarity of these results, we only present the estimation that uses workforce in the industrial and service sector as economic variable in Table 2.

Table 2 Estimation of the gravity model
(Standard errors in parenthesis)

	$\ln(x_i)+\ln(x_j)$	$\ln(d_{i,j})$	r_i	r_j	Constant	R^2	Mean sq. error
A. Network time (Unrestricted)							
OLS	1.151	-0.731	-0.494	-0.207	-21.909	0.8096	0.8689
	(0.17)	*(0.11)*	*(0.07)*	*(0.06)*	*(4.07)*		
2SLS/[a]	1.151	-0.731	-0.494	-0.207	-21.909	0.8096	0.8685
	(0.17)	*(0.11)*	*(0.07)*	*(0.06)*	*(4.07)*		
B. Network time (Restricted)							
OLS	1.166	-0.543	-0.441	-0.178	-22.744	0.8092	0.8698
	(0.16)	*(0.09)*	*(0.06)*	*(0.05)*	*(4.03)*		
2SLS/[a]	1.166	-0.543	-0.441	-0.178	-22.744	0.8092	0.8698
	(0.16)	*(0.09)*	*(0.06)*	*(0.05)*	*(4.03)*		
C. Network length (City average)							
OLS	1.151	-0.658	-0.148	-0.188	-20.270	0.8072	0.8744
	(0.17)	*(0.11)*	*(0.05)*	*(0.05)*	*(4.66)*		
2SLS/[a]	1.151	-0.658	-0.148	-0.188	-20.270	0.8072	0.8744
	(0.17)	*(0.11)*	*(0.05)*	*(0.05)*	*(4.66)*		
D. Great circle length (City average)							
OLS	1.148	-0.596	-0.282	-0.166	-20.862	0.8038	0.8820
	(0.17)	*(0.11)*	*(0.05)*	*(0.05)*	*(4.69)*		
2SLS/[a]	1.148	-0.596	-0.282	-0.166	-20.862	0.8038	0.8820
	(0.17)	*(0.11)*	*(0.05)*	*(0.05)*	*(4.69)*		
E. Great circle length (Geographic centroid)							
OLS	-0.872	-0.596	0.679	0.269	26.335	0.8065	0.8760
	(0.32)	*(0.10)*	*(0.13)*	*(0.08)*	*(7.58)*		
2SLS/[a]	-0.872	-0.596	0.679	0.269	26.335	0.8065	0.8760
	(0.32)	*(0.10)*	*(0.13)*	*(0.08)*	*(7.58)*		

/[a] IV Population 1940
x_k: Workforce in the secondary and tertiary sector

All the coefficients of the regression in Table 2 are statistically significant at a confidence level of 1 percent. They also present the expected sign and value with the only exception of regression E. According to the theoretical formulation that derives the gravity equation from micro foundations, the value of the coefficient of the economic variable must be equal to one. This hypothesis is not rejected at a confidence level of 1 percent for regressions A to D. The effect of remoteness is also negative and highly significant for these equations.

The effect of the instrument in the first stage regression is positive and significant. The values of the instrumented variable are very close to the originals. Therefore, for all the regressions the results of the OLS and the 2SLS are virtually identical suggesting that endogeneity is not a problem in the estimation.

The anomalous behaviour of regression E can be attributed to the poor explanatory variable of the remoteness index associated to the distance measure used in this regression. The similarity of the elasticity of traffic flow with respect to distance in regression D can be explained by two factors: first, the correlation between the two grand circle distances is close to one, and second, remoteness and distance in regression E are poorly correlated (-0.3).

The explanatory power measured through the R^2 is almost identical for all the models, with a marginal increment for the time-based regression. The most important difference among the regressions in Table 2 is the estimated elasticity of traffic flow with respect to distance. We introduce the concept of average speed in a given route as the ratio of the route's length and its transit time. Since elasticity is a non-dimensional measure, its value should be the same regardless of whether it is calculated using distance measured in length or time. This property is satisfied only if average speed is the same across the network (the speed effect is captured by the constant) or is the same across the state network (the speed effect is captured by the state fixed effects).

Given the fact that the coefficients are different (*ie.* speed is not the same across the network) we can conclude the following. First, the omitted speed effect is correlated with the route's length. This is not surprising, especially for regression C. The definition of distance used in this regression is the only one that estimates the routes' length as an implicit function of network speed. Despite this property, the difference between the elasticity of regression C and the time-based regression A, is lower than the one between great circle-based regressions (E and D) and A. This implies that the linear correlation between the omitted speed effect and length is higher in the great circle-based regression (even though the omitted speed effect and length do not depend implicitly on average route's speed).

The result strongly suggests that the omitted fixed effect and the length of the route that minimises time between any two nodes on the network do not follow a linear relationship. This possibility has an important implication. Let l_k be average length and v_k be average speed in section k of any route. Variables t and l are defined as average transit time and length of the whole route. V is the implicit average speed along the route. Given these definitions, Equation 4 presents a non-linear relationship between the route's average speed and its length.

Equation 4

$$\ln(t) \equiv \ln\left(\sum_k \alpha \frac{l_k}{v_k}\right)$$

$$\ln(l) \equiv \sum_k \alpha \ln(l_k)$$

$$V \equiv \sum_k \alpha \ln(v_k)$$

For this case Jensen's inequality implies that $\ln(t) \geq \ln(l) - V$. Defining $\varepsilon_z = y - x\beta - \theta \ln(z) - C$, where $z=\{l,t\}$, C is a set of constants such that $C=\{C_1, C_2+V\}$ and ε_z is an error term, it follows that $(\varepsilon_d)^2 > (\varepsilon_t)^2$. Therefore, the mean square error of a time-based estimation is lower than the length based model. The result is compatible with the empirical estimation as shown in Table 2.

The second observation is that the scale of the speed effect is countrywide (*ie.* across states). Average speed depends on characteristics of the road network such as its structure, internal links, and road quality. The primary trunklines, as well as a large number of secondary roads of the National Road Network were built and are currently maintained by the federal government. Therefore, it is natural to assume that the network determinants that influence average speed are not state-specific. If this were the case, the effect of the omitted average speed would be captured by the state fixed effect in the length-based regressions (C, D, and E). This would generate similar elasticities in both the time-based and length-based regressions. An important implication of this observation is that the main features of the road network are not related to particular characteristics of the states. It is also worth mentioning that this excludes the possibility of variations in road network characteristics at local level. This is because average variations in the speed network at local level for each state must also be captured by the fixed effect.

5 FORECASTING

Under similar conditions to those of the year for which the model was estimated, the coefficients allow to forecast, with a reasonably error margin, statewide freight flow. These forecasts enable us to estimate an OD matrix, given an economic variable, a distance, a remoteness index, and a state fixed value. We estimate the square of the difference between the observed and predicted values of total freight flow $T_{i,j}$ at state level. This aggregated squared sum of errors at state level (SLSSE) is similar for all the estimated models. The worst performance is that of regression E, which uses the great circle length based on geographic centroids. None of these quantities exhibit spatial autocorrelation. Results are presented in Table 3.

Table 3 SSE aggregated at state level
(Variance in parenthesis)

	With fixed effects		No fixed effects		
	SLSSE	Moran's I	SLSSE	Moran's I	
A. Network Time (U)	9,692	0.032	41,211	0.094	*
		(0.00)		*(0.00)*	
B. Network Time (R)	10,066	0.033	39,337	0.070	*
		(0.00)		*(0.00)*	
C. Network Length	10,515	0.031	20,574	-0.054	
		(0.00)		*(0.00)*	
D. Great Circle Length (CA)	9,670	0.032	26,537	-0.022	
		(0.00)		*(0.00)*	
E. Great Circle Length (GC)	11,399	0.020	107,312	-0.063	
		(0.00)		*(0.00)*	

*Significant with a confidence of 1%
OLS regressions

Freight flow data is only available at state level. The estimated model can be used to estimate freight flows at a lower geographical scale, such as for example, regions, and metropolitan areas. This exercise may generate information that is not available by other means. However, the use of the model would require to allocate to each observation a fixed effect. This would be particularly problematic for the case of regions that overlap one or more states. Therefore, in practice this kind of estimation would omit at the fixed effects for each variable.

To asses the accuracy of the model under these circumstances, we estimate the SLSSE omitting the fixed effects. We find that, the worst model by far is regression E, which is based on great circle-length using geographic centroids. The best performance is shown by length-based regressions C and D. In this exercise, spatial autocorrelation is positive and significant at 1 percent level for the SLSSE of the two time-based regressions. Despite exhibiting a low value, these results suggest the forecast is geographically biased (*ie.* it is not randomly distributed across states). Table 3 presents these results.

The geographic bias of the forecast without fixed effects should be related to the variables of the gravity model as long as these variables are correlated to state fixed effects. To investigate this relationship we regress the SLSSE of the forecast without fixed effects on remoteness and economic mass. The results of the regressions indicate that only remoteness is a statistically significant determinant (Table 4). Under these results, we can conclude that forecasts omitting fixed effects will be more biased for remote states. This characteristic leads to the positive spatial autocorrelation presented above, since remote states are geographically clustered.

Table 4 Determinants of SSE aggregated by State

(Standard Error in Parenthesis)

	r	$\ln(x_i)+\ln(x_j)$		$R^{2\,/a}$
Network Time				
A. Unrestricted	-14.95	6.02	ns	0.51
	(2.71)	*(5.78)*		*(0.48)*
B. Restricted	-12.11	6.22	ns	0.49
	(2.32)	*(5.80)*		*(0.45)*
Distance				
C. Network	-4.16	6.83	ns	0.12
	(2.39)	*(5.52)*		*(0.06)*
D. Great Circle (CA)	-8.49	6.35	ns	0.29
	(2.50)	*(5.62)*		*(0.24)*
E. Great Circle (GC)	22.55	-56.22		0.78
	(3.19)	*(6.43)*		*(0.76)*

[/a] Adjusted R^2 in parenthesis

ns: Not significant at 1 percent level.

r: remoteness

x_k: logarithm of workforce in the secondary and tertiary sector

6 DISTANCE BEYOND GRAVITY

Spatial econometrics is a field where the application of empirical measurements of distance is highly relevant. In the simplest spatial autoregressive model y_i depends on y_{-i} weighted by a spatial weight $w_{i,j}$. Several definitions of spatial weights have been proposed in the literature, such as distance between locations, lengths of common boundaries, contiguity criterions, and several non-Euclidean measures.

In this section, we use the five definitions of distance presented in Section 3 to estimate a row normalised weight distance matrix at state level. We define $w_{M,i,j}$ as the i,j component of the weight matrix W_M, built using the definition of distance $M=\{A,B,C,D,E\}$. To compare how similar any two matrices are we define $q_{[M,N],i,j}=(w_{M,i,j}-w_{N,i,j})^2$. Table 5 presents the estimated values for $Q_{[M,N]}={}^1/_{IJ}\Sigma_i\Sigma_j\, q_{[M,N],i,j}$ for all the definitions of distances. Since $q_{[M,N],i,j}\cong 0$ if $W_M\cong W_N$, values of $Q_{[M,N]}$ close to 0 imply that definitions of distances M and N generate similar row normalised weight distance matrices. The results show that the unrestricted time-based weight matrices are significantly different to the length-based matrices (definition A in comparison to C, D, and E). The restricted time-based matrix (B) presents the largest differences with respect to the rest of the definitions. Length-based matrices C and D generate the most similar weight matrices; however, the difference between them is statistically significant. Finally, the matrix that uses the great circle distance based on geographic centres (E) exhibits significant differences with respect to the other length-based matrices.

These differences are not symmetric across states. We estimate $Q'_{[A,N],i}={}^1/_J\Sigma_j\, a_{[A,N],i,j}$ for $N=\{C,D,E\}$ to compare time-based weight matrix A with the length-based matrices row by row, where each row represents the values of a state. The value of Q' is significantly lower for states located in the extremes of the country and larger for the central states. This pattern is reflected in a high degree of spatial autocorrelation measures through Moran's *I*. The distance between the states located in the extremes and the rest of the country presents the largest values. This is true for both time-based and length-based distances. The results imply that changes in the definition of distance do not affect this kind of links. An important implication is that long distance measurements are not sensitive to distance definitions; however, for the case of short distances variations can be significant.

Table 5 Average value of the square of the difference of row normalised weight matrices under different definitions of distance

(Standard error in parenthesis)

	A.	B.	C.	D.	E.
A. Network Time (U)	n.a.	0.69	0.56	0.59	0.50
		(15.1)	*(5.8)*	*(6.7)*	*(8.8)*
B. Network Time (R)	0.69	n.a.	1.57	1.58	1.54
	(15.1)		*(6.8)*	*(7.2)*	*(9.1)*
C. Network Distance	0.56	1.57	n.a.	0.06	0.21
	(5.8)	*(6.8)*		*(12.7)*	*(13.6)*
D. Great Circle Distance (CA)	0.59	1.58	0.06	n.a.	0.16
	(6.7)	*(7.2)*	*(12.7)*		*(11.4)*
E. Great Circle Distance (GC)	0.50	1.54	0.21	0.16	n.a.
	(8.8)	*(9.1)*	*(13.6)*	*(11.4)*	

7 FINAL REMARKS

This paper presents an econometric analysis based on a gravity model to assess the performance of different empirical measures of distance at state level in Mexico. The estimation shows that at this scale the definition of distance does not affect the explanatory power of the model significantly. However, time-based definitions of distance have a marginal enhancement on the model fit. When geographic fixed effects are unknown, traditional length-based measures of distance perform poorly.

The estimated elasticities of traffic flow with respect to distance are very sensitive to the definition that is used. This behaviour implies that the features of the road network that determine average speed on a particular route are not determined homogenously at state level.

Finally, a comparison of row normalised distance weight matrices shows that time and length-based estimations are statistically different. However, the divergences in the matrix' rows are lower for long distances than for short ones.

Time-based definitions of distance present several advantages in comparison to traditional length-based definitions. In particular, among the length-based definitions, the commonly used great circle distance between geographical centroids is the one that presents the poorest performance. At state level, both definitions generate similar results; however, this may not be the case at a lower scale. Another implication is that at large geographic scale, where relative distances between every geographic unit are large, the difference between time and length based distances tends to be lower. This result validates the use of traditional great circle distances in contexts such as the international trade literature.

Acknowledgements and disclaimer

The authors are grateful to two reviewers and to the attendees to the 40th Annual Universities' Transport Study Group Conference, Southampton, January 3-5 2008, where an earlier version of this paper was presented. This study was financed by the Mexican Federal Government through the National Council of Science and Technology (CONACYT). Any opinions, findings, conclusions and recommendations expressed in this paper are those of the author alone and should not be attributed to any other person or entity.

REFERENCES

Anderson, J. and E. van Wincooop (2003) "Gravity with gravitas: A solution to the border puzzle" American Economic Review (93:1) pp. 170-192.

Angel, S., and Hyman, G.M. (1976) "Urban Fields: A Geometry of Movement for Regional Science". Pion, London.

Couclelis, H. (2000) "From Sustainable Transportation to Sustainable Accessibility: Can we Afford a New Tragedy of the Commons?" Information, place, and cyberspace: Issues in accessibility. Springer Berlin, Berlin, Germany

Disdier, A.C. and K. Head (2006) "The puzzling persistence of the distance effect on bilateral trade" WP No. 186, Centro Studi Luca d'Aglian, University of Milan, Italy.

Duran Fernandez, R. and G. Santos (2014) "A GIS Model of the National Road Network in Mexico" *Research in Transportation Economics, **THIS** ISSUE.*

Duran Fernandez, R. (2007) "A Benchmark for a Metropolitan Geo-Statistical Framework in Mexico" WP No. 1022 Transport Studies Unit, Oxford, UK. http://www.tsu.ox.ac.uk/pubs/1022-duran-fernandez.pdf

Head, K. (2003) “Gravity for Beginners. Faculty of Commerce” University of British Columbia,.

INEGI (1999) "Censo Económico 1999: Origen y destino de los pasajeros y la carga transportada." Publication from the Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.

Lopez-Cordova, E. (2002) "NAFTA and Mexico's Manufacturing Productivity: An Empirical Investigation Using Micro-Level Data" Mimeo. Inter-American Development Bank, DC, USA.

McCallum, J. (1995) "National borders matters: Canada-US regional trade patters" American Economic Review (85:3) pp. 615-623.

Miller, H. (2004) “Tobler’s First Law and Spatial Analysis” Annals of the Associations of American Geographers (94:2) pp. 284-289.

Montenegro, C.E. and I. Soloaga (2006) "NAFTA’s trade effects: New evidence with a gravity model" Estudios de Economía (33:1) pp. 45-63.

Soloaga, I. Wilson, J.S. and A. Mejia (1996) "Moving forward faster: Trade facilitation reform and Mexican competitiveness" WP No. 3953, World Bank Policy Research, World Bank, DC, USA.

Wilson, A. G. (1970) “Entropy in Urban and Regional Modeling” Pion, London.

Wolf, H.C. (2000) "International home bias in trade" The Reviews of Economics and Statistics (82:4) pp. 555-563.