Towards a direct demand modeling approach

Georgios Sarlas, IVT, ETH – Zürich
Kay W. Axhausen, IVT, ETH – Zürich

Conference paper STRC 2014

STRC | 14th Swiss Transport Research Conference
Monte Verità / Ascona, May 14-16,2014

Towards a direct demand modeling approach

Georgios Sarlas
IVT, ETH Zürich
CH 8093, Zürich

Phone: +41 44 633 37 93
Fax: +41-44-633 10 57
email: georgios.sarlas@ivt.baug.ethz.ch

Kay W. Axhausen
IVT, ETH Zürich
CH 8093, Zürich

Phone: +41-44-633 39 43
Fax: +41-44-633 10 57
email: axhausen@ivt.baug.ethz.ch

May 2014

Abstract

The aim of this paper is to employ spatial regression modelling as a form of direct demand modelling where the speed of each link is the dependent variable of interest. The variables that their inclusion in the regression model is investigated, correspond to only aggregated values where no personal data information can be traced back. More specifically, socio-demographic variables along with variables that represent the network characteristics are taken into account for that purpose. A particular focus is given on the identification and the construction of the spatial weighting matrices. Three different spatial autoregressive models are estimated and compared to the ordinary linear regression to highlight their capability of explaining transport related phenomena.

Keywords

Spatial regression – demand modelling –speed regression

1. Introduction

Travel demand models have increased their data demands massively both in scope and scale, and in addition their complexity has increased in a similar way. The obvious reluctance of the practice to adopt such advanced models, raises the concern that the gap between academia and practice has become wider than ever. On the one hand, the increased data collection abilities of the field, along with the expected wave of "big data" might allow the (academic) field to continue on its current trajectory, but on the other hand the use and abuse of big data raises the danger of a sudden change in the course of public policy and the sudden lack of high quality alternatives to the existing state-of-the-art (i.e. academic) models. At this point, it is tempting to contradict this trend and explore the formulation of an alternative direct travel demand model structure that requires only aggregate and anonymous data. Nevertheless, the model structure should be able to make statements about the speed and the traffic volume on a link level, that constitute the minimum requirements for the transport project appraisal. The alternative that is checked is spatial econometrics techniques, and more specifically spatial regression.

Spatial econometrics was popularized by Anselin (1988), defined as the *domain that deals with the peculiarities caused by space in the statistical analysis of regional science models*. More specifically, these peculiarities are caused by the dependence and the heterogeneity of data in space (spatial effects). *As spatial dependence, it can be considered to be the existence of a functional relationship between what happens at one point in space and what happens elsewhere. Spatial heterogeneity is considered to be the lack of structural stability of the various phenomena over space, and also the lack of homogeneity of the spatial units of the observations.* (Anselin, 1988)

A number of applications of spatial regression models can be found in the urban and modelling area. A comprehensive review of the application of such models is presented by Paez and Scott (2004). The presence of spatial effects constitutes a dimension which normally is neglected in the existing transport modelling approaches. There is a relatively limited number of applications employing spatial regression models for the explanation of how transport related phenomena, such as speed or flows, occur and evolve over the space. The correlation of speed observations was demonstrated by Bernard et al. (2006) and pointed out the necessity of accounting for spatial dependency when it comes to the estimation of speed or flows. Hackney et al. (2007) demonstrated the plausibility of accounting for the spatial dependence in the estimation of speed where three spatial autoregressive models were estimated and compared. Cheng et al. (2011) examined the spatio-temporal dependence structure of road networks.

In this paper, the first steps towards a simplified direct demand modelling approach is presented along with the relevant theoretical background of spatial regression that is the employed modelling approach. In particular, three different spatial simultaneous autoregressive (SAR) models are estimated and compared to an ordinary linear regression in order to highlight and evaluate the impact of utilizing spatial regression models as a simplified direct travel demand model approach. The SAR models are constructed to offer a structural explanation of the speed on the links, and subsequently their predictive power is assessed to draw conclusions regarding the plausibility and the effectiveness of the approach.

2. Spatial regression models

As mentioned above, the alternative of spatial econometrics techniques constitutes the option that is examined to be employed in an coherent framework of a direct demand modelling approach. Conceptually, it is arguable that a simplified approach cannot exhibit the predictive accuracy and the sensitivity of the existing approaches (4-step and agent-based models), however it cannot be overseen the fact that when it comes to the appraisal of public transport projects, as Flyvbjerg et al. (2005) argue, the quality of the demand forecasts has not been improved over the years even though more complex and advanced models have been employed. Driven by this, the option of spatial regression as a simplified demand modelling approach is examined in the context of transport project appraisal. The advantage of that choice in comparison to the "classical" approaches is that it is significantly less cumbersome to apply, less data demanding, and also offers a structural explanation of the observed transport phenomena such as speed and flows in a direct way. The underlying assumption and hypothesis is that by accounting properly for the impact of spatial effects in the context of regression modelling, accurate demand forecasts can be provided. Nevertheless, the value of the research towards this direction is not only as a competing alternative of the existing methods but it can also point directions regarding the importance of accounting properly for the spatial effects and thus can provide insight on how the existing approaches can be improved.

As spatial regression models is defined the use of regression models by accounting for the impact of spatial effects in their specification and estimation, avoiding to give rise to statistical problems such as unreliable statistical tests and biased and inconsistent estimated parameters. This is accomplished by incorporating in the model the information about the spatial structure of the data, in the form of a contiguity matrix. Spatial simultaneous autoregressive (SAR) models is a popular category of such models that they have been applied in many cases. As suggested by Ord (1975), their estimation can be conducted by means of maximum likelihood since the ordinary least square (OLS) estimation produces inconsistent estimates. The assumption of these models is that the response variable at each location is a combination of the explanatory variables at that location but also of the response of neighbouring locations (Löchl and Axhausen, 2010).

Three main types of SAR models can be found in the literature, each one having different characteristics based on their underlying assumptions about where the autoregressive occurs (Kissling and Carl (2007), LeSage and Pace (2004)). At first, the spatial error autoregressive model (SARerr) assumes that the spatial dependence is in the error term of the model, and thus the spatial autoregressive process is applied to it. The formulation of the model is:

$$Y = \beta X + u \ (1)$$

$$with \ \ u = \lambda W u + \varepsilon \ (2)$$

where Y is a vector with N values of the dependent variable, β is a vector with the regression coefficients, X is a matrix with the independent variables, u the error term, λ the spatial autoregressive coefficient, W a matrix with the contiguity structure having dimensions N x N, and ε a vector of independent and identically distributed (iid) error terms.

The spatial lag autoregressive model (SARlag) assumes that the spatial dependence exists in the response variable, and applies the spatial autoregressive process to the response variable and treats it as a lagged variable. The formulation of the model is:

$$Y = \rho W Y + \beta X + \varepsilon \ (3)$$

where ρ is the spatial autocorrelation parameter, and WY is the term for the lagged variable.

The spatial mixed autoregressive model (SARmix, also denoted as spatial Durbin model in some application (eg LeSage and Pace (2004)) assumes that the spatial dependence exists in both the response and the independent variables. The formulation of the model is:

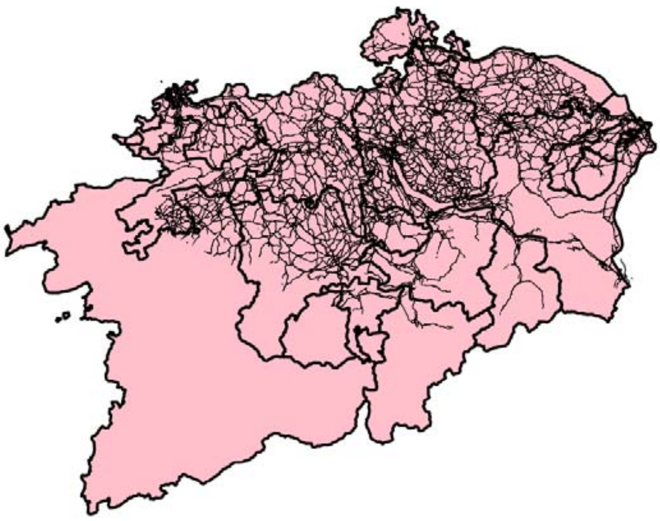
$$Y = \rho W Y + \beta X + W X \gamma + \varepsilon \ (4)$$

$$with \ \ \gamma = -\rho \beta (5)$$

3. Case study

In order to assess the plausibility of applying SAR models for speed prediction purposes, a case study scheme is established. A part of the national network of Switzerland is selected, including the canton of Zurich and extending to the neighboring cantons as well. In particular, the full road network of the North-East Switzerland is included in the chosen network. A navigational network is used, commercially available by Tom-Tom, including hourly speed estimations based on GPS measurements for the majority of the links. In detail, the study network includes approximately *190.000* links (having excluded the secondary, or less important links) while the remaining links are classified based on their type (5 available types). In addition to the estimated speeds, the set speed limit is available for each link along with a dummy variable denoting if it is an on/off ramp. A visual representation of the study network can be seen in Figure 1.

Figure 1: Case study network with administrative borders of cantons



The average speed for the morning peak hour (8-9pm) of a typical weekday is the dependent variable of interest for the regression. This choice is made in order to ensure that there is sufficient variation of the speed values on the links, in comparison to their reported free flow speeds. The regression yields two speed components; first, the average road speed which is a function of the speed limit and the link type and it is a non-spatial quantity. Spatial variation is added to the link speed estimates in the second component via the spatially resolved explanatory variables. Spatially resolved road and public transport network densities represent the effect of road supply on speed. Spatial data on population and employment densities are taken to be indicative of the intensity of local activities, reflecting travel demand locally (Hackney et. al, 2007).

3.1 Spatially resolved variables

Apart from the network data that presented above, the spatial resolved variables constitute an important component of the regression model since they introduce variation on the estimated average values, as resulted from the non-spatial component. At first, the road and public transport densities are of apparent interest since they represent the effect of road supply and also the spatial competence between the private and public modes, especially in the urban areas, on speed. The road density is estimated as the total length of links within a given area and it is calculated for different radii. The full navigational network is used for that density calculation. Besides the full network, the densities of ramp links is calculated as well as it is expected to have local impact on speed. In the case of accounting for the impact of the public transport network on speed, it is less straightforward the way that an appropriate variable can be constructed. As an approximation, the density of public transport stops within a given area, is considered to be the most appropriate variable for that purpose.

Another source of spatially resolved variables corresponds to the impact of socio-demographic data on speeds. More specifically, the socio-demographic data of interest are the population and the employment positions for the whole area of Switzerland, aggregated per hectare, and they are available from the Swiss Federal Statistical Office (BFS: Bundesamt für Statistik).. The population data correspond in the year 2011, taken from the "Statistics of Population and Households 2011" ("Statistik der Bevölkerung und der Haushalte 2011", date of version: 30 August 2012), while the employment data are taken from the "Federal Business Census" of 2008 (Eidgenössische Betriebszählung 2008, date of version: 29 March 2008). Given the disaggregate level of these data (hectar based), they are taken into account as densities over different radii. In addition to the normal densities, kernel densities are calculated as well to account for the diminishing impact of the socio-demographic data over the space.

At last, the spatially resolved variables need to be associated to the links of the network. Thereupon, each link of the network associate with the hectare (cell) values of each spatial variable, closest to the upstream endpoint of the link.

4. Estimation of models

In this section the different regression models estimations are presented and compared to exhibit the impact of accounting properly for the spatial dependency of speeds. More specifically, a standard linear regression model is estimated in terms of ordinary least squares (OLS), while three SAR models are estimated as well. A comparison of the estimated models is conducted in order to shed some light on the plausibility of the SAR models to predict traffic related variables, as speed, and also to what extent they can accomplish that. At first, OLS

4.1 Linear regression model

At first, a linear regression model is estimated (ordinary least squares estimate) to serve as the basis for testing the necessity of accounting properly for the spatial association (auto-correlation). It is expected that the OLS model is going to give rise to biased and inconsistent estimates and thus the resulted adjusted coefficient of determination will be inconsistent and not true. Additionally, OLS predicted values are going to be used for testing if spatial association exists in the residuals by estimating Moran's I measure. Depending on the results of the Moran's I, a justified explanation of whether or not it is needed to account for the spatial dependence properly is going to be provided. The independent variables that are included in the model are determined based on their predictive power and in accordance to the appropriate statistical tests , avoiding to give rise to multicollinearity issues. The specification of the model and the estimated coefficients are presented in Table 1.

As it can be seen, the adjusted R square is extremely high while the estimated parameters are all statistically significant. Employment positions and population densities are not used at the same time due to high correlation. Notably, a differentiation of the employed densities radius for different links' types is found to be more appropriate and thus chosen, instead of a fixed radius density for all links' types. This finding reflects that depending on the type of the link, the impact of spatial resolved variables on speed is not homogeneous, indicating a rather localized impact in the case of lower link types. In the case of ramps, with the exception of other roads type, all the rest seem to have a positive impact on speed, a possible explanation can be that ramps are unlikely to have congestion because of their sort length and thus their speed remains high and close to the existing speed limit. However, that is not the case for the lower classified ramps that seem to have a negative impact on speed. Ramps' density variables have a negative impact on speed besides the cases of collector and trunk roads. This can be possibly explained by the fact that the higher the density of ramps, more vehicles are exiting the roads and thus causing alleviation of traffic. In the case of highways, the magnitude of ramps' density variable is reasonable, since the highest the number of on/off ramps can lead to higher disruption of the traffic flow, and thus decreased speed.

Table 1: Estimated OLS parameters

Y = Morning Peak-Hour Speed Explanatory variables	Estimate	Std. Error	t value	Pr(>|t|)	Signif.
Speed-limit	0.513	0.003	195.572	0.00E+00	***
Highways: Constant	64.956	0.956	67.932	0.00E+00	***
Trunk roads: Constant	43.558	1.124	38.768	0.00E+00	***
Collector roads: Constant	41.744	0.879	47.472	0.00E+00	***
Distributor roads: Constant	41.284	0.234	176.106	0.00E+00	***
Other roads: Constant	34.014	0.197	172.901	0.00E+00	***
Distributor: PuT stops density,r=0.5km	-0.320	0.010	-31.021	9.68E-211	***
Other roads: PuT stops density, r=0.2km	-0.177	0.005	-38.391	3.26e-321	***
Highways: ln(popul, r=5km)	-3.529	0.144	-24.443	9.63E-132	***
Trunk roads: ln(popul,r=2km)	-3.954	0.175	-22.578	1.01E-112	***
Collectors: ln(employm,r=2km,kernel)	-3.723	0.120	-30.897	4.44E-209	***
Distributor: ln(employm,r=1km,kernel)	-2.402	0.026	-92.469	0.00E+00	***
Other roads: ln(employm,r=0.5km,kernel)	-1.531	0.018	-84.887	0.00E+00	***
Trunk roads: Ramp dummy	11.285	0.948	11.907	1.12E-32	***
Collector roads: Ramp dummy	11.722	1.041	11.264	2.02E-29	***
Type 4: Ramp dummy	5.270	0.201	26.206	4.30E-151	***
Type 5: Ramp dummy	-2.822	0.499	-5.653	1.58E-08	***
Highways: Ramps' dens, r=1km	-2.087	0.224	-9.329	1.08E-20	***
Trunk roads: Ramps' dens, r=1km	6.805	0.825	8.246	1.64E-16	***
Collector roads: Ramps' dens, r=1km	3.886	0.764	5.084	3.70E-07	***
Distributor roads: Ramps' dens, r=0.5km	-0.311	0.048	-6.419	1.38E-10	***
Other roads: Ramps' dens, r=0.5km	-0.197	0.073	-2.679	7.39E-03	**
Distributor roads: Road density, r=100 m	-0.173	0.003	-56.836	0.00E+00	***
Other roads: Road density, r=100 m	-0.171	0.003	-56.501	0.00E+00	***
adjusted R-square	*0.9673*				
Log-Likelihood (x 10^4)	*-684431*				
observations	*188428*				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

4.2 Spatial regression models

The first key aspect of proceeding to the estimation of the spatial regression models is to determine the underlying spatial dependence, if any, and incorporate it accordingly in the spatial regression models, in the form of a spatial weight matrix. In order to assess the impact of different weighting matrix type, two categories of weighting matrices are tested thoroughly; one that assigns a uniform weight on all the k-nearest neighbours based on the Euclidean distance, and one similar where once the k-nearest neighbours are identified, the assigned weight is calculated as the inverse squared distance, aiming in capturing the diminishing dependence of links over the distance (distance decay). Subsequently, each row of the latter matrix is standardised to one. In addition, the issue of identifying the extent of the neighbourhood needs to be treated. Driven by this, three SAR models are estimated for different number of neighbours, namely the SARerr, the SARlag, and the SARmix models. The optimum number of neighbours for each model is identified on the basis of minimizing the Akaike Criterion (AIC). The spatial autocorrelation of the OLS residuals is estimated as well in terms of the Moran's I measure, for each weighting scheme. The estimation of the SAR models and the construction of the weighting matrices was conducted in R (R development team, 2005), making use of the package "spdep" (Bivand et al., 2011). It should be noted that for facilitating computationally the estimation of the SAR models, the LU method for the decomposition of sparse matrix is used (LeSage and Pace, 2009).

Table 2: Measures of quality of fit for SAR models for different weight matrices

Number of Neighbors	OLS residuals autocorrelation Binary	Inv. sqr. dist.	SARerror AIC Binary	Inv. sqr. dist.	SARlag AIC Binary	Inv. sqr. dist.	SARmix AIC Binary	Inv. sqr. dist.
3	0.709	0.662	1256831	1272235	1282845	1293182	1245797	1260697
4	0.680	0.653	1245499	1261144	1276645	1285096	1236069	1249901
5	0.660	0.646	1243602	1256478	1276704	1282107	1235519	1245892
6	0.635	0.638	***1241721***	1253803	***1276195***	1280352	***1234832***	1243703
7	0.615	0.631	1243688	1252396	1278225	1279639	1237746	1242814
8	0.593	0.624	1245333	1251527	1280106	***1279362***	1240238	***1242372***
9	0.574	0.618	1248958	1251186	1283485	1279503	1244532	1242429
10	0.554	0.613	1251851	***1250988***	1286159	1279654	1247991	1242542
11	0.537	0.608	1255871	1251017	1289591	1279956	1252482	1242878
12	0.518	0.603	1259222	1251052	1292591	1280288	1256256	1243204
13	0.502	0.599	1263237	1251232	1296075	1280690	1260598	1243640
14	0.486	0.595	1266602	1251387	1298935	1281014	1264229	1244000
15	0.471	0.591	1270264	1251570	1302068	1281400	1268111	1244370
OLS AIC	***1368911***							

As it can be seen in the table above, there is evidence of strong spatial autocorrelation on the residuals of the OLS regression, and naturally the spatial autocorrelation decreases as the number of nearest neighbours increases. The optimum number of k-nearest neighbours, in terms of AIC, for the binary weighting scheme is found to be equal to 6 for all three SAR models. In the case of the inverse weighting scheme, the optimum number differs and it is found to be equal to 10 for the SARerr model, while for the case of the rest two SAR models is found to be 8. It can be concluded that the binary weighting matrix scheme outperforms the inverse square distance weighting scheme and due to that it is the employed one for the comparison and the evaluation of the models that follows.

4.3 Comparison of models

In Table 3, the estimated coefficients, along with the relevant goodness of fit measurements, can be seen. In summary, the coefficients of the OLS model are higher than the corresponding ones in the SAR models, reflecting that in the absence of accounting properly for the spatial dependence, more weight is assigned mistakenly on the structural variables. SAR models are significantly better than the OLS one, all of them having smaller values (in absolute terms) of both the AIC and the Log-likelihood measure. It should be noted that the formulation of the model remains the same in the different model estimations in purpose, in order to allow a comparison of all models in terms of identifying the impacts that the three different SAR models have both on the estimated coefficients and on the results.

Table 3: Estimated coefficients for the different models

Y = Morning Peak-Hour Speed	OLS	SARerr	SARlag	SARmix	
Explanatory variables	Coeff.	Coeff.	Coeff.	Coeff.	Lag.Coeff.
Speed-limit	0.513	0.421	0.282	0.38	-0.038
Highways: Constant	64.956	83.675	31.739	60.38	-8.007
Trunk roads: Constant	43.558	58.389	13.67	32.497	-4.054
Collector roads: Constant	41.744	63.514	18.141	40.257	-5.287
Distributor roads: Constant	41.284	47.721	12.342	23.668	-2.377
Other roads: Constant	34.014	34.477	9.862	8.088	NA
Distributor: PuT stops density,r=0.5km	-0.32	-0.282	-0.099	-0.149	0.010 ‘*’
Other roads: PuT stops density, r=0.2km	-0.177	-0.094	-0.066	-0.039	-0.001‘**’
Highways: ln(popul, r=5km)	-3.529	-4.97	-1.486	-4.282	0.616
Trunk roads: ln(popul,r=2km)	-3.954	-4.038	-1.486	-2.879	0.345‘**’
Collectors: ln(employm,r=2km,kernel)	-3.723	-4.792	-2.06	-3.919	0.504
Distributors: ln(employm,r=1km,kernel)	-2.402	-2.298	-0.858	-1.663	0.189
Other roads: ln(employm,r=0.5km,kernel)	-1.531	-1.061	-0.492	-0.49	0.020‘***’
Trunk roads: Ramp dummy	11.285	6.234	2.278	5.052	0.272‘^’
Collector roads: Ramp dummy	11.722	8.583	5.584	8.2	-0.528‘***’
Distributor roads Ramp dummy	5.27	2.085	-1.597	1.395	0.283
Other roads: Ramp dummy	-2.822	-1.5	-5.069	-1.15	0.243‘^’
Highways: Ramps' dens, r=1km	-2.087	-3.902	-5.275	-3.739	0.544
Trunk roads: Ramps' dens, r=1km	6.805	5.537	2.916	4.889	-0.769‘^^’
Collectors: Ramps' dens, r=1km	3.886	2.936 ‘****’	-0.325 ‘^’	1.631 ‘^’	-0.199‘^^’
Distributors: Ramps' dens, r=0.5km	-0.311	-0.059 ‘^^’	-0.673	-0.149	-0.018‘^’
Other roads: Ramps' dens, r=0.5km	-0.197 ‘****’	-0.443	-0.924	0.045	0.001‘^’
Distributors: Road density, r=100 m	-0.173	-0.078	-0.074	-0.063	0.0003‘^^’
Other roads: Road density, r=100 m	-0.171	-0.129	-0.095	-0.095	0.007
ρ	-	-	0.101		0.126

λ	-	0.131		-	
AIC	1368911	1241721	1276195		1242372
adjusted R-square	*0.9673*	-	-	-	
Log-Likelihood (x 10^4)	-684431	-620834	-638072		-617367
Signif. codes: 0.001 ‘****’ 0.01 ‘***’ 0.05 ‘**’ 0.1 ‘*’ 0.5‘^^’ 1‘^’					

The predictive power of each model is calculated in order to facilitate their comparison and draw some conclusions with respect to their ability to make accurate predictions. The predictive accuracy in terms of predicted values that are within different specified ranges is presented in Table 4. As it can be seen, OLS model performs relatively bad since less than 50% of the predictions fall within a range of 10%. On the other hand the predictive accuracy of the SAR models is much better and it is clearly reflected that accounting for the spatial dependence of data, in addition to the structural variables can lead to significantly improved predictions. This finding shows that in the case of OLS models, the estimated coefficients are inconsistent and biased since more explanatory power is attributed on them. Among the SAR models, clearly the SARerr model is better than the SARlag model, indicating that accounting for the spatial dependence in the error terms of the model is more important than accounting for the spatial dependence in the response variable. Nevertheless, the SARmix model that accounts for both, is slightly better than SARerr model. However, in terms of AIC values, SARerr model seems to be better. The disadvantage of the estimated SARmix model is that many of the constructed spatial lagged variables' coefficients are insignificant while their interpretation is less straightforward than of the other models and should be conducted with caution. In summary, the results of the SAR models can be considered that they highlight the impact of accounting properly for the spatial dependence of transport related data.

Table 4: Predictive accuracy of estimated models in terms of predicted speeds within specified range of actual speeds

Model	2% range	5% range	8% range	10% range	15% range	20% range	30% range
OLS	9.43%	23.16%	36.39%	44.71%	62.38%	74.87%	86.62%
SARerrr	20.04%	45.51%	63.05%	71.13%	83.20%	89.26%	94.54%
SARlag	16.15%	38.57%	56.51%	65.64%	79.87%	86.93%	93.13%
SARmix	20.45%	46.10%	63.66%	71.64%	83.32%	89.31%	94.60%

5. Conclusion

In the present paper, the first steps of building up a simplified model towards a direct demand modelling approach were presented. The first results showed the plausibility of the spatial regression models to be used in that context and highlight how the careful set up of the weighting matrices can contribute to obtaining better predictions.

The next steps will be in the direction of further investigating issues regarding the form of the employed spatial regression models and also the form of the weighting matrices. An apparent improvement would be to account for the network connectivity of the links, in line with the approach presented by Hackney et al.(2007). Moreover, modelling approaches that account for structural instability in space (eg geographically weighted regression) will be tested as well to conclude in the best modelling approach for modelling transport related phenomena. The modelling approach will be extended and applied for the estimation of the average daily volume on the links, in order to be able to form a complete framework that can provide the essential answers for transport project appraisal purposes. In addition, the results of the developed approach will be compared with the results of existing modelling techniques (MATSim and VISUM) and an analytical comparison will be done within the context of a cost-benefit analysis for a given set of transport projects and policies to identify the trade-offs between the modelling technique and the supported choices.

6. References

Anselin L. (1988) Spatial econometrics: methods and models, Kluwer Academic, Dordrecht.

Bivand, R., L. Anselin, O. Berke, A. Bernat, M. Carvalho, Y. Chun, C. F. Dormann et al. (2011) spdep: Spatial dependence: weighting schemes, statistics and models.

Bernard, M., J. Hackney and K. W. Axhausen (2006) Correlation of link travel speeds, *6th Swiss Transport Research* Conference, Ascona.

Cheng, T., J. Haworth and J. Wang (2011) Spatio-temporal autocorrelation of road network data, *Journal of Geographical Systems*, *14*(4), 389–413.

Flyvbjerg, B., M. K. Skamris Holm and S. L. Buhl (2005). How (In) accurate are demand forecasts in public works projects, *Journal of the American Planning Association*, *71*.

Hackney, J. K., M. Bernard, S. Bindra and K. W. Axhausen (2007) Predicting road system speeds using spatial structure variables and network characteristics, *Journal of Geographical Systems*, *9*(4), 397–417.

Kissling, W. D. and G. Carl (2007) Spatial autocorrelation and the selection of simultaneous autoregressive models, *Global Ecology and Biogeography*, *17*(1), 59–71.

LeSage, J. and R. Pace (2004) Spatial and spatiotemporal econometrics, *Advances in Econometrics*, *18*(4), 1–32.

LeSage, J. and R. K. Pace (2009) Introduction to Spatial Econometrics. CRC Press, Taylor and Francis Group.

Löchl, M. and K. W. Axhausen (2010) Modelling hedonic residential rents for land use and transport simulation while considering spatial effects, *Journal of Transport and Land Use*, *3*(2), 39–63.

Ord J. K. (1975) Estimation Methods for Models of Spatial Interaction, *Journal of the American Statistical Association*, 70, 120-126.

Páez, A. and D. M. Scott (2004) Spatial statistics for urban analysis: A review of techniques with examples, *GeoJournal*, *61*(1), 53–67.

R Development Core Team (2011) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.