EFFECTS OF MODEL STOCHASTICITY ON UNCERTAINTY IN PREDICTED DESTINATION TOTALS AND TRAFFIC VOLUMES

Soora Rasouli, Eindhoven University of Technology, s.rasouli@tue.nl

Harry J. P. Timmermans, Eindhoven University of Technology

ABSTRACT

This paper reports results of uncertainty analysis of the Albatross model system. More specifically, the effects of model uncertainty on destination totals and traffic volumes, predicted by Albatross are investigated, using the city of Rotterdam as an example. The study involves 1000 runs of model system for a synthetic population of 41,668 individuals. Results indicate that the average uncertainty in the predicted OD matrices due to model uncertainty is 45 per cent, and 0.13 per cent for destination totals based on these simulation runs. In general, uncertainty is lower for the destinations with higher traffic volumes. Uncertainty in predicted traffic volumes, represented by the cells of the OD matrix, tends to be higher. Finally, for both types of indicators, there is evidence of spatial variability in coefficients of variation (CV), capturing respectively uncertainty in destination totals and traffic volumes. Generally, uncertainty is a non-linear function of the number of samples.

Keywords: Model uncertainty, Activity-based model, Trip frequency, Coefficient of variation

INTRODUCTION

Research on model uncertainty of activity-based models of travel demand is still in its infancy. Little is known about the effects of simulation error on the forecasts of these models. Investigating uncertainty in forecasts of travel demand model is of the utmost importance to differentiate policy effects from simulation error and better understand the probabilistic nature of travel demand forecasts. In addition, knowing uncertainty associated with a model's forecasts allows improving the model and make better operational decisions in model applications.

Most previous research has been concerned with the traditional four-step model of travel demand and tour-based models (Rasouli and Timmermans, 2012). Examples include Rodier and Johnston (2002), Zhao and Kockelman (2001), De Jong *et al*. (2007), Walker (2005) and Yang and Chen (2010, 2011). For example, Zhao and Kockelman (2001), investigating input uncertainty and error propagation of a conventional four-step model for an 818-link network covering 25 zones in the Dallas-Fort Worth metropolitan region, found that the coefficients of variation of two link flows were larger than the 0.3 for combination of model and input uncertainty. They analyzed input uncertainty by varying the coefficients of variation for the number of households and different employment types from 0.1, via 0.3 to 0.5 and model uncertainty by considering CV of model parameters as 0.30 and later varied to 0.1 and 0.50, assuming multivariate normally distributions and a correlation of + 0.3 across all variables. The four-step model was run 100 times with different input values, drawn from these distributions. Similarly, De Jong *et al*. (2007) conducted an uncertainty analysis of the Dutch national and regional travel demand model used the standard deviations and correlations of 20-year moving averages of some input data to extract values from a multivariate normal distribution. They ran the model 100 times, 50 times for a reference scenario and 50 for a new infrastructure project. For each of these 50 runs, 20 were made for varying input variables, 20 with varying model coefficients and 10 in which input variables were combined with model coefficients. Uncertainty of the LMS (the National Model System) forecasts was assessed at the level of aggregate travel indices (such as number of tours and passenger kilometers by mode and purpose) and the link level (traffic flows in passenger car equivalents, travel times and vehicle hours lost on a number of selected links). They found input uncertainty to be higher than model uncertainty. Standard deviations of the link flows were between 4% and 9% for input uncertainty, and around 1% for model uncertainty. The most elaborate study to date was conducted by Yang and Chen (2010, 2011). They investigated uncertainty and error propagation in Oppenheim's combined travel demand model (Oppenheim, 1995). Both input and model uncertainty in model forecasts, considering travel demand, traffic flows, and travel costs as output variables, was assessed for the Sioux Falls network, consisting of 24 nodes and 76 links, reduced to two modes (car and transit). Inputs were assumed independently and normally distributed, with a coefficient of variation of 0.3. Results indicated that the CV of travel demand and traffic flows was almost identical to input uncertainty. The coefficient of variation of total travel time and total vehicle miles are lower than the coefficient of variation of inputs. Contrary to De Jong et al. They found the effects of model uncertainty higher.

Research on rule-based travel demand forecasting models is even more limited. It has been primarily concerned with uncertainty of aggregate travel indicators such as daily total travel distance and daily time use (e.g., Cools et al., 2011; Rasouli et al., 2011, 2012a, 2012b). Cools *et al.* (2011) examined uncertainty of the FEATHERS model, the Flemish version of the Albatross model. They ran the model 200 times for the same 10% fraction of the synthetic population. Uncertainty, measured in terms of the coefficient of variation, was assessed for the average daily number of trips per person and the average daily distance travelled per person, both for the entire sample and for segments, defined by mode choice, age and gender. Calculated coefficients of variation based on the 200 runs were compared against a 1.27% threshold error rate, which corresponds to the corresponding 95% confidence bounds of a 5% deviation. Results showed that this threshold value was often exceeded for public transport. Overall, the effects of model uncertainty on aggregate travel indices were small: 0.12 for the daily number of trips and 0.20 for daily distance travelled per person. Rasouli et al. (2011, 2012a, 2012b) found uncertainty in number of trips is higher than uncertainty in distance travelled in an application of the Albatross system to the Rotterdam area. Thus, compared to the already limited research on uncertainty analysis on four-step and tour-based models of travel demand, knowledge about uncertainty in rule-based models of travel demand is very limited. Yet, the study of model uncertainty in rule-based models is of particular interest because different constraints and non-linearity in responses may affect predicted choices in distinct ways.

To contribute to this limited line of research, this paper reports the effects of model uncertainty on predicted destination totals and traffic volumes by the Albatross model system (Arentze and Timmermans, 2000, 2004, 2005). Albatross is a rule-based model. It consists of 27 decision tables that are activated according to some sequence, which reflects an assumed priority ordering in the scheduling of activities and travel. Each decision table uses a set of socio-demographic characteristics of travellers and the state of the environment as input. Moreover, the outcomes of previous scheduling decisions are used as input in subsequent decision trees, representing the next scheduling decisions. Albatross also incorporates the decisions of other household members to model task allocation, joint activity participation and resource allocation decisions.

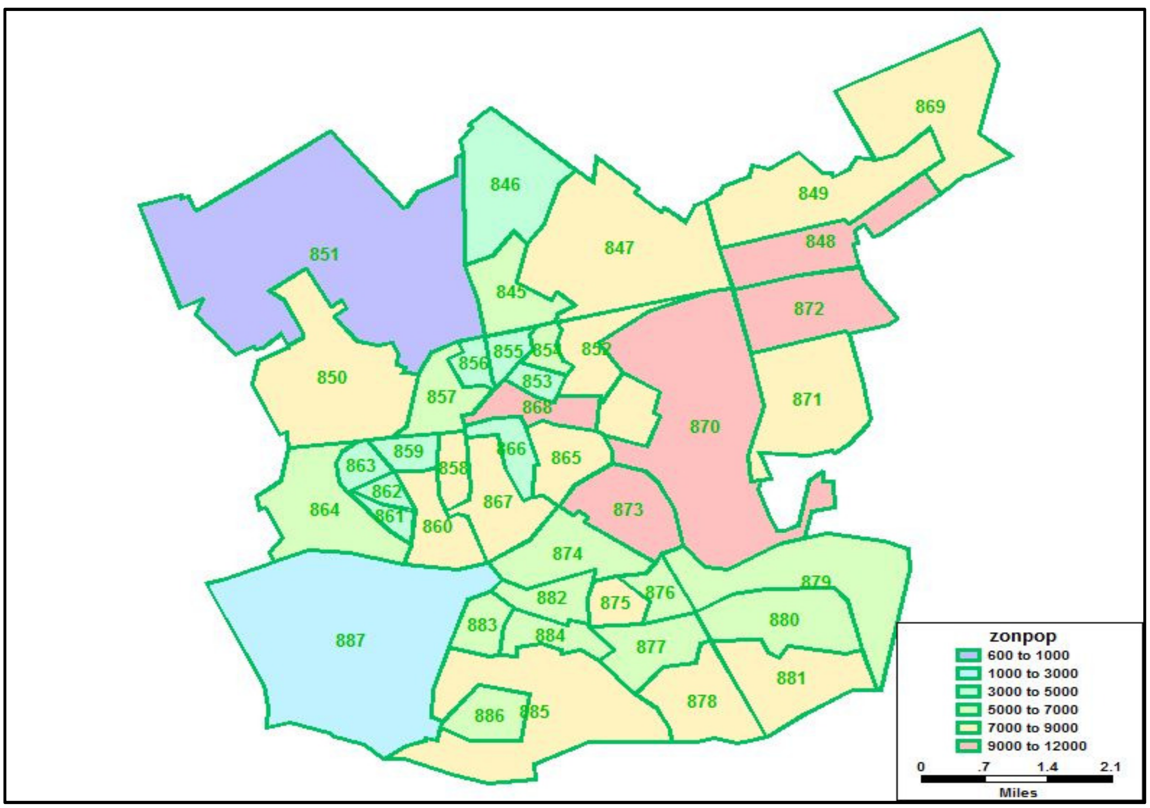

Figure 1: Population distribution across zones in Rotterdam

DATA AND METHODOLOGY

The Rotterdam study area

The uncertainty analysis of the Albatross model system was conducted for the city of Rotterdam, the second largest city in the Netherlands with a population of approximately 600,000 inhabitants in 2005. Rotterdam is a harbour city, which was destroyed during WW II. Consequently, it is a modern city, with a city centre, more or less in the middle of the city. Major highways runs North-South and from the East to the harbour. Figure 1 depicts household population distribution across zones in Rotterdam.

The synthetic population

Because Albatross simulates activity-travel patterns of individuals and households as a function of their socio-economic profile, individual and corresponding household profiles are required for every individual and household in the Rotterdam study area. Such population data are, however, not available. The only information that is available concerns marginal distributions of socio-economic variables for Rotterdam. As a solution to this problem, a synthetic population is created. It is based on the notion to derive individual and household profiles such that (i) aggregations of the derived data are consistent with available distributions for the city and (ii) the correlations in the derived profiles are consistent with those observed in sample data. In the present case, the correlations observed in the National Travel Survey for similar cities were used to accomplish the latter requirement.

In the literature, several algorithms have been suggested to create a synthetic population. For example, Beckman et al. (1996) suggested using the technique of iterative proportional fitting using PUMS data. Other recent examples include Guo and Bhat (2007), Auld and Mohammadian (2010), Mohammadian et al., (2010) and Abraham et al. (2012). As evidenced by these examples, the specific approach often depends on the available data, although the algorithms have more general meaning. In the present study, the synthetic population was created as follows. Attributes of households/individuals that are included are household type (single non-worker, single worker, double non-workers, double one-worker, double two workers), age of oldest member of household (younger than 25 years, 25 – 44, 45 – 64, 65 or older), age of youngest child (no children, younger than 6 years, 6 – 11 years, 12 or older), socio-economic class (very low household income, low, average, high income), number of bikes in household, number of cars in household, gender of person, driver's license, car availability, bike availability and number of weekly working hours of person. The data used to create a synthetic population are based on the Dutch National Travel Survey and a division of the Netherlands into a 1308 zones. Iterative proportional fitting was used to synthesize individual data. The concept of a relation matrix was used to convert individual count data to household count data for age groups and work status. More specifically, an age-group relation matrix specifies the relation between our age groups and three household status positions for females and males. Again, the IPF technique is used to derive estimated cell frequencies, based on the National Travel Survey sample data and marginal constraints, derived from demographic data. The frequencies of a similar work-status relation matrix were obtained in a similar vein. This matrix distinguishes no work, part-time work and full-time work, which are linked to the household status categories. It results in a distribution of 15 household types comprising 3 × 3 double work status groups, 3 single female work status groups and 3 single male work status groups (Arentze et al., 2007).

Model predictions

The actual uncertainty analysis involved the following procedure. First, a random sample was drawn from the synthetic population. More specifically, a 10% fraction of the synthetic population, consisting of 41,668 persons and 27,961 households, was randomly selected. To rule out the possibility that results are influenced by the sampled fraction, it was kept constant for all analyses in the present study. Next, for each sampled individual of this fraction, the Albatross model was run 1000 times. In each run, the 27 decision trees making up the Albatross model system were activated according to the sequence underlying the assumed process model. The simulated realizations of previous decision tables were used as input for the next decision tables. Choices were simulated by Monte Carlo draws from the probabilistic action tables. These runs thus result in different individual-level simulated activity-travel patterns. Running the model multiple times allows one to analyse the effects of the number of runs on the uncertainty of a set of performance indicators and travel indices. In this study, the focus is on predicted OD matrices, which were derived by aggregating predicted individual space-time trajectories into origin-destination tables.

The coefficient of variation (CV) was used as a measure of uncertainty. If *x* is a normal random variable with mean μ and variance σ^2, then the parameter

$$\kappa \equiv \frac{\sigma}{\mu} \qquad (1)$$

is called the population coefficient of variation. A point estimate of (1) is given by

$$K \equiv \frac{s}{\overline{x}} \qquad (2)$$

where $\overline{x}$ and s are respectively the mean and standard deviation of the sample.

RESULTS

The study area was divided into 66 postal areas. Thus, uncertainty at the cell level of the OD matrix, depicting the number of trips between an origin and a destination can be calculated for 66 x 66=4357 cells. In addition, uncertainty can be calculated for 66 destination totals. Uncertainty estimates at the cell level are relevant for transportation planning in the sense that these cells of the OD matrix constitute input for traffic assignment and are thus influential to simulated traffic flows on the network. Uncertainty estimates at the level of destination totals are more relevant to urban planning applications. These destination totals are often used to assess the feasibility of new plan proposals. For example, if the destination total relates to shopping, this indicator represents the number of predicted people to shop in a certain postcode area. This number when combined with expenditure data can be used to assess floor productivity of retail space in that zone (e.g. Rasouli and Timmermans, 2013).

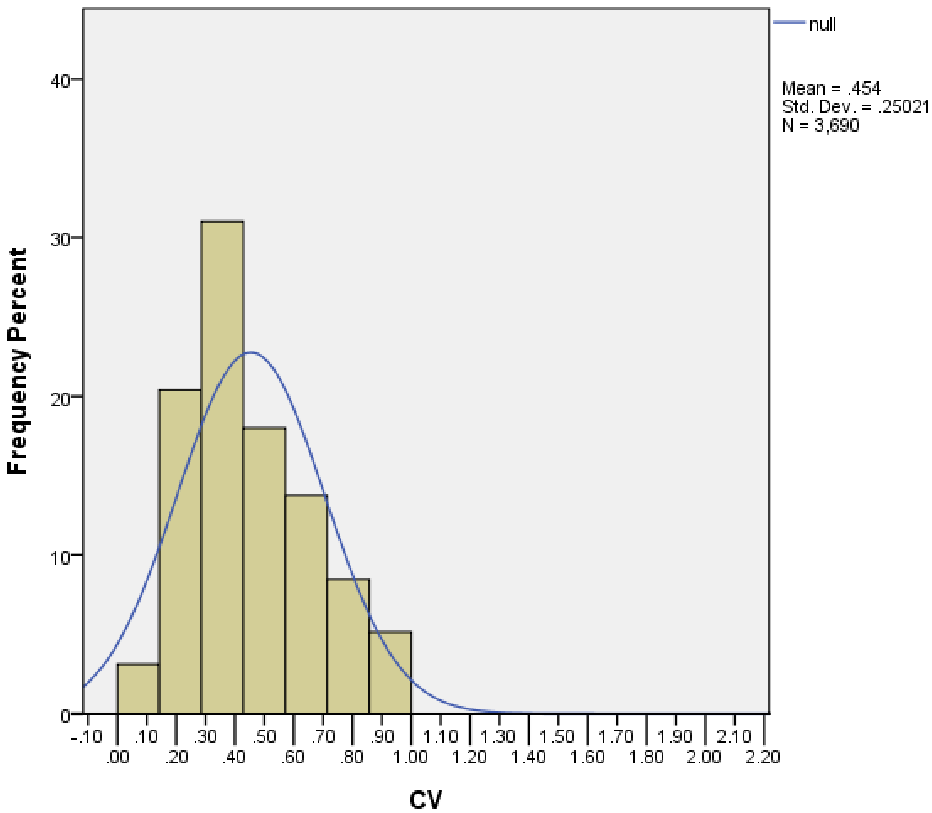

Figure 2 - Distribution of coefficient of variation for traffic volume on OD pairs

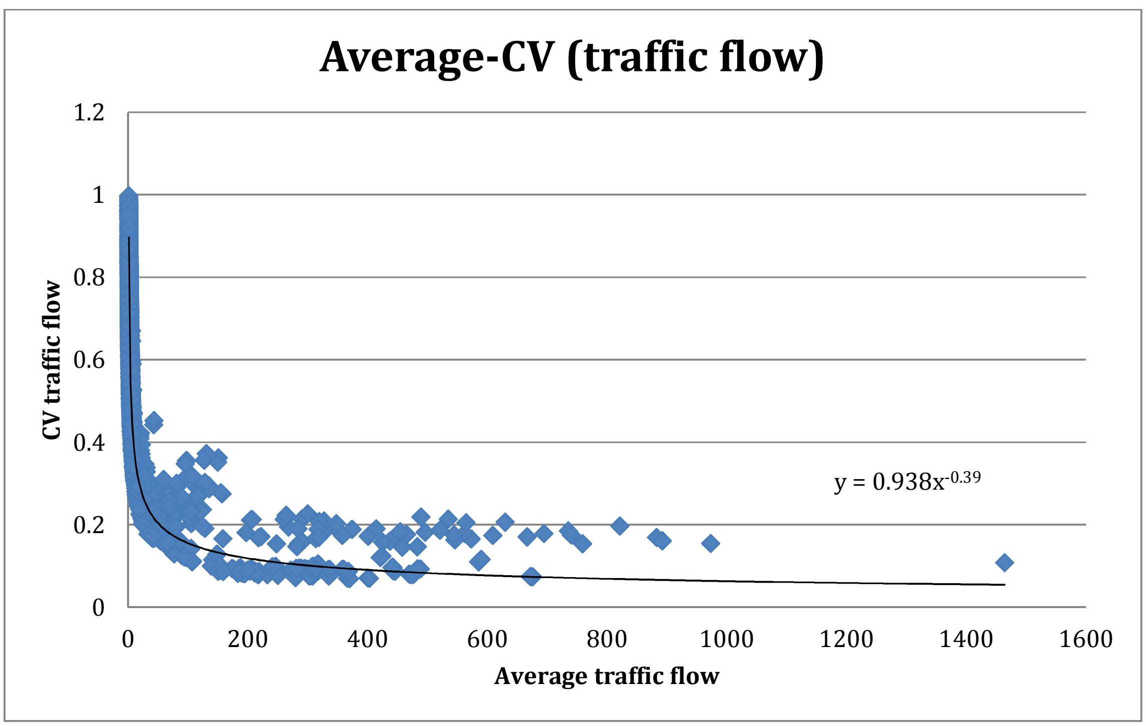

Figure 3 - Relation between average traffic volume and CV

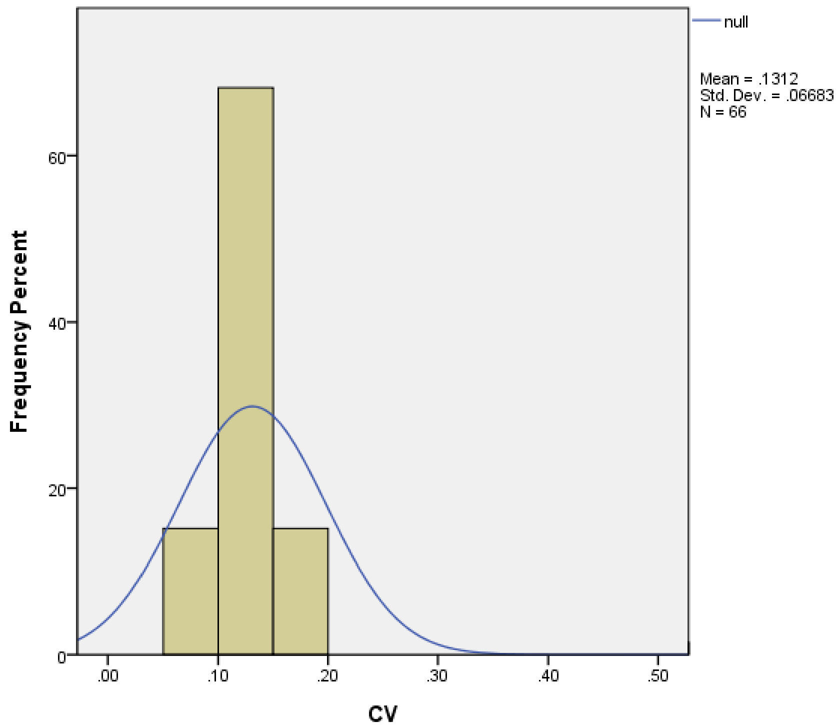

Figure 4- Distribution of coefficient of variation for destination totals

Figure 2 represents the CV distribution over all OD pairs. The average and standard deviation of CV are 0.45 and 0.25. Because as one might expect that model uncertainty is related to the number of trips, which varies across the study area, in addition to the aggregate results, the spatial distribution of the coefficient of variation was investigated. To that end, the calculated CV for all pairs in the OD matrix were graphed against the total number of trips observed for that pair. The result is visualised in Figure 3. Note that considering the population fraction of 10 per cent in this analysis, some OD pairs have very low predicted trip frequencies (the average less than 1 trip in some cases). Therefore, variation is relatively high as one might expect. 15 per cent of PCA pairs had very low predicted trip frequencies and were therefore removed from the data.

Uncertainty in destination totals

The predicted number of trips between origins and destinations can be summed across origins to derive for each destination the total number of trips arriving in that destination. As discussed, this statistic is often crucial for assessing the spatially varying impact of plans. For each of the 1000 runs, the predicted origin-destination tables were aggregated to arrive at destination totals for each run. Next, the coefficient of variations and their standard deviations were calculated for each destination. Table 1 provides an overview of the resulting CV. Figure 4 depicts the distribution of CV for destination totals. The table and figure show that the coefficient of variation varies between 0.09 and 0.19 with the exception of PCA 3059 with a CV of 0.62. The highest number of destinations falls in the 0.09 - 0.13 category.

As expected, these findings indicate that uncertainty of model forecasts, due to the inherent stochastic nature of the Albatross model system, is lower for destination totals, compared to traffic volumes. This finding is the immediate result of the larger number of observations for the destination totals. To explore spatial effects, Figure 5 maps the results of the coefficient of variation for the destination totals for the various destinations. The map clearly shows that most destinations (55 out of the 66 postal code areas) have a CV of less than 0.15. Since the CV was calculated based on the traffic of Rotterdam, some PCA's which are located at the border of the study area tend to have higher coefficients of variation due to the fact that a large portion of traffic of these areas is generated from outside of Rotterdam.

IMPICATIONS FOR RESEARCH/POLICY

This paper has reported the main results of an analysis of model uncertainty of the Albatross model system for a synthetic population of the Rotterdam analysis. It contributes to the increasing interest in examining uncertainty in model forecasts of models of travel demand.

Policy makers have realised that uncertainty in model forecasts in some cases may be as relevant as predicted averages. Moreover, the increasing complexity of models of travel demand and the inherent stochastic nature of the latest generation of discrete choice and activity-based models have made researchers realize that formal uncertainty analysis is required to differentiate policy effects from inherent probabilistic forecasts of these models.

While uncertain analysis has been applied to four-step and discrete choice models, its application to an activity-based model is a relatively novel feature of this research. Relatively few prior studies have been concerned with activity-based models. The number of studies on rule-based model systems in transportation research is even smaller. Moreover, the very limited empirical evidence obtained thus far has been concerned with aggregate travel indices. Thus, this study is to the best of our knowledge, the first study examining the effects of model uncertainty of a rule-based model system of travel demand on predicted destination totals and traffic volumes.

The findings of this study indicate that the average uncertainty in the predicted OD matrices due to model uncertainty is 45 per cent, and 13 per cent for destination totals, based on 1000 simulation runs. In general, this uncertainty is lower for the destinations with higher traffic volumes and higher for less frequently visited zones. In that sense, clear spatial effects could be discerned. Uncertainty in predicted traffic volumes, represented by the cells of the OD matrix, tends to be higher

The implications of these findings for transportation policy are manifold. First, the importance of conducting a formal uncertainty analysis is evidenced by the results of this study. The substantial size of the coefficient of variation under particular conditions suggests that a single run may deviate substantially from the average prediction. Second, results indicate that a small number of simulation runs may not be sufficient for assessing uncertainty in OD matrices. It implies that either an even larger number of model runs would be required, or that the sample for the concerned cells of the OD matrix or the concerned destinations, should be increased or a combination of these strategies would be required. Third, transportation policy should develop and explore approaches of how to include this uncertainty in policy evaluations.

The results also point at interesting avenues of future research. It will be evident that the approach reported in this paper, which involved 1000 model runs is very time and resource

Table 1 - Coefficient of variation for uncertainty in destination totals

Dest	Average	Std	CV	Dest	Average	Std	CV	Dest	Average	Std	CV
3011	2356.17	301.37	0.128	**3038**	1206.10	150.21	0.125	**3067**	2889.40	443.02	0.153
3012	1944.02	252.72	0.130	**3039**	1712.53	213.62	0.125	**3068**	3033.39	307.71	0.101
3013	647.51	62.044	0.096	**3041**	106.45	19.81	0.186	**3069**	2353.54	301.87	0.128
3014	1282.04	146.34	0.114	**3042**	1228.24	142.96	0.116	**3071**	2634.43	306.72	0.116
3015	1195.59	109.17	0.091	**3043**	853.01	87.39	0.102	**3072**	2004.59	230.21	0.115
3016	545.63	59.805	0.110	**3044**	439.10	28.99	0.066	**3073**	2003.09	250.82	0.125
3021	1855.14	233.59	0.126	**3045**	298.76	51.33	0.172	**3074**	1827.00	260.64	0.143
3022	1383.96	153.70	0.111	**3046**	61.86	12.21	0.197	**3075**	1862.07	242.45	0.130
3023	1251.46	153.09	0.122	**3047**	126.49	16.39	0.130	**3076**	2170.06	260.68	0.120
3024	1161.09	110.84	0.095	**3051**	1232.96	121.03	0.098	**3077**	2052.95	233.99	0.114
3025	1005.80	144.21	0.143	**3052**	892.55	91.00	0.102	**3078**	1888.11	333.35	0.177
3026	1009.80	122.28	0.121	**3053**	1256.25	152.63	0.122	**3079**	2219.38	277.62	0.125
3027	1309.63	142.64	0.109	**3054**	1142.64	143.49	0.126	**3081**	1656.58	209.59	0.127
3028	1086.19	144.77	0.133	**3055**	983.39	118.14	0.120	**3082**	1612.72	209.97	0.130
3029	744.29	70.785	0.095	**3056**	176.88	27.47	0.155	**3083**	2433.26	438.97	0.180
3031	1375.10	174.39	0.127	**3059**	3.21	2.01	0.625	**3084**	505.62	68.48	0.135
3032	1126.14	122.68	0.109	**3061**	2347.42	347.02	0.148	**3085**	1823.84	275.76	0.151
3033	788.23	90.43	0.115	**3062**	1407.29	155.79	0.111	**3086**	1692.57	257.30	0.15
3034	1975.46	231.90	0.117	**3063**	1717.91	157.32	0.092	**3087**	378.73	46.49	0.123
3035	1155.84	160.42	0.139	**3064**	249.84	27.97	0.112	**3088**	244.64	20.72	0.085
3036	1350.46	176.13	0.130	**3065**	974.82	113.69	0.117	**3089**	525.48	49.30	0.094
3037	1211.39	187.20	0.155	**3066**	1559.39	173.63	0.111	**9999**	16572.01	975.10	0.059

demanding. If, as the results suggest, more runs are needed, the approach may be too demanding. Hence, it is interesting and relevant in future research to find ways to reduce this computational burden. Technological solutions such as cloud and parallel computing can be explored, but research into clever sampling methods should be stimulated.

ACKNOWLEDGEMENTS

This work has been funded by the European Commission, under the Seventh Framework Programme, by Contract no. 248488 within project The Uncertainty Enabled Model Web. The views expressed herein are those of the authors and are not necessarily those of the European Commission.

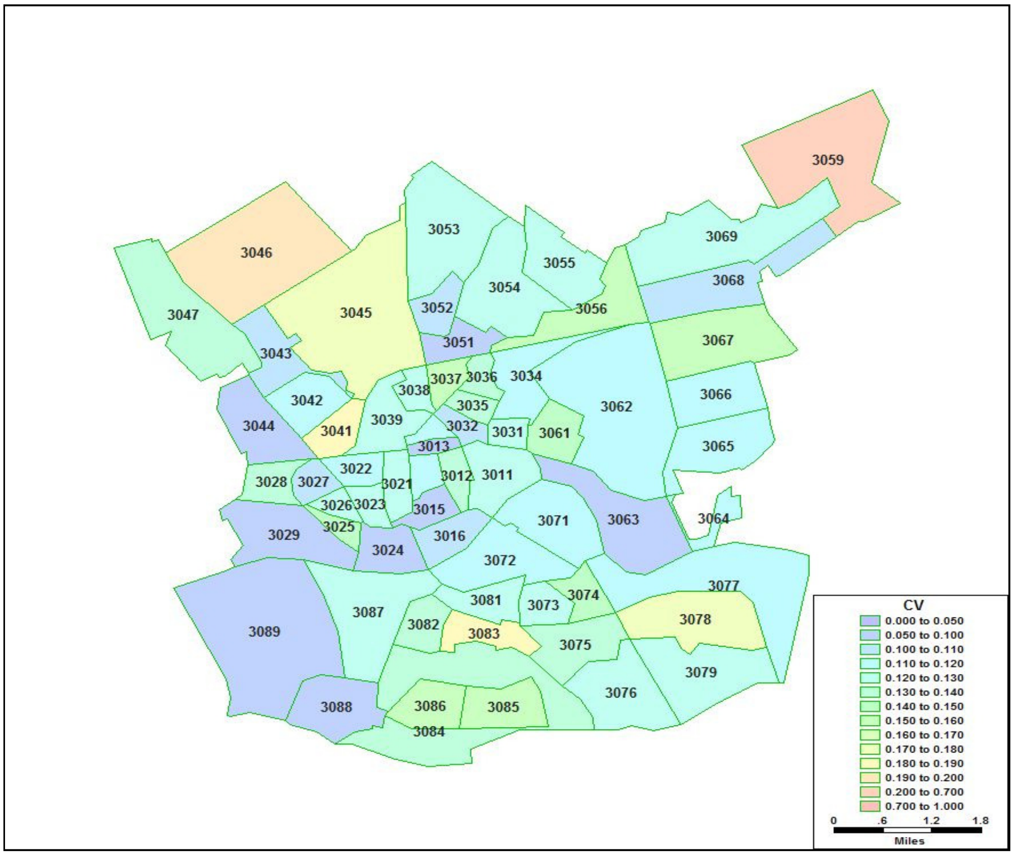

Figure 5 - Spatial distribution of coefficient of variation

REFERENCES

Abraham, J. E., K. J. Stefan J. D. Hunt (2012). Population synthesis using combinatorial optimization at multiple levels, Proc. TRB Annual Meeting, Washington D. C.

Arentze, T. A. and H. J. P. Timmermans (2000). Albatross: A Learning-Based Transportation Oriented Simulation System, EIRASS, Eindhoven University of Technology, Eindhoven, The Netherlands.

Arentze, T. A. and H. J. P. Timmermans (2004). A learning-based transportation oriented simulation system, Trans. Res. B, 38, 613-633.

Arentze, T. A. and H. J. P. Timmermans (2004). Re-induction of Albatross' decision rules using pooled activity-travel diary data and an extended set of land use and costs-related condition states, Trans. Res. Rec., 1831, 230-239.

Arentze, T. A. and H. J. P. Timmermans (2005). Albatross V2: A Learning-Based Transportation Oriented Simulation System, EIRASS, Eindhoven University of Technology, Eindhoven, The Netherlands

Arentze, T. A., H. J. P. Timmermans and F. Hofman (2007). Creating synthetic household populations: problems and approach. Proc. 86th Annual Trans. Res. Board, Washington D. C.

Auld, J. and A. K. Mohammadian (2010). An efficient methodology for generating synthetic populations with multiple control levels, Proc. 89th Annual Trans. Res. Board, Washington D. C.

Beckman, R. J., K. A. Baggerly and M. D. McKay (1996). Creating synthetic baseline populations. Trans. Res. B, 30, 415-429.

Bradley, M., M. Outwater, N. Jonnalagadda and E. Ruiter (2001). Estimation of an activity-based microsimulation model for San Francisco, Proc. 80th Annual Trans. Res. Board, Washington D. C.

Cools, M., B. Kochan, T. Bellemans, D. Janssens and G. Wets (2011). Assessment of the effect of microsimulation error on key travel indices: Evidence from the activity-based model Feathers, Proc. 90th Annual Trans. Res. Board, Washington D. C.

Guo, J.Y. and C. R. Bhat (2007). Population synthesis for microsimulating travel behavior, Proc. 86th Annual Trans. Res. Board, Washington D. C.

Jong, G. C. de, A. J. Daly, M. Pieters, S. Miller, R. Plasmeijer and F. Hofman (2007) Uncertainty in traffic forecasts: Literature review and new results for The Netherlands, Trans., 34, 375-395.

Mohammadian, A. K., M. Javanmardi and Y. Zhang (2010). Synthetic household travel survey data simulation, Trans. Res. C, 18, 869-878.

Rasouli, S., T. A. Arentze and H. J. P. Timmermans (2011). Error propagation in complex large-scale computational process models of activity-travel behavior. In: Transportdynamics (W. Y. Szeto, S. C. Wong and N. N. Sze (eds.), pp. 291-298. HKSTS, Hong Kong, China,

Rasouli, S., T. A. Arentze and H. J. P. Timmermans (2012a). Analysis of uncertainty in performance indicators of a complex activity-based model: The case of the Albatross model system. Proc. Innovations in Travel Demand Modelling Conf., Tampa, USA (On-line: 7 pp).

Rasouli, S., M. Cools, B. Kochan, T. A. Arentze, T. Bellemans and H. J. P. Timmermans (2012b). Uncertainty in forecasts of complex rule-based systems of travel demand: Comparative analysis of the Albatross/Feathers model system. Proc. Int. Ass. of Travel Behavior Res. Conf., Toronto, Canada (CD-Rom: 17 pp).

Rasouli, S. and H. J. P. Timmermans (2012). Uncertainty in travel demand forecasting models: Literature review and research agenda, Trans. Lett., 4, 55-73.

Rasouli, S. and H. J. P. Timmermans (2013). Assessment of model uncertainty in destination and travel forecasts of models of complex spatial shopping behaviour, J. of Ret. and Cons. Serv., to appear.

Rodier, C. J. and R. A. Johnston (2002). Uncertain socioeconomic projections used in travel demand and emissions models: could plausible errors result in air quality nonconformity?, Trans. Res. A, 36, 613-631.

Yang, C. and A. Chen (2010). Uncertainty analysis of a combined travel demand model. In: Transportation and Urban Sustainability (Sumalee, A., W. H. K. Lam, H. W. Lo and B. Siu (eds.), HKSTS, Hong Kong, pp. 441-450.

Yang, C. and A. Chen (2011). Sensitivity-based uncertainty analysis of a combined travel demand model. Proc. 90th Annual Trans. Res. Board, Washington D. C.

Zhao, Y. and K. M. Kockelman (2001). The propagation of uncertainty through travel demand models: An exploratory analysis, Ann. of Reg. Sci., 36, 909-921.

Ziems, S., S. Bhargava, J. Plotz and R. M. Pendyala (2011). Stochastic variability in microsimulation modeling and convergence of corridor-level characteristics. Proc. 90th Annual Trans. Res. Board, Washington D. C.