

# City-scale Traffic Simulation From Digital Footprints

Gavin McArdle  
National Centre for Geocomputation  
National University of Ireland Maynooth  
Maynooth, Co. Kildare, Ireland  
gavin.mcardle@nuim.ie

Eoghan Furey  
National Centre for Geocomputation  
National University of Ireland Maynooth  
Maynooth, Co. Kildare, Ireland  
eoghan.furey@nuim.ie

Aonghus Lawlor  
National Centre for Geocomputation  
National University of Ireland Maynooth  
Maynooth, Co. Kildare, Ireland  
aonghus.lawlor@nuim.ie

Alexei Pozdnoukhov  
National Centre for Geocomputation  
National University of Ireland  
Maynooth, Co. Kildare, Ireland  
alexei.pozdnoukhov@nuim.ie

## ABSTRACT

This paper introduces a micro-simulation of urban traffic flows within a large scale scenario implemented for the Greater Dublin region in Ireland. Traditionally, the data available for traffic simulations come from a population census and dedicated road surveys which only partly cover shopping, leisure or recreational trips. To account for the latter, the presented traffic modelling framework exploits the digital footprints of city inhabitants on services such as Twitter and Foursquare. We enriched the model with findings from our previous studies on geographical layout of communities in a country-wide mobile phone network to account for socially related journeys. These datasets were used to calibrate a variant of a radiation model of spatial choice, which we introduced in order to drive individuals' decisions on trip destinations within an assigned daily activity plan. We observed that given the distribution of population, the workplace locations, a comprehensive set of urban facilities and a list of typical activity sequences of city dwellers collected within a national road survey, the developed micro-simulation reproduces not only the journey statistics but also the traffic volumes at main road segments with surprising accuracy.

## Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications—*data mining, spatial databases and GIS*; I.6.5 [Simulation and Modelling]: Model Development—*modelling methodologies*

## Keywords

urban mobility, spatial choice, location based social networks, agent based traffic modelling

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM UrbComp'12, August 12, 2012, Beijing, China  
Copyright 2012 ACM 978-1-4503-1542-5/08/2012 ...\$15.00.



Figure 1: The road network in Greater Dublin region used for modelling.

## 1. INTRODUCTION

New opportunities in the study of human mobility arise from the availability of digital traces of movement such as the check-in data of location-based social network services or Call Detail records of cell phone usage. They allow for uncovering details about urban mobility previously unavailable from traditional travel surveys, such as an evidence of a long-tail in the daily trip distribution indicating that some individuals cover distances orders of magnitudes larger in their typical trips than a majority of other users [7]. Check-in services such as Foursquare [19] or geo-referenced Twitter [21, 20] provide new insights for transportation modelling by providing quantifiable evidence about the purpose of individuals' travel whether for shopping, leisure or recreation, or meeting friends and visiting family. These activities generate a considerable amount of road traffic which needs to be accounted for in transportation models. However, they are not covered in detail by traditional travel surveys [3] and rely on perceived rather than measured trip lengths and are likely to contain biases.

## 1.1 Contributions of this work

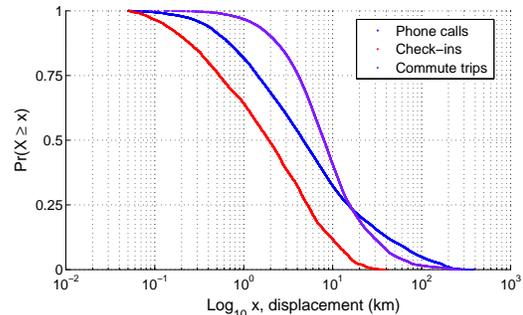
In this paper we investigate the usefulness of digital footprints of individual movement for calibrating human mobility models within an urban traffic micro-simulation framework. We implemented a large scale realistic working day scenario for the Greater Dublin region in Ireland. Particularly, the presented approach includes the following novel contributions.

- We introduced a spatial choice model of the radiation type for selecting destinations of individual trips (Section 2.1), with interpretable parameters and a simple calibration scheme (Section 2.3).
- The model is applied for facility choice based on a dataset of points of interest and transitions statistics observed via geo-referenced Twitter messages and Four-square check-ins in Ireland (Section 2.2).
- Geographical layout of a social network observed in country-wide cell phone data is used as a proxy for modelling destination choice of the socially related trips (Section 2.4).
- The developed methodology is applied for destination choice in shopping, leisure and socially related journeys which account for major part of the traffic flows but are not available from traditional surveys.
- These activities are integrated into a realistic traffic scenario calibrated on the daily plans generated in accordance with a census of population, workplace locations, daily activities and departure times (Section 3), and validated on the measured traffic volume counts at major roads in Greater Dublin region (Section 4).

The paper is organised as follows. Section 2 gives an overview of spatial interaction approaches to urban mobility studies with particular focus on spatial choice modelling. We describe the developed adaptation of the radiation model in Section 2.1, which is then applied within a comprehensive framework of activity-based micro-simulation of traffic flows. This framework is built on the MatSim platform [4] and is described in Section 3. It uses a dataset of places of work locations to model commute flows. The necessary technical details on the datasets used in model development are given in Appendix 1. Our experimental results presented in Section 4 show that the proposed spatial choice model produces accurate estimates of the daytime traffic volumes at major roads. We highlight and interpret some characteristic traffic volumes patterns and compare this model to a naive baseline nearest neighbour method where all individuals choose a closest facility for their destination. We discuss the possible origins of the surprisingly accurate predictions in Section 5 which concludes the paper.

## 2. URBAN MOBILITY

Traditional transportation planning and forecasting frameworks stem from travel surveys on origin-destination flows and apply gravity laws [28], intervening opportunities [24], competing destinations [13] or an overarching constrained entropy maximisation framework [27] to investigate the trip distribution. A more flexible approach using activity-based models, focuses on modelling travel demand based on the activities that people need to perform in the course of a



**Figure 2: A probability of observing a commute trip, a phone call or a check-in displacement longer than  $x$  kilometers.**

day. This framework is usually implemented in a micro-simulation system where each agent is assigned an activity chain and performs destination choice in the context of this agenda. Meanwhile, the mobility of individuals is still relatively unexplored within these frameworks. Empirical evidence confirms that regular commute is a dominating mobility pattern [23], which also governs occasional fluctuations as people tend to arrange their travel plans by considering accessibility and convenience with regard to their primary locations such as home and work. This rational paradigm and availability of digital footprints opens new ways to enrich activity-based models in transportation modelling and urban planning. Facility choice thus becomes a key element of the model performance.

Attempting to predict the locations where people travel for work, recreation, shopping and to live is a significant challenge with a long research history. Both professionals and academics have carried out considerable work over the last century in dealing with this challenge and many positive findings have emerged. Factors that are taken into consideration regarding the choice of location include: travel distance and time, size of the store or facility, range of products or services and overall quality and price considerations. Models of estimating residential location choice include various logit models [2, 16, 10]. In modelling the choice of leisure facilities some recent developments are based on hollow space time prisms which are derived from leisure trip length statistics [14].

### 2.1 Radiation model of spatial choice

The radiation model [22] is inspired by the theory of intervening opportunities [24] and applies emission-absorption ideas to compute probabilities of interactions for a set of origins and destinations of known capacities. It is a destination-constrained parameter free model where distance decay is replaced with rank-based decay similarly to intervening opportunities. We applied this idea at an individual level to derive a probability of choosing a particular facility from a set of facilities of the given type with known capacities.

In our model we assign every individual an emission threshold  $z^i$  which determines a minimum level above which a particular driving trip will become worthwhile. For example, in case of shopping destination choice process, an individual with a large threshold  $z$  who is planning a shopping trip would have high or perhaps very specific demands which would have to be overcome and so is less likely to be ab-

sorbed by a nearby facility or shop. We assume there is some preselected distribution which describes this demand,  $p(z)$ . As there is no information on which kind of shopping trip an individual plans to undertake, we consider that a particular location choice of an individual at location  $i$  is based on the probability  $P_{m_i}(z)$  that a maximum threshold drawn from  $p(z)$  after  $m_i$  repetitions is equal to  $z$ . Suppose that each possible destination facility at location  $j$  has a certain probability to satisfy that demand  $P_{n_j}(z)$ , which is given by a maximum threshold extracted from  $p(z)$  after  $n_j$  repetitions, where  $n_j$  is the capacity of a facility at  $j$ . We must also account for the probability that none of the intervening facilities could absorb the traveller  $P_{s_{ij}}(<z)$  where  $s_{ij}$  is the total facility capacity in a circle centred on  $i$  of the radius equal to the distance between locations  $i$  and  $j$ . Then, the probability that a person at location  $i$  with a demand threshold  $m_i$  makes a trip to a facility at  $j$  with capacity  $n_j$  and no other closer facility, is given by

$$P(1|m_i, n_j, s_{ij}) = \int_0^\infty dz P_{m_i}(z) P_{s_{ij}}(<z) P_{n_j}(z), \quad (1)$$

We perform the integral in a similar fashion to the radiation model [22] and find

$$P(1|m_i, n_j, s_{ij}) = \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}. \quad (2)$$

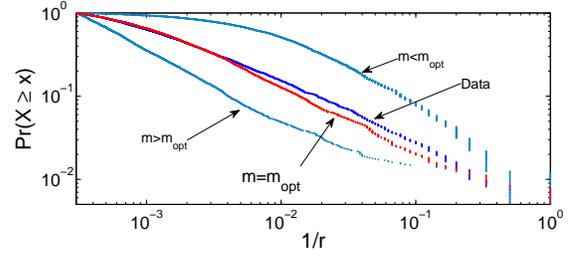
In our adaptation we aimed at a model where an unknown distribution  $p(z)$  of the demands of individuals deciding to commence a car trip can be integrated out.

## 2.2 Mobility data

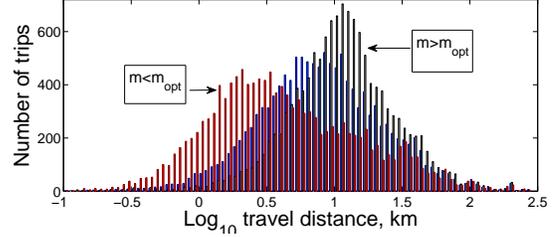
The movement dataset we used to model shopping, leisure and recreation trips is a combination of Twitter data collected in [20] and an Irish subset of the Foursquare dataset described in [8]. It contains a total of 107218 check-in events posted by 5287 unique users. Characteristic trip lengths contained in this dataset as compared to commute distances and call lengths are presented in Figure 2. Other summary statistics plots including temporal descriptors of users activities and their mobility are presented in Figure 5. We use this data to assess the parameters of the characteristic trip length for non-working activities.

## 2.3 Facility choice and parameter fitting

In contrast to the original radiation model where the inputs are the known populations of the origin and destination, we have a quantity  $m_i$  which relates to the choice of facilities in a region. Good quality public datasets on facility capacities are not readily available. We can make some estimate of our parameter  $m_i$  for a given region from the user-inputted data on OpenStreetMaps, but the overall coverage of this dataset is somewhat sparse. Instead we have found that we can substitute the  $m_i$  for a given location with an average facility choice  $m_{opt}$  for the entire region. We have devised a simple method to determine the optimum value for this parameter (Figure 3). The facilities are ranked according to the distance to the trip origin and for each facility we use the radiation model (2) to find the probability that a trip to the facility will be made. The data clearly show a long tail, confirming that longer trips to lower capacity facilities become increasingly unlikely. In trying to find a good value of  $m$  we see that if we set it too large,  $m > m_{opt}$ , this implies the high or specific demand which can not be satisfied by nearby facilities and indeed it can be seen in Figure 3 the probability



**Figure 3: Fitting the  $m$  parameter in the inverse rank cumulative probability plot,  $\log$ - $\log$  scale.**



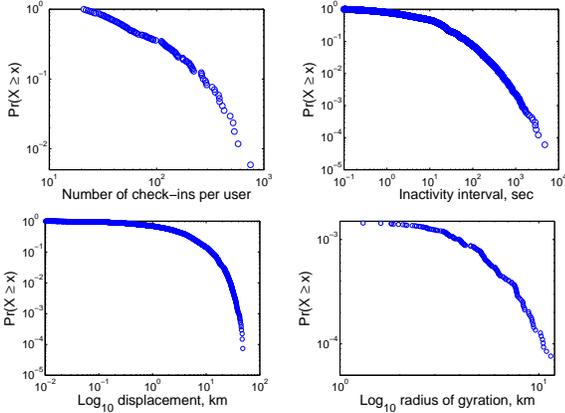
**Figure 4: Impact of  $m$  parameter on the trip length histogram. X-axis is in  $\log_{10}$  scale, km.**

to undertake a trip to a distant facility is significantly increased. Conversely, if  $m$  is too small, an individual is much more likely to make a trip to a nearby facility. The optimal value of the parameter  $m_{opt}$  is found by minimising the difference between the rank cumulative probability as found by the radiation model and the observed data. The trip length distributions resulting from a facility choice dictated by a radiation model are shown in Figure 4. Again, it is seen that the  $m < m_{opt}$  results in a shorter average path length, and conversely for  $m > m_{opt}$ . The optimal  $m_{opt}$  reproduces the trip length distribution which we find from a database of check-ins. An example of the theoretical analysis of the trip length distribution under generic multiplicative spatial choice models can be found in [25].

## 2.4 The geography of social networks

Empirical evidence [3, 21, 9] suggests the importance of social influence on the formation of atypical patterns of mobility. People visit family members or friends, and join them in recreation, leisure, tourism or shopping trips. It was observed that a probability of befriending a person is inversely proportional to the number of closer people, i.e. a spatial rank of the person [18]. Social networks also possess distinct community structure which often show geographical patterns both at inter-city [11] as well as intra-city scales [26]. One can use the characteristic distances and geographical layout of these interactions as a proxy for socially related travel such as journeys to visit friends and family.

The geographical layout of the major communities detected in a cell phone communication network in Greater Dublin area is presented in Figure 6 (taken from [26]). The community structure is clearly influenced by the underlying geography. Given that it is much more likely to observe social links between members of a community than across different ones, we have simulated a social network for the population of the agent which reproduces the characteristic link length distribution, node degree and community struc-



**Figure 5: Empirical CDF for the observed check-ins dataset show lack of power law tails due to limited sampling time, relatively low number of messages registered per user, and bounded geographical area of observation.**



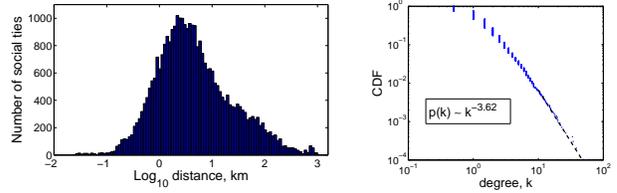
**Figure 6: Geographical layout of the major communities detected in a cell phone communication network in Greater Dublin area [26].**

ture which we find in the communication network. We will present further details on this social network generation algorithm elsewhere, and instead show the resulting statistics of the network in Figure 7.

### 3. TRAFFIC SIMULATION

Agent-based micro simulation is an effective way to model and predict traffic. In this approach, each agent is considered as an individual with an ability to make their own decisions and manage their daily activities to get the greatest return. MATSim [4] and SUMO (Simulation for Urban MObility) [6] are two examples of software frameworks implementing agent-based traffic simulations.

In MATSim each agent is assigned a plan which represents the desires of that agent for the day, for example, one desire is the departure time for work. The plan is altered through different iterations of the simulation in order to maximise an individual agent’s utility score. Travelling is seen as having a negative or neutral score whereas spending time at home has a positive score. Each iteration tries to minimise travel time to increase the overall utility score. The iterations should continue until the system has reached a relaxed state, known as a Nash Equilibrium where future iterations will produce little improvement in the utility scores



**Figure 7: A histogram of a social tie length (left) and a node degree (right) of the simulated social network.**

of agents. The variables which MATSim can alter during iterations are the departure time, the route choice and optionally, the location choice for secondary activities, such as leisure and shopping [14]. Route choice is determined using the A-Star algorithm [17], while time choice is achieved using local random mutation [5].

Activity chains, which represent the sequence of activities undertaken by individuals, are a key input for micro simulations. Generally they are derived from data recorded from travel surveys and reveal patterns of activities that people carry out on a normal day. For example, the percentages of people who travel to a shop immediately after work can be calculated and integrated into traffic simulations in order to predict which shopping locations will receive which portions of the road traffic. MATSim has been used by Horni et al. [14] in this way. By combining data from the Swiss census regarding work, education and home locations of citizens, with a activity chains collected by the Swiss National Travel Survey, traffic flows were produced for an average day. Originally MATSim employed an entirely time based utility function to calculate where individuals could travel within the time allocated for shopping trips, however this was discovered to be insufficient. Therefore, the model was extended to consider further variables such as shop size or the density of shops in a given area [14].

### 3.1 Implementation

We use MATSim to generate a traffic simulation for the Greater Dublin Area and compare two location choice models in the simulating traffic flows. One approach considers a nearest neighbour algorithm while the other uses a variant of the radiation model detailed in Section 2.1. MATSim has specific data requirements, including the road network and agent plans for the study area. Below, the details of the how this data was prepared for the simulation of traffic in the Greater Dublin Area are provided.

#### 3.1.1 Network

MATSim requires a network consisting of nodes and links. The nodes represent road intersections while the links are the road segments joining these intersections. Using tools provided by MATSim, the OpenStreetMap(OSM) road network for Dublin was extracted and transformed into the appropriate structure. All roads in an area of approximately 200kms squared around Dublin City were extracted from OSM, Figure 1. Additionally, all major roads (national routes and motorways) in Ireland were obtained. Additional information provided by OSM including the speed limits, class of road and type of road were also obtained. This permits the simulation to determine the flow capacity of road segments which is used in route choice.

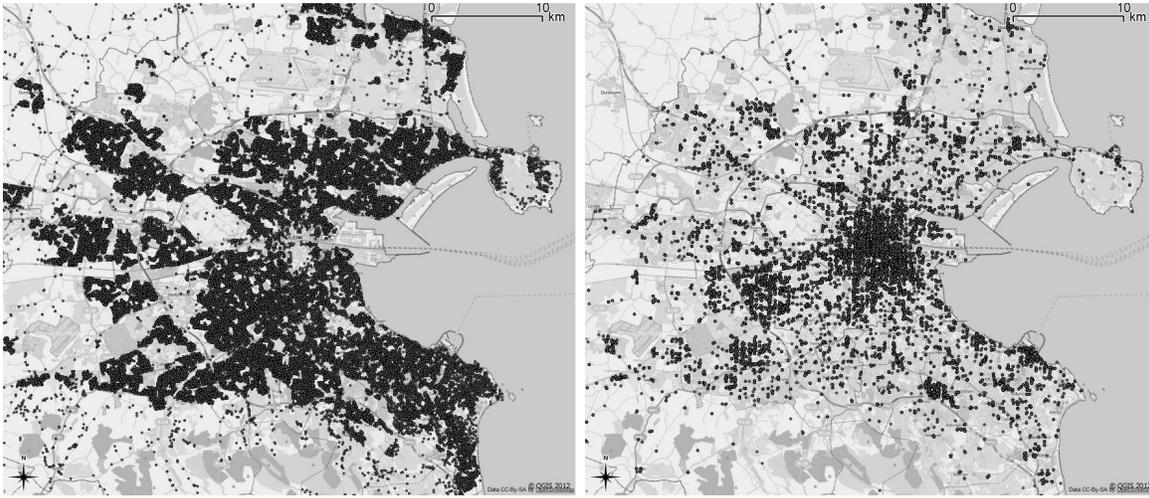


Figure 8: Home (left) and work (right) locations.



Figure 9: A close up view on home (left) and work (right) locations highlight a typical segregation of residential areas and industrial zones.



Figure 10: Locations of facilities in the area. The estimated capacities are denoted by circle size (largest correspond to major shopping malls).

### 3.1.2 Population and Demand

MATSim also requires a population, which it will model. The population represents the home and work locations of individuals within the study area. The locations are shown in Figure 8. Figure 9 highlights the contrast between a residential and industrial area. Additionally each individual needs to have a plan, which consist of the desired activities they will perform during the day. The plans or activity chains which include the sequence and duration of activities represent the demand on the network. Further details on the datasets used within the micro simulation framework are given in Appendix 1. One important aspect of creating the demand is the choice of location where activities will occur.

### 3.2 Facility choice implementation

The location of various activity types (schools, gyms, pubs, restaurants, shops, etc.) were extracted from public datasets including the OSM and a points of interest database of an in-car GPS navigator. Additionally, the capacity of each of these facilities was estimated. The resulting facilities are shown in Figure 10 where the size of the circle on the map represents the capacities. This formed the input for determining the location choice during the generation of the day plan for each agent. The plans were assigned according to the survey as described in Appendix 1 and contained the following activity types: school/education, shopping, personal business, visiting family/friends, social/entertainment, sport/leisure, and doctor/medical facility. The radiation model was applied for each individual choice over all alternatives amongst the facilities of a given type. That is, the same model was applied both at a strategic choice (a school's location) and a tactical choice level (a pub or a restaurant). For a social visit, a location to visit was assigned by sampling a home or work location of a friend from the simulated social network (Section 2.4). In addition to the radiation model for location choice described in Section 2.1, a nearest neighbour location choice was used as a naïve baseline approach. This model randomly selects a facility within a 4km threshold of the agents' current or future location. A sample of 50,000 individuals was randomly selected from the available data for home and work locations. In total 50 route replanning

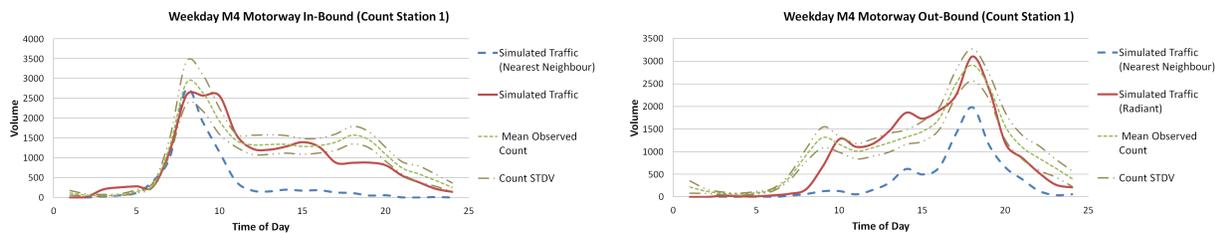


Figure 11: Traffic volume counts and simulation results for location 1.

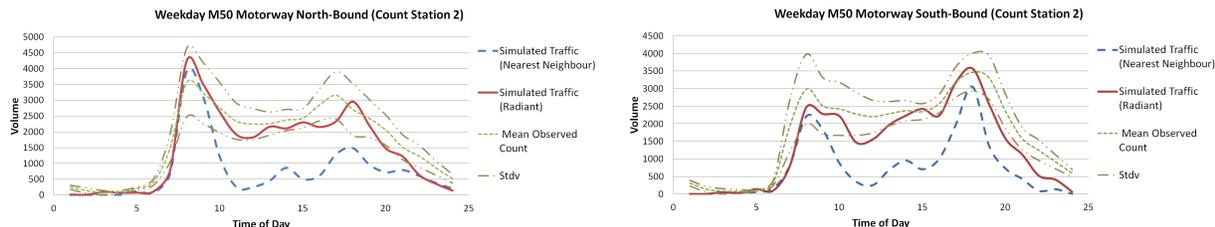


Figure 12: Traffic volume counts and simulation results for location 2.

iterations of the simulation were used.

MATSim outputs several pieces of data which are useful to assess the effectiveness of the simulation. Firstly an animation which shows the movement of the 50,000 agents over the road network can be rendered to assess traffic volumes at different periods of the day. The total distance travelled and trip duration of individuals is also produced. Finally, count data which shows the number of cars passing each road segment (link on the network) for each hour of the day is provided and served as a key validation measure for assessing the model performance.

## 4. RESULTS

The count data obtained with a simulation is compared to observed count information provided by the National Roads Authority (NRA) in Ireland which counts traffic using hardware embedded in road surface at specific locations. We have aggregated hourly average volumes for a typical working day over a summer period of 2006. Figure 14 shows a summary of the volume of cars using each road segment on the network during a 24 hour period returned by the simulation with radiation location choice. Not surprisingly, the motorway (M50 motorway) which surrounds the city sees the highest volume of traffic and so is assigned the darkest colour. Figure 14 also shows the physical location of all NRA count stations used to validate the simulation and highlights the three (labelled 1 - 3) that are presented in the paper in Figures 11-13. These were chosen due to their diverse geographic locations. The count information for these count stations was extracted from the data returned by MATSim. As 50,000 agents represents approximately 10% of vehicular traffic, the counts were scaled appropriately. The results were plotted alongside the mean NRA observed count data which were calculated by averaging the count data for weekdays (Tuesday to Thursday) from the published count statistics.

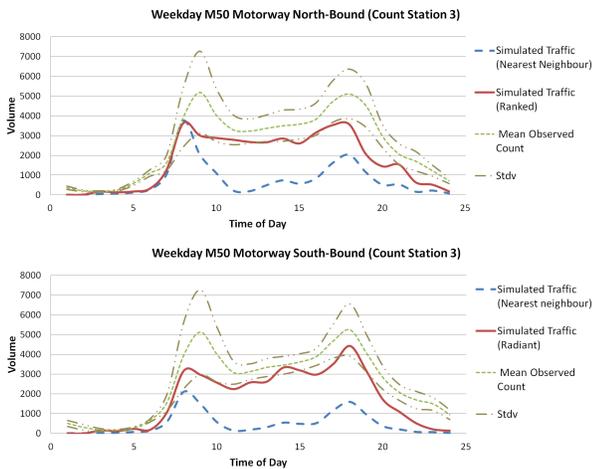
Figures 11 to 13 show the count data for each count station. All the graphs emphasise the importance of location choice for secondary activities. The nearest neighbour approach (dashed line) successfully detects the time of day that

the morning and evening peak in traffic occurs. This is due to the fact that home and work locations are obtained from census data. For the remainder of the day when secondary activities are occurring, the nearest neighbour model significantly underestimates the volume of traffic as individuals fail to travel for better opportunities and instead select activity locations which are in close proximity to them.

This is in contrast to the radiation model (solid line) which produces accurate count data throughout the day. Significantly, the volumes at the morning and evening peaks occur within 2 standard deviations of the mean observed counts for traffic on the M50 motorway (Figures 12,13). Similarly for the remainder of the day, the volume closely follows the mean observed count data. Figure 11 shows the count data for station 1 which is on the M4 motorway that connects Dublin to cities in the West of Ireland. Here, the peak for the morning out-bound traffic appears later than the observed mean values and the evening peak for in-bound traffic appears absent. This anomaly can be explained by the experiment set-up in which only individuals working in Dublin are considered. The simulation does not capture those that live in the Greater Dublin Area and work outside the city. Therefore this road is underused in the morning for out-bound commuter traffic and likewise in the evening for in-bound traffic.

## 5. CONCLUSIONS

It is not uncommon to observe the accuracy of models and the forecasts of volumes within 40% interval of the measured flows [12], and the observed fit can be considered as surprisingly good for a generic approach undertaken in this study. A major impact on the quality of the results is due to the amount, high detail and spatial resolution of the home to work data available for the region. Nevertheless, the newly introduced universal radiation spatial choice model was shown to perform superior to the nearest facility choice and was able to reproduce midday traffic volumes at a variety of major roads. The exact geography of social ties makes an essential contribution to its performance. We will study the influence of model components on the traffic system in



**Figure 13: Traffic volume counts and simulation results for location 3.**

more detail. Particularly, we are interested in the impact of stratification effect that emerges in the coupled consideration of mobility and social influence on facilities choice [15]. Also, the simulated volumes will be increased to the levels at which the impact of congestion on the route and destination facility choice can not be neglected.

## Acknowledgments

Research presented in this paper was funded by a Strategic Research Cluster grant (07/SRC/I1168), Stokes Lectureship award by Science Foundation Ireland under the National Development Plan, and the Complexity-NET programme Complexity in Spatial Dynamics (COSMIC). The authors gratefully acknowledge this support. We would like to thank Fergal Walsh for producing Figure 6 and Caroline Treacy for assistance in spatial data handling.

## Appendix 1.

The population data was obtained from the Irish National Census which is conducted every five years. The most recent census was conducted in 2011, however as the data is still being collated, the 2006 Census was used for generating the Dublin Simulation. POWCAR (Place of Work - Census of Anonymised Records) is a subset of the full census which provides the home location and work location of individuals, the mode of transport used to commute and the time at which individuals leave their residence. The home location is anonymised by giving it at an electoral division level while the work location is presented a 250 metre grid level. The time of departure is presented as discrete 30 minute intervals for the morning period and several modes are encoded in the means of transport.

To simulate movement within the Greater Dublin Area, individuals whose place of work is within the Dublin are extracted. Furthermore, only those who use a private car or van to get to work are considered. For finer grained location data, we assign each individual in the POWCAR dataset buildings to represent their home and work locations. GeoDirectory, a commercial database, which contains the location of every building in Ireland was used. The database contains the coordinates of buildings, the electoral

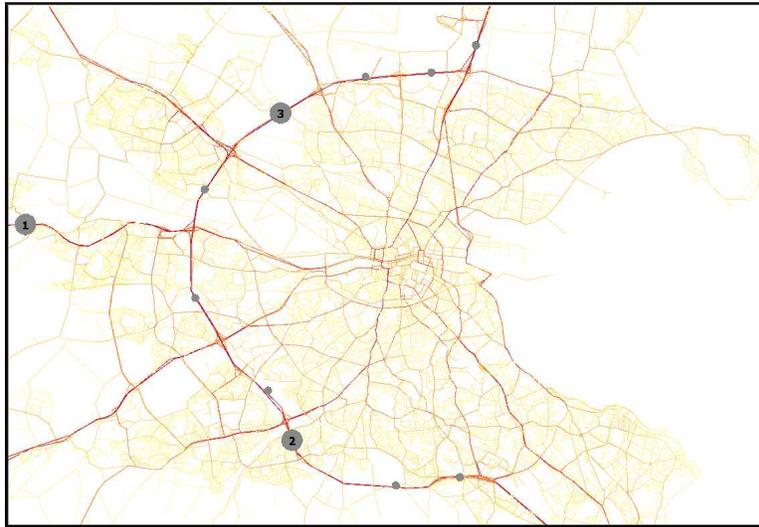
division they are in and the class of building (commercial, residential or both). Using this database, each individual is assigned a random residential building in their electoral district and a random commercial building with a 250 metre buffer of the work location declared in the POWCAR dataset. This data was combined with the departure time information. For the discrete values, a random time instant in the 30 minute departure segment was selected.

The demand is represented by activity chains and day plans collected via The Irish National Travel Survey (INTS) [1]. This Survey was carried out in 2009 as part of the Quarterly National Household Survey. Over 7000 participants were randomly selected and issued with a travel diary to record all journeys for a period of twenty four hours on a day that was allocated to them. The information gathered included journey origin and destination type (home, work, school, etc.), time of departure and arrival, mode of transport, purpose of trip, distance travelled and the time of each journey.

Activity chains, with durations were extracted from the INTS and relative frequencies of all travel sequences were calculated. This enabled a probability to be applied to each one so that for each individual in the POWCAR dataset, a day plan was generated. Once the sequence of events is determined, the duration of the activities needs to be defined by randomly selecting from all of the durations associated with the specific activity chain that has been chosen.

## 6. REFERENCES

- [1] *National Travel Survey Report*. Central Statistics Office, Government of Ireland, 2009.
- [2] J. Abraham and J. Hunt. Specification and estimation of nested logit model of home, workplaces, and commuter mode choices by multiple-worker households. *Transportation Research Record: Journal of the Transportation Research Board*, 1606(-1):17–24, 1997.
- [3] K. W. Axhausen. Social networks, mobility biographies, and travel: survey challenges. *Environment and Planning B: Planning and Design*, 35:981–996, 2008.
- [4] M. Balmer, K. Meister, M. Rieser, K. Nagel, and K. Axhausen. *Agent-based simulation of travel demand: Structure and computational performance of MATSim-T*. ETH, Eidgenössische Technische Hochschule Zürich, IVT Institut für Verkehrsplanung und Transportsysteme, 2008.
- [5] M. Balmer, B. Raney, and K. Nagel. Adjustment of activity timing and duration in an agent-based traffic flow simulation. *Progress in activity-based analysis*, pages 91–114, 2005.
- [6] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz. SUMO - simulation of urban mobility: An overview. In *SIMUL 2011, The Third International Conference on Advances in System Simulation*, pages 63–68, Barcelona, Spain, October 2011.
- [7] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.
- [8] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui. Exploring millions of footprints in location sharing services. In *ICWSM*, 2011.



**Figure 14: The modelled traffic volumes in Greater Dublin region and locations of counters used for model validation.**

- [9] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 1082–1090, New York, NY, USA, 2011. ACM.
- [10] N. Eluru, C. Bhat, R. Pendyala, and K. Konduri. A joint flexible econometric model system of household residential location and vehicle fleet composition usage choices. *Transportation*, 37(4):603–626, 2010.
- [11] P. Expert, T. S. Evans, V. D. Blondel, and R. Lambiotte. Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences*, 108(19):7663–7668, May 2011.
- [12] B. Flyvbjerg, M. K. Skamris Holm, and S. L. Buhl. Inaccuracy in traffic forecasts. *Transport Reviews*, 26(1):1–24, 2006.
- [13] A. S. Fotheringham. A new set of spatial-interaction models: the theory of competing destinations. 1(15):15–36, 1983.
- [14] A. Horni, D. Scott, M. Balmer, and K. Axhausen. Location choice modeling for shopping and leisure activities with matsim. *Transportation Research Record: Journal of the Transportation Research Board*, 2135(-1):87–95, 2009.
- [15] A. Lawlor, C. Coffey, R. McGrath, and A. Pozdnoukhov. Stratification structure of urban habitats, June 2012. Pervasive Urban Applications workshop (PURBA'12) at PERVASIVE'2012.
- [16] B. Lee and P. Waddell. Residential mobility and location choice: a nested logit model with sampling of alternatives. *Transportation*, 37(4):587–601, 2010.
- [17] N. Lefebvre and M. Balmer. *Fast shortest path computation in time-dependent traffic networks*. ETH, Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme, 2007.
- [18] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33):11623–11628, 2005.
- [19] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo. A tale of many cities: universal patterns in human urban mobility. *arXiv:1108.5355v4 [physics.soc-ph]*, 2011.
- [20] A. Pozdnoukhov and C. Kaiser. Space-time dynamics of topics in streaming text. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, LBSN '11, pages 8:1–8:8, New York, NY, USA, 2011. ACM.
- [21] A. Sadilek, H. Kautz, and J. P. Bigham. Finding your friends and following them to where you are. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12, pages 723–732, New York, NY, USA, Feb. 2012. ACM.
- [22] F. Simini, M. C. Gonzalez, A. Maritan, and A.-L. Barabasi. A universal model for mobility and migration patterns. *Nature*, (484):96–100, 2012.
- [23] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, Feb. 2010.
- [24] S. A. Stouffer. Intervening opportunities: A theory relating mobility and distance. *American Sociological Review*, 5(6):845–867, 1940.
- [25] D. Veneziano and M. C. Gonzalez. Trip length distribution under multiplicative spatial models of supply and demand: Theory and sensitivity analysis. *CoRR*, abs/1101.3719, 2011.
- [26] F. Walsh and A. Pozdnoukhov. Spatial structure and dynamics of urban communities, June 2011. The First Workshop on Pervasive Urban Applications (PURBA).
- [27] A. G. Wilson. *Entropy in Urban and Regional Modelling*. Pion, London, United Kingdom, 1970.
- [28] G. K. Zipf. The p1 P2/D hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6), 1946.