APPROXIMATION OF PEDESTRIAN EFFECTS IN URBAN TRAFFIC SIMULATION BY DISTRIBUTION FITTING

Andreas D. Lattner, Jörg Dallmeyer, Dimitrios Paraskevopoulos
Goethe University Frankfurt
P.O. Box 11 19 32, 60054 Frankfurt, Germany
[lattner|dallmeyer|paraskev]@cs.uni-frankfurt.de

Ingo J. Timm
University of Trier
54296 Trier, Germany
ingo.timm@uni-trier.de

KEYWORDS

Multimodal Traffic Simulation, Pedestrian Impact Approximation, Distribution Fitting

ABSTRACT

For a proper simulation of urban traffic scenarios, besides cars other road users, namely bicycles and pedestrians, have to be modeled. In scenarios where a whole city is modeled, a detailed actor-based simulation of pedestrians leads to expensive extra computational load. We investigate to what extent it is possible to capture traffic effects imposed by simulated pedestrians and then perform simulations without pedestrians. We propose to collect information about pedestrian impacts in a simulation with pedestrians, estimate underlying probability distributions and finally, use a simplified model where only these effects are generated probabilistically. We investigate two approaches – a best-fit distribution fitting and a histogram-based distribution approximation – using synthetic data as well as simulated traffic scenarios. The experiments show that using the proposed approximations can lead to similar average cars' travel times.

INTRODUCTION

The continuous increase of traffic has led to the situation that in many areas congestions occur frequently. Traffic as a field of research has been addressed by many researchers and among the goals of such studies are to get insights about traffic effects as well as to investigate strategies to improve the situation. The modeling of traffic dates back to the beginning of the last century (Greenshield, 1935). One famous model that has been widely used for the simulation of traffic – mainly on motorways – has been introduced by Nagel and Schreckenberg. The Nagel-Schreckenberg model (NSM) is a time and space discrete model which has been used to analyze certain traffic effects (Nagel and Schreckenberg, 1992).

For a proper simulation of urban scenarios, besides cars other road users, namely bicycles and pedestrians, have to be modeled. Although there exist some simulation systems which aim at the simulation of urban scenarios, they usually either focus on a detailed reproduction (including visualization) of rather small scenarios or they do not take into account requirements of multimodal traffic with varying behaviors and characteristics (like velocities and acceleration) of different road users.

One particular model for the simulation of multimodal traffic including different cars (passenger cars and trucks), bicycles, and pedestrians has been introduced by (Dallmeyer et al., 2011). In contrast to the NSM, it uses a continuous space model and thus, provides means for the representation of different road users' characteristics. In this work we use and extend this simulation system.

In an earlier study, we have investigated the impact of pedestrians to urban traffic and the results indicated that actually simulating pedestrians has influence to car traffic with a tendency to increase car travel times if more pedestrians are in transit (Dallmeyer et al., 2012b). Therefore, an analysis of the effects including pedestrians in the simulation is out of scope of this paper. Simulating a whole city with potentially one million residents or more leads to a large number of road users including a significant fraction of pedestrians. A detailed actor-based simulation of all these pedestrians leads to expensive computational load.

As in many studies the focus is set on strategies of how to arrange traffic in a way that motorized traffic is more efficient, we investigate in this paper if it is possible to capture effects imposed by pedestrians and then perform simulations without actually simulating pedestrians. We propose to first collect information about pedestrian impacts in a simulation with pedestrians, estimate underlying probability distributions of these impacts and finally, use a simplified simulation model where only these effects are generated probabilistically.

RELATED WORK

Traffic simulation is done for different scenarios ranging from the simulation of whole countries (Voellmy et al., 2001) to high-fidelity simulation of small areas (Bönisch and Kretz, 2009). The simulation of pedestrians has also been addressed by different works including high-fidelity models like the social force model Helbing and Molnár (1995), macroscopic models like Hughes (2003) or microscopic models built on cellular automatons like Blue and Adler (2001). Our work focuses on a level of detail enabling simulation of whole cities and still regarding multi modality (Dallmeyer et al., 2012a).

Techniques of distribution fitting are used to establish a probability distribution function F from a sample (a set of observations) (e.g., (Law, 2007)). Additionally, distri-

Proceedings 26th European Conference on Modelling and Simulation ©ECMS Klaus G. Troitzsch, Michael Möhring, Ulf Lotzmann (Editors)
ISBN: 978-0-9564944-4-3 / ISBN: 978-0-9564944-5-0 (CD)

butions' parameters, e.g., mean and standard deviation of a normal distribution, can be assigned. In Goodness-of-Fit tests, statements as to whether a sample might have been drawn from a certain distribution are made. In these tests, the null hypothesis is that the considered values have been generated by the distribution function F. Common tests are the Chi-Square test, the Kolmogorov-Smirnov test, the Anderson-Darling test and the Poisson-Process test; for more information about these tests see, e.g., (Law, 2007, p. 340-353). Regarding the Chi-Square Test, it is necessary to divide the data in adjacent intervals. In the next step, respecting the specific distribution function, the probability for several values falling into each interval has to be determined. Finally, after computing the test statistic χ^2, the null hypothesis will be rejected, if differences appear to be too large and thus, the probability of a type 1 error is below a threshold α.

Distribution fitting is widely applied to input data in the context of simulation in order to set up probability distributions, e.g., in manufacturing or arrival patterns of patients in healthcare (cf. Kuhl et al. (2007)). To the best of our knowledge the approximation of pedestrian effects in traffic simulation by distribution fitting has not been addressed so far.

TRAFFIC SIMULATION SYSTEM

The traffic simulation system used in this work is MAINS^{2}IM (MultimodAl INnercity SIMulation). It can automatically build up simulation models using cartographical material from *OpenStreetMap*[1] (OSM). A study about OSM map quality can be found, e.g., in Haklay (2010). At first, an arbitrary area is extracted from an OSM file. The included geoinformation is separated into logical layers for different types of geometries. The system generates a graph data structure from a layer representing roads and incorporates information from additional layers like, e.g., the areas of cities. Several analysis and correction steps are performed in order to build a valid graph (e.g., putting nodes at intersections of edges, considering bridges and tunnels and determining roundabouts). The system is built on a geographical information system on the basis of *GeoTools*[2].

Mixed traffic is simulated under usage of models for cars (passenger cars, trucks and buses), as well as bicycles and pedestrians. The models are discrete in time with (default) time steps of 1s per simulation iteration. Movement is modeled continuously in space. Pedestrian movements are simulated using two models: one for sidewalk movements on a graph structure, and one for crossing roads at pedestrian crossings. Interaction of pedestrians with other road users takes place when pedestrians cross roads – either at a pedestrian crossing or at some other position of a road – and other road users need to brake to avoid collisions.

The simulation system is implemented in Java and can be used on an off-the-shelf PC. The system is capable for the simulation of the traffic of whole cities. Further information about MAINS^{2}IM can be found in (Dallmeyer et al., 2012a,b, 2011; Lattner et al., 2011) and on the website `www.mainsim.eu`.

[1] http://www.openstreetmap.org

[2] http://www.geotools.org

PEDESTRIAN EFFECT APPROXIMATION

The pedestrian model used in MAINS^{2}IM is introduced in (Dallmeyer et al., 2012b). This work analyzes the interaction between pedestrians and road traffic. These interactions occur whenever a pedestrian crosses a road or uses a traffic light and a car needs to slow down because of this action. Pedestrians cross roads when a perceptual sufficient gap in traffic exists. The perceptions may be faulty and lead to pedestrians crossing the road without sufficient gaps. Traffic lights and crosswalks are stored in nodes of the simulation graph. A pedestrian crossing a node blocks the affected edges at this position. The impact is measured by capturing how often and how long pedestrians interfere with other road users. This section discusses two methods to approximate the influences of pedestrians on urban traffic. Both approaches use observed data from simulations with pedestrians as input. The approaches are independent of specific data but can be used to generate approximations of probability distributions from provided samples.

Best-fit Distribution Fitting

The first approach is the best-fit distribution fitting. In this approach, the samples are passed to a distribution fitting algorithm based on maximum likelihood estimation for distributions' parameters. We use the *MASS* package of R Project in order to fit the distributions (Venables and Ripley, 2002) (also cf. (Ricci, 2005)). In our approach, we take into account Gaussian, exponential, and uniform distributions. For the latter case, no distribution fitting is used but minimal and maximal values of the sample are taken as parameters for the approximated distribution. Of course, the set of distribution types to be tested can be extended as desired.

In order to decide which distribution to use, a χ^2 test is performed for all distribution types. The range of the sample $s = (s_1, \ldots, s_l)$ is divided into n equidistant intervals with $n = 1 + log_2(l)$ (Sturges' formula). These intervals are used for the performance of the χ^2 tests and the fitted distribution with highest p value is then used as approximation (the lower the p value, the lower the probability having a type 1 error, i.e., selecting the alternative hypothesis assuming that the sample has not been generated by the given distribution although actually the null hypothesis is valid). Figure 1 illustrates the approach where the distribution fitting results for three samples from different distribution types are shown.

Histogram-based Distribution Approximation

As the best-fit distribution fitting approach only takes into account (a selection of) univariate distributions and

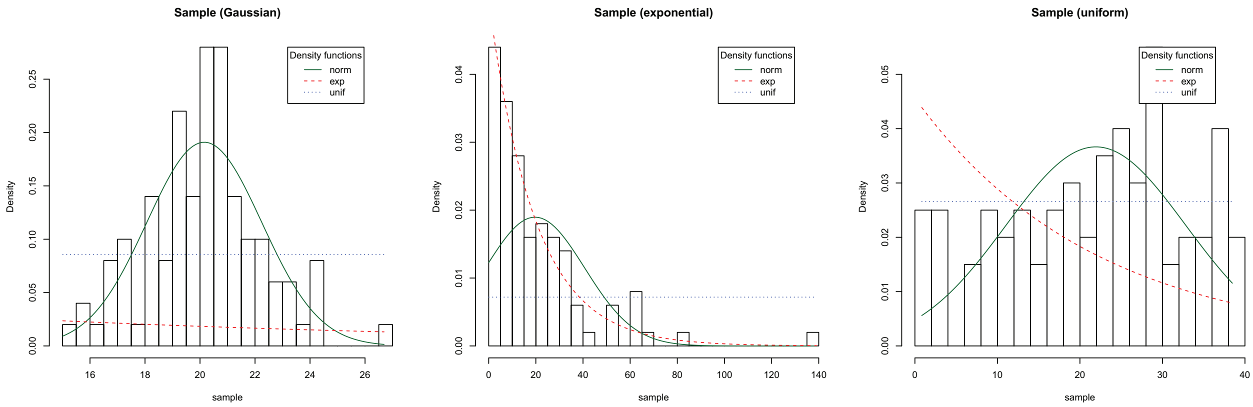


Figure 1: Distribution fitting results for three samples

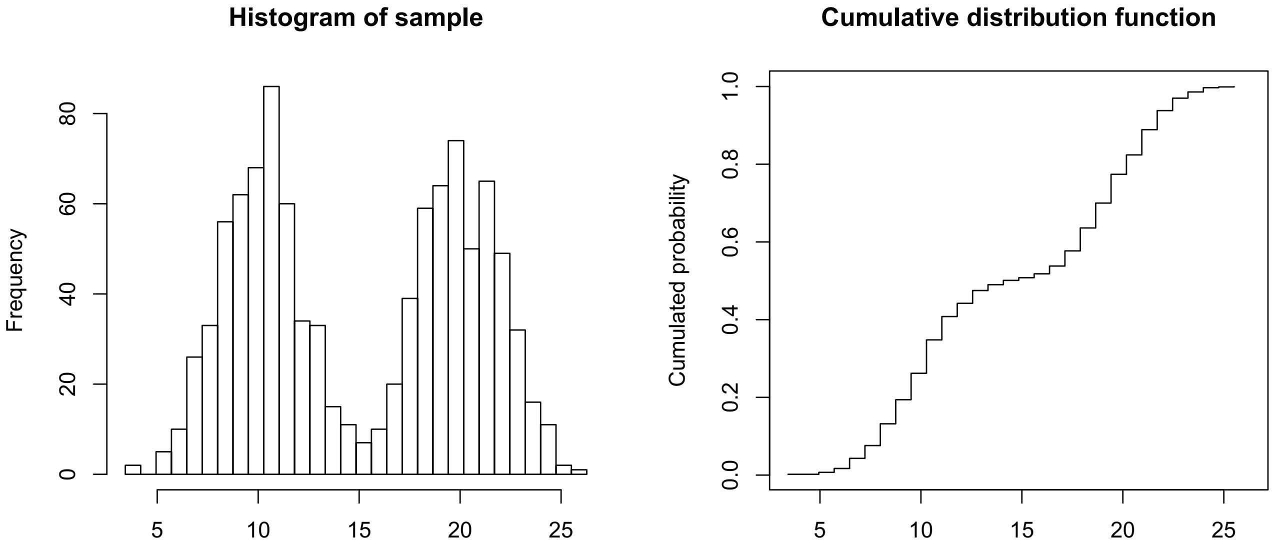


Figure 2: Illustration of histogram-based distribution approximation

as it might be that the probability distributions under consideration exhibit other characteristics, we also analyze a second approach in this paper. In histogram-based distribution approximation, an observed sample $s = (s_1, \ldots, s_l)$ is divided into n categories. While different ways to select n can be chosen, we use the following formula as it leads to a sufficient minimal number of categories and grows logarithmic with the sample size l: $n = max(10, round(10 \times log_{10}(l)))$. The range of the values is split into n intervals and frequencies are measured for each bin. The relative frequency is used to generate a cumulative distribution function (Figure 2). Similar to existing approaches, the inverse of the cumulative distribution function can be used for the generation of random numbers following this distribution. In this case, a uniform random number in the interval $[0, 1]$ is generated and the corresponding x value is identified. A random number is generated using a uniform random number generator in the interval $[x - x_{step}, x]$ with $x_{step} = ((max(s) - min(s))/n)$.

EVALUATION - SYNTHETIC DATA

In the first part of the evaluation, synthetic data using random samples from artificial probability distributions are used in order to evaluate the best-fit distribution fitting approach. An explicit generation of random samples from given distributions has the advantage that the underlying type of probability distribution as well as its parameters are known in advance. Therefore, it is possible to assess if the approach is able to identify the correct type of distribution. Furthermore, using parameter estimation, a statement about the difference of real and estimated values can be performed.

For evaluation, we have generated a number of samples with different sizes and using different probability distributions (Gaussian, exponential, and uniform). For each distribution, we randomly generate parameters and use these parameters to generate random samples of different sizes (20, 50, and 100 values) in order to find out how the approach works in dependence of the size of samples. For each distribution and sample size combina-

	Sample size: 20				Sample size: 50				Sample size: 100			
	Gauss	Exp.	Unif.	Recall	Gauss	Exp.	Unif.	Recall	Gauss	Exp.	Unif.	Recall
Gauss	874	0	126	87.40	983	0	17	98.30	1000	0	0	100
Exp.	155	841	4	84.10	35	965	0	96.50	1	999	0	99.90
Unif.	471	22	507	50.70	263	1	736	73.60	111	0	889	88.90
Prec.	58.27	97.45	79.59		76.74	99.90	97.74		89.93	100	100	

Table 1: Confusion matrices of the best-fit approach using synthetic data

tion, 1000 random values are generated. The generation of parameter values is done as follows (using uniform probabilities within the specified ranges):

- Gaussian distribution: Mean μ is chosen randomly as $\mu = 1 + unif(0.0, 20.0)$ and standard deviation σ is chosen randomly using $\sigma = unif(0.0, 0.2\mu)$.
- Exponential distribution: The rate λ is chosen using the same range for the expected value as in the Gaussian distribution: $\lambda = \frac{1}{1+unif(0.0,20.0)}$.
- Uniform distribution: For the uniform distribution, the minimal value (left limit) min_{unif} is chosen randomly with $min_{unif} = unif(1, 40)$ and the maximal value (right limit) with $max_{unif} = unif(min_{unif}, 40)$.

The function $unif(min, max)$ denotes a random number generator using a uniform probability distribution. For the generation of the samples we have used *R* (R Development Core Team, 2011). Interaction with *R* is done via the *rJava* library (Urbanek, 2011).

For the evaluation with synthetic data, we investigate in how many cases the correct type of distributions has been identified and capture how the actual parameters differ from the estimated ones. Table 1 shows the classification results for the three different sample sizes. The results are presented as confusion matrices where the lines represent the actual distribution and the columns the classification decision. If 20 values are used, the recall values for the Gaussian and exponential samples are above 84% and for the uniform distribution $\sim$ 50%. The precision values using the Gaussian, exponential, and uniform samples are approximately 58%, 97%, and 80%, respectively. It can be observed that the uniform samples are classified as Gaussian distributions in many cases. With increasing sample sizes, the misclassification rates decrease. Using samples with 50 or 100 values leads to recall values above 96% for the Gaussian and the exponential samples and to precision values above 97% for the exponential and uniform samples. Especially, the confusion among Gaussian and uniform distributed values is smaller if more values are used. For sample size 100, only 11.1% of the uniformly distributed samples are misclassified as Gaussian while all but one sample of the other two distribution types are classified correctly.

Table 2 presents the average p values of the χ^2 test and the differences to the original distributions' parameters, subdivided into the samples which have been classified correctly and incorrectly as well as the joint values ("total"). Additionally, it is shown in how many cases the p value of the χ^2 test was below $\alpha = 0.05$ for the different distribution types, indicating that the alternative hypothesis would be chosen, i.e., that it cannot be assumed that the sample was drawn by the corresponding distribution. The results show that on average the p values of the χ^2 test with the (true) distribution type are above 0.5 in all cases and that the average p values are smaller for the misclassified samples in comparison to the correctly classified ones. Average (absolute) differences between fitted and actual distributions' parameters are below 0.22 for Gaussian mean, below 0.16 for Gaussian standard deviation, below 1.88 for the inverse of the exponential rate ($1/\lambda$), below 0.46 and 0.43 for the minimal and maximal value of the uniform distribution. The α threshold comparisons show that the tests with the correct distribution type reject the null hypothesis in all cases in less than 5%. It also shows that with increasing sample sizes, the rejection rates of the wrong distribution types increase.

EVALUATION - TRAFFIC SIMULATION

In the second part of the evaluation, distribution fitting is applied to the results of simulation runs with simulated pedestrians. The simulation is done in the area of Hanau, shown in figure 3.

Each simulation run starts with a settlement phase of 1,000 iterations, followed by a measurement phase of 50,000 iterations. The amount of road users is held constantly at 2,500 with 33% cars, 7% bicycles and 60% pedestrians. Whenever a pedestrian p crosses a road r, its position, the time period since the last crossing pedestrian on r, and the time duration the crossing took will be stored in an observance object for r. Crossing actions at nodes of the simulation graph are handled similarly. After a simulation run, the measured data is analyzed and probability distributions for crossing time intervals, crossing positions and crossing durations for each node and edge of the graph are estimated.

A second run is performed with identical copies of the used cars and bicycles of the former run, but pedestrians are now approximated by dummy pedestrians, influencing traffic using the estimated probability distributions.

The whole setting is repeated 100 times with different initial seed values for the random number generator. The quantity for a comparison of the different settings in

Sample	correct classif.			incorrect classif.			total			χ^2 test		
	avg. p	avg. diff. mean	avg. diff. sd	avg. p	avg. diff. mean	avg. diff. sd	avg. p	avg. diff. mean	avg. diff. sd	$p_{Gauss} <$ α (in %)	$p_{exp} <$ α (in %)	$p_{unif} <$ α (in %)
Gauss 20	0.69	0.22	0.15	0.56	0.21	0.17	0.68	0.22	0.16	0.60	100.00	15.70
Gauss 50	0.66	0.12	0.09	0.46	0.19	0.10	0.66	0.12	0.09	2.30	100.00	72.20
Gauss 100	0.62	0.09	0.06	n/a	n/a	n/a	0.62	0.09	0.06	3.30	100.00	99.80
	avg. p	avg. diff. $1/\lambda$		avg. p	avg. diff. $1/\lambda$		avg. p	avg. diff. $1/\lambda$		$p_{Gauss} <$ α (in %)	$p_{exp} <$ α (in %)	$p_{unif} <$ α (in %)
Exp. 20	0.62	1.88		0.35	1.70		0.58	1.85		24.40	3.10	77.30
Exp. 50	0.57	1.25		0.20	0.85		0.56	1.24		76.80	4.80	99.90
Exp. 100	0.55	0.86		0.01	0.06		0.55	0.86		98.80	5.00	100.00
	avg. p	avg. diff. min	avg.diff. max	avg. p	avg. diff. min	avg.diff. max	avg. p	avg. diff. min	avg.diff. max	$p_{Gauss} <$ α (in %)	$p_{exp} <$ α (in %)	$p_{unif} <$ α (in %)
Unif. 20	0.61	0.47	0.39	0.40	0.45	0.48	0.51	0.46	0.43	1.10	92.80	3.80
Unif. 50	0.56	0.21	0.21	0.36	0.17	0.19	0.51	0.20	0.20	6.30	99.80	3.90
Unif. 100	0.53	0.09	0.08	0.25	0.09	0.11	0.50	0.09	0.09	21.70	100.00	4.70

Table 2: p-values and differences to original distributions' parameters of the best-fit approach using synthetic data

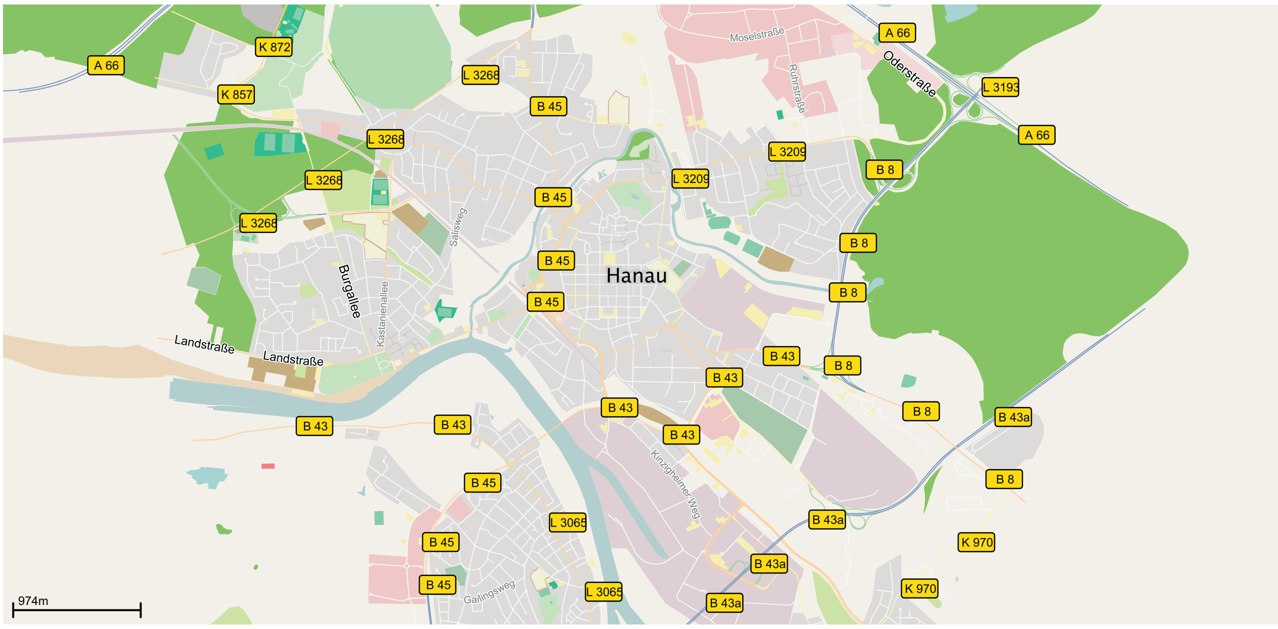
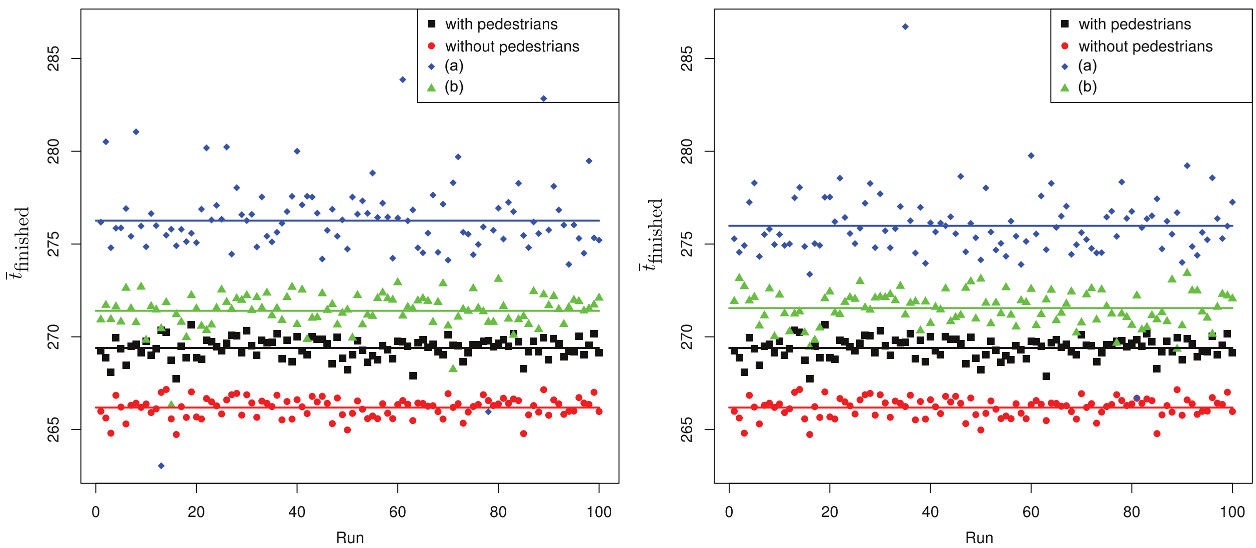


Figure 3: Map area for simulation (4,201 nodes, 5,758 edges, 548km total length of roads)

this experiment is the mean travel time for each setting. Whenever a car or bicycle leaves the simulation, its travel time will be recorded and averaged at the end of the run.

Two methods for estimation of the probability distributions are used: (a) A distribution taking into account all pedestrians crossing roads (blue, rhombus) and (b) one for those pedestrians which are crossing roads and are actually influencing traffic (green, triangle). A pedestrian p influences traffic whenever a road user ru has a distance smaller or equal to its velocity to p, because ru needs to brake in order to avoid a collision with p. Because the pedestrian dummies are set on roads without taking into account the current traffic situation, this might lead to a situation where road users have to brake strongly, because the dummies do not look for sufficient gaps, they just appear. This leads to an overestimation of travel times for road users. Method (b) does not overestimate as much, because the number of dummies is reduced to the number of pedestrians, who have in fact forced road users to brake. The estimated travel times overlap with the original travel times with simulated pedestrians (black, square) in some cases. In comparison to the runs with pedestrians, results of a simulation with no pedestrians and no dummies are shown (red, circle).

Figures 4(a) and 4(b) show the simulation results of the best-fit approach and of the histogram-based approach. It is obvious that pedestrians increase travel times of cars and bicycles. The simple method of calculating probability distributions for each pedestrian crossing roads and then setting dummies without taking care of traffic conditions (a) leads to an overestimation of travel times, as assumed. Method (b) performs better. In a few runs, (b) nearly reproduces the results of the runs with simulated pedestrians. Table 3 shows the mean travel times of cars and bicycles, averaged over all simulation runs. Apparently, method (b) leads to an overestimation of travel times ($\approx 2s$), which is not as great as the underestimation when pedestrians are not respected ($\approx 3s$). Approximations using the best-fit and the histogram-based approaches lead to similar mean travel times.



(a) Best-fit (b) Histogram based

Figure 4: Comparison of different setups with average travel times of road users $\bar{t}_{\text{finished}}$ $[s]$. Lines show mean values.

	$\bar{t}_{\text{finished}}$ $[s]$	
	Best-fit	Histogram based
(a)	276.26	275.98
(b)	271.40	271,55
with pedestrians	269.34	
without pedestrians	266.19	

Table 3: Mean travel times for cars and bicycles in different settings.

CONCLUSION

We have investigated two approaches for approximation of pedestrian impacts in urban traffic simulation. The first approach is based on distribution fitting using the distribution with the lowest expected error probability based on a χ^2 test. The second approach is a histogram-based approximation of a probability distribution. The realization of the best-fit approach – if implemented from scratch – needs more effort as specific details about all different distributions have to be taken into account and some technique for parameter estimation needs to be integrated. The histogram-based approach can be implemented with less effort as it is mainly based on counting how many values belong to each of the bins. In some contexts it might be more interesting to get information about what kind of distribution is expected to have generated observed data as provided by the best-fit approach. The histogram-based approach is only reflecting the distribution without providing any additional information.

The best-fit approach has been additionally evaluated using synthetic data. The evaluation has shown that when using sample sizes of 20 that Gaussian and uniform probability distributions are mistaken in many cases, while with increasing sample sizes the distribution types can be distinguished with higher accuracies.

The experiments approximating pedestrian effects in traffic simulation have shown that both approaches, best-fit distribution fitting and the histogram-based approach, exhibit similar mean travel times. The mean differences regarding the observed average travel times in comparison to the simulation with pedestrians are 2.06s and 2.21s. The approximation approaches are overestimating the mean travel times but the differences are smaller than not taking into account pedestrians at all (3.15s).

Using this approximation leads to an about two times faster simulation as detailed pedestrian behavior needs not to be computed. The proposed approach could also be used to extend other traffic simulation systems if the approximated effects of pedestrian travels are integrated.

For future work, it would be interesting to extend the best-fit approach and to take into account further distributions as well as multivariate distributions. Furthermore, a better selection of the "relevant" pedestrians who actually have impact to motorized traffic could lead to even smaller differences in the simulation results. It could also be investigated to what extent different numbers of bins – in the histogram-based approach as well as for the Goodness-of-Fit test – have influence on the results. It would also be interesting to address how more realistic data about pedestrian behavior could be collected – preferably in an automated way.

REFERENCES

Blue, V. J. and Adler, J. L. (2001). Cellular automata microsimulation for modeling bi-directional pedestrian walkways. *Transportation Research Part B: Methodological*, 35(3):293 – 312.

Bönisch, C. and Kretz, T. (2009). Simulation of pedestrians crossing a street. *Proceedings of the Eighth International Conference on Traffic and Granular Flow*, 8.

Dallmeyer, J., Lattner, A. D., and Timm, I. J. (2011). From GIS to Mixed Traffic Simulation in Urban Scenarios. In Liu, J., Quaglia, F., Eidenbenz, S., and Gilmore, S., editors, *4th International ICST Conference on Simulation Tools and Techniques, SIMUTools '11, Barcelona, Spain, March 22 - 24, 2011*, pages 134–143. ICST, Brüssel.

Dallmeyer, J., Lattner, A. D., and Timm, I. J. (2012a). *Data Mining for Geoinformatics: Methods and Applications*, chapter GIS-based Traffic Simulation using OSM. Springer. (accepted).

Dallmeyer, J., Lattner, A. D., and Timm, I. J. (2012b). Pedestrian simulation for urban traffic scenarios. *Proceedings of the 44rd Summer Simulation Multi-Conference (SummerSim'12)*. (submitted).

Greenshield, B. D. (1935). A study of traffic capacity. *Proceedings of the 14th Annual Meeting Highway Research Board 14*, 468:448–477.

Haklay, M. (2010). How good is volunteered geographical information? a comparative study of OpenStreetMap and ordnance survey datasets. *Environment and Planning B: Planning and Design*, 37(4):682–703.

Helbing, D. and Molnár, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, 51(5):4282–4286.

Hughes, R. L. (2003). The flow of human crowds. *Annual Review of Fluid Mechanics*, 35(1):169–182.

Kuhl, M. E., Lada, E. K., Steiger, N. M., Wagner, M. A., and Wilson, J. R. (2007). Introduction to modeling and generating probabilistic input processes for simulation. In *Proceedings of the 39th Winter Simulation Conference*, WSC '07, pages 63–76, Piscataway, NJ, USA. IEEE Press.

Lattner, A. D., Dallmeyer, J., and Timm, I. J. (2011). Learning dynamic adaptation strategies in agent-based traffic simulation experiments. In *Ninth German Conference on Multi-Agent System Technologies (MATES 2011)*, pages 77–88. Springer: Berlin, LNCS 6973, Klügl, F.; Ossowski, S.

Law, A. M. (2007). *Simulation Modeling & Analysis*. McGraw-Hill, 4th, internat. edition.

Nagel, K. and Schreckenberg, M. (1992). A cellular automaton model for freeway traffic. In *Journal de Physique I*, volume 2, pages 2221–2229. ISSN 1155-4304.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ricci, V. (2005). Fitting distributions with R. Release 0.4-21, `http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf`, accessed: 04/13/2012.

Urbanek, S. (2011). *rJava: Low-level R to Java interface*. R package version 0.9-0.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition.

Voellmy, A., Vrtic, M., Raney, B., Axhausen, K., and Nagel, K. (2001). *Status of a TRANSIMS implementation for Switzerland*. Swiss Federal Institute of Technology. `http://dx.doi.org/10.3929/ethz-a-004339673`.

AUTHOR BIOGRAPHIES

ANDREAS D. LATTNER received the Diploma (2000) and the doctoral degree (2007) in computer science at the University of Bremen, Germany. From 2000-2007 he was working as research scientist at the Center for Computing Technologies (TZI) at the University of Bremen. He works now as a postdoctoral researcher at the chair for "Information Systems and Simulation" at the Goethe University Frankfurt. His research interests include knowledge discovery in simulation experi-ments, temporal pattern mining, and multi-agent systems.

JÖRG DALLMEYER received the bachelor degree (2008) as well as the master degree (2009) in computer science at the Goethe University of Frankfurt am Main, Germany. Since December 2009, he has been a PhD student at the chair for "Information Systems and Simulation" at the Goethe University Frankfurt. His research interests are actor based simulation for the field of traffic simulation under consideration of multi modal traffic and the building of simulation systems from geographical information.

DIMITRIOS PARASKEVOPOULOS received the diploma degree (2010) in computer science at the Goethe University of Frankfurt am Main, Germany. Since February 2011, he has been a PhD student at the chair for "Information Systems and Simulation" at the Goethe University Frankfurt. His research interests are Process Mining and IT Risk Management.

INGO J. TIMM held various positions in research at University of Bremen, Technical University Ilmenau, and Indiana University Purdue University - Indianapolis (IUPUI) from 1998 until 2006 before being appointed full professor for Information Systems and Simulation at Goethe-University Frankfurt/Main. Since 2010, he holds a chair for Business Information Systems at University of Trier. Ingo Timm is speaker of the German SIG on Distributed Artificial Intelligence (FG-VKI) of German Informatics Society. His special interest lies on business information systems, intelligent assistance systems, (actor-based/multiagent-based) simulation, and knowledge-based support to simulation system.

Acknowledgement This work was made possible by the *MainCampus* scholarship of the *Stiftung Polytechnische Gesellschaft Frankfurt am Main*. Parts of this paper are results of the EDASim project (Hessen Agentur Project No.: 260/11-06), funded by the German State Hesse in context of the Hessen ModellProjekte ("LOEWE - Landes-Offensive zur Entwicklung Wissenschaftsökonomischer Exzellenz").