A Framework to Determine the Spatial Variation of the Optimal Paths

Andreas PARTUSCH[1], Gerhard NAVRATIL[2] and Hans FIBY[1]

[1]ITS Vienna Region / Austria · andreas.partusch@its-viennaregion.at
[2]Vienna University of Technology / Austria

This contribution was double-blind reviewed as full paper.

Abstract

Optimal paths in networks are determined by costs attached to nodes and edges. However, these costs are uncertain. The question raised in this paper is how to determine the effect that this uncertainty has on the optimal path, i.e., the spatial distribution of the optimal path. The method used is Monte Carlo Simulation. The simulation was done for the road network of Vienna. Travel time serves a cost function and the temporal distribution of travel time is derived from floating car data. The experiment shows that typically several different paths are possible with similar travel times.

1 Introduction

Today, a large variety of routing systems and services is available to determine the optimal path between a starting point and a destination in a given network. Such routing systems play an important role in individual traffic, and most people in developed countries have already used them. The definition of the optimal path in such systems is typically based on shortest path, minimum travel time, or similar criteria (WATERS 1999). Commonly used algorithms like Dijkstra (DIJKSTRA 1959) or A* (HART et al. 1968) apply the costs to the edges of the network and then search for the path with the minimum accumulated costs.

Determining the costs is simple if the used property is stable. An example for such a property is the geometrical distance. Other properties like travel time are time-dependent and are difficult to predict exactly. These properties can only be modelled statistically. Typically, an expected value is used to describe the travel time and the variation is ignored. This leads to uncertainty in the results and the goal of this paper is to show how to estimate this uncertainty. The procedure can then be used for different applications. Traffic planners may want to check if the principal roads have sufficient capacity, and navigational devices could use the method to compute alternative paths that they propose to the driver. In the experiment we used Monte Carlo Simulation (RICE 2007) to determine the spatial variation caused by the variation in the travel time. The road network used in the experiment is the road network of Vienna. The driving speed and its variation was determined using available floating car data.

2   Experimental Setup

The adopted routing software was the ITS router, developed by David Geroe in his master's thesis for ITS Vienna Region, the traffic management project of Vienna, Lower Austria, and Burgenland (GEROE 2010). It uses the graph produced by the Graph Integration Platform (GIP), which combines data from various sources like federal provinces and traffic infrastructure providers into a unified, intermodal transportation network. The feasibility of the data set for routing applications has been shown by the web-based routing application www.AnachB.at which began in 2009. The data set contains all relevant legal information like one-way streets or turn restrictions and the ITS router respects them. The relevant legal information is updated through a standardized interface by the organization defining these restrictions. Thus the resulting path is legally correct. However, since the legal rules may change, a subsequent repetition of the experiment may result in slightly different paths.

Monte Carlo Simulation is based on distribution functions for the variable parameters. Realistic distribution functions are necessary to provide realistic simulation results. Thus, realistic data for driving speed distribution is necessary. It was derived from floating car data. It has already been shown that floating car data provides a good estimation of real traffic speed distribution (GRASER et al. 2012a; GRASER et al. 2012b). The distribution function is determined assuming that observations from different days may be combined and are representatives for the same distribution. Since weather has a significant influence on driving speed (LITZINGER et al. 2012), a period with stable weather conditions was necessary. The following conditions for the data set were stipulated:

- The data has to cover the time period without gaps.
- The data set should not be too old to be able to verify it with recent observations.
- The week shall not contain public holidays because this would significantly change the driving behavior.
- There should be no extreme weather conditions (e.g., no heavy rain or snowfall).

Typically, the months April and October are well suited to fulfill these criteria. In April 2013 two weeks without holidays (08.04.2013 – 21.04.2013) with an average daily temperature between 6 and 18 degrees Celsius and almost no precipitation were found. Floating car data of these two weeks was used and evaluated for building the distribution functions as an input for the simulation.

The FLEET service provided by the Austrian Institute of Technology connects travel speed to the GIP links. These values are calculated by the weighted harmonic mean speed of the single probe cars detected per link. However, this aggregated travel speed on links is not useable for generating travel speed distributions since information on the variation is lost. Therefore, the raw data, respectively, the single vehicle speeds of the probe cars were used to calculate the distribution functions.

Each entry in the raw data consists of several fields including identification of the trip, the GIP link on which the speed value of the trip is projected, travelling direction, and the travelled distance on the link. This last information is relevant, if the transmitted position point of the probe cars is in the middle of a GIP link, and not the whole link is passed. The covered length is an important parameter for the evaluation. Speed values that are only assigned to a short segment (e.g. 15% of the total link length) are not reliable estimates for the speed on the whole link. The evaluation consisted of three steps:

- **Matching the raw data on the GIP network:** The floating car data set contains the information about which GIP links the data points have been projected to. However, the GIP is continuously updated and so the currently used graph version of FLEET and the version used for the simulation may not be identical. It would not be possible to interpret link IDs that have changed in the GIP. Hence, the GIP data and the floating car data were imported in a PostGIS data-base and joined for further analysis.
- **Defining time slices with similar travel speed characteristics:** In order to assess the influence of the time of day on the routing distribution, time slices were defined. Each slice should have as little variation in travel speed characteristic as possible.
- **Building distribution functions using the matched floating car data:** The floating car data matched on the graph are then split according to the time slices. This allows the computation of characteristic travel speed distribution functions for different link categories and time slices.

In order to identity the spatial variation, test cases are necessary. Nine pairs of start and destination segments were defined. The pairs can be grouped into tangential paths, radial paths, and longer paths across the city. Each of the paths was then determined 10,000 times with varying estimations of the travel speed. During the simulation, the resulting paths were stacked and thus it is possible to determine how often each path segment was used in the resulting paths. The maximum is obviously 10,000 and the minimum is 0.

Figure 1 shows the processes and components of the model, and shows how the Monte Carlo Simulation is integrated into the existing routing software.
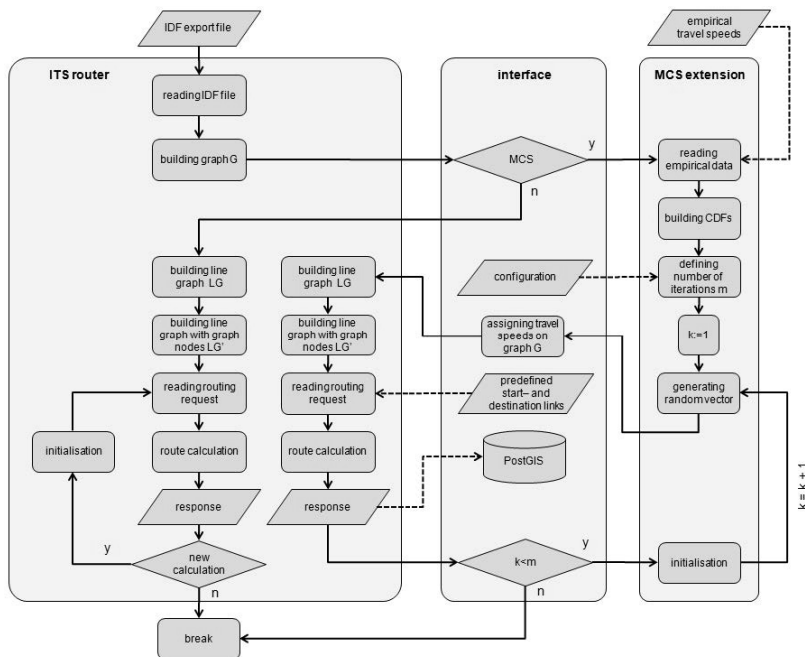

Fig. 1: Flow diagram of the simulation with its components and processes (Source: PARTUSCH 2013, 38)

3   Input Data

The most important data set is obviously the road network. However, since it is intended for both motorized and non-motorized traffic, in also contains nodes that are not permitted for motorized traffic. These irrelevant segments were eliminated from the network to improve the performance. Additionally, some roads are for parking purposes only and it is prohibited to drive through them. These were also eliminated. The final network consists of 41,505 directed links.

The original floating car data set contained 22,258,859 entries on 55,610 directed links. 19,849,709 entries could be matched on the 34,981 directed links of the GIP network used. One reason for mismatches could be a change of the link ID. However, there are other reasons as well. Many entries relate to links that have been eliminated during the network cleaning process. A significant portion of the cars providing data in the FLEET service are taxies and they are allowed to use some roads that are otherwise restricted (e.g., bus lanes). Since such links have been dropped, the matching is no longer possible. In addition, the start or end of the trip could lie on a road segment for parking purposes only, and these links have been dropped as well. Almost 85% of the directed links in the network for the simulation could be covered with data points, and approximately 8,400 directed links contain more than 500 floating car data points.

Traffic speed varies with the time of day, since traffic density changes (compare Figure 2). Thus the floating car data needs to be divided into different time slices. In a first step, 24 time slices were created. Data for different days were merged in these slices. Since weekend traffic is differently distributed, weekends had to be excluded. Figure 2 shows a case covered well by floating car data. However, many links did not have such a good coverage. In order to provide good coverage of speed distribution, the road network was clustered with the K-means algorithm (TAN et al. 2006, 496 ff) and the number of clusters was determined by evaluating the David Bouldin Index (JAIN & DUBES 1988, 186) for different numbers of clusters. Since city highways were an obvious class unto themselves, they were excluded from the clustering. Figure 3 shows two of the resulting clusters. Cluster 1 contains important roads with a speed limit of 50km/h. Cluster 5 contains traffic-reduced roads with a speed limit of 30km/h. In both cases, the average speed is lower during the daytime than during the night time. Obviously, the speed limit is sometimes exceeded and the largest outliers occur during the night time. This problem needs to be addressed during the modelling process.

These results can then be used to separate traffic characteristics in different time slices. Since the average traffic speed and its variation changes during the day, daytime and night time was separated (6AM to 6PM vs. 6PM to 6AM). Two special time slices can be treated separately as well: In general, the highest speeds were observed between 1AM and 4AM because the amount of traffic is at the lowest level in that period. The other extreme is morning traffic between 6AM and 9AM. It shows the lowest speed due to the high traffic density.
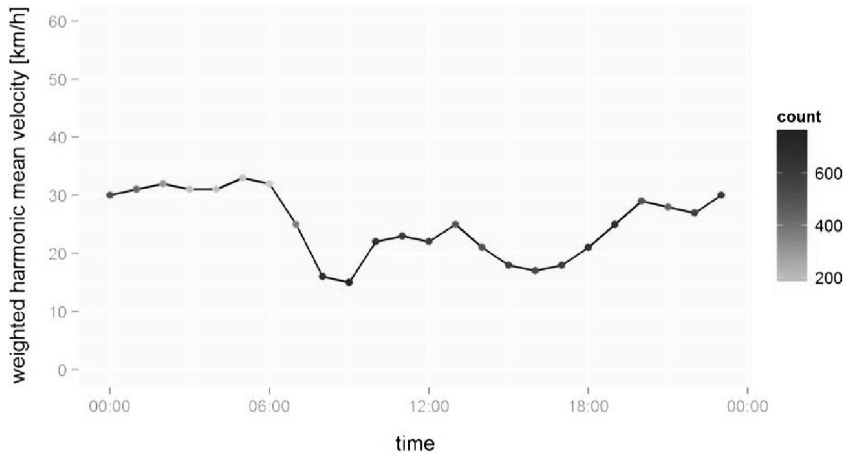
Fig. 3: Variation in the average speed along a link in the city centre (Museumsplatz to Getreidemarkt), darker points represent average speed computed from more floating car data (Source: PARTUSCH 2013, 47)
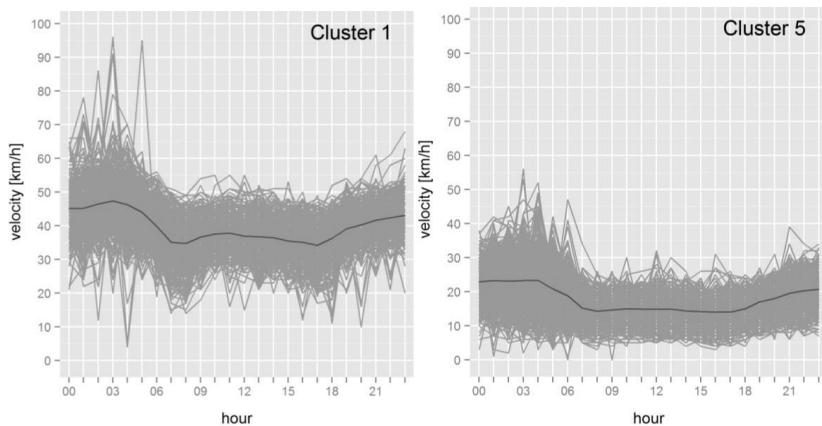

Fig. 4: Cluster plots of the speed profiles from the cluster analysis (Source: PARTUSCH 2013, 49)

The next step is the determination of the cumulative distribution functions. A necessary decision for the cumulative distribution function is whether speeding should be taken into consideration or not. Especially during the night, travel speeds exceeding the speed limit are observed. Using these higher speeds in the cumulative distribution function would provide a realistic model of driver behaviour. However, it could also lead to the conclusion that some paths are better than others because speeding is simpler along these paths. Restricting travel speed to the speed limit leads to less realistic modelling but the presented paths are legally correct. In this study, the second approach was taken because the results of the study shall be used for the improvement of www.AnachB.at and an official navigational support system should not propose speeding.

Since not all links in the network were covered by data points, a conceptual approach of applying density functions to road segments was necessary. The concept used here is functional road classes (FRC). This classification ranges from FRC0 (highways) to FRC7 (minor local roads) and is available for the used road network. Table 1 summarizes the number of data points in each of the FRC classes existing in the road network. 34,981 of the 51,505 links (almost 85%) in the road network are covered by the data. However, there are still a number of roads with low coverage. In order to eliminate this flaw, a longer observation period would be necessary. However, this could limit the homogeneity of the driving behaviour. Figure 5 shows the spatial distribution of the coverage. Darker colour represents better data coverage. Although there are large numbers of data points for the main roads, there are still many small roads with limited or no data. Still, the number of floating car data points of approximately 8,500 directed links exceeds 500, which provides a reasonable statistical basis for the creation of the distribution functions. Figure 6 shows an example for the resulting cumulative distribution functions.

Tab. 1:   Detailed view on number of floating car data (FCD) points per FRC

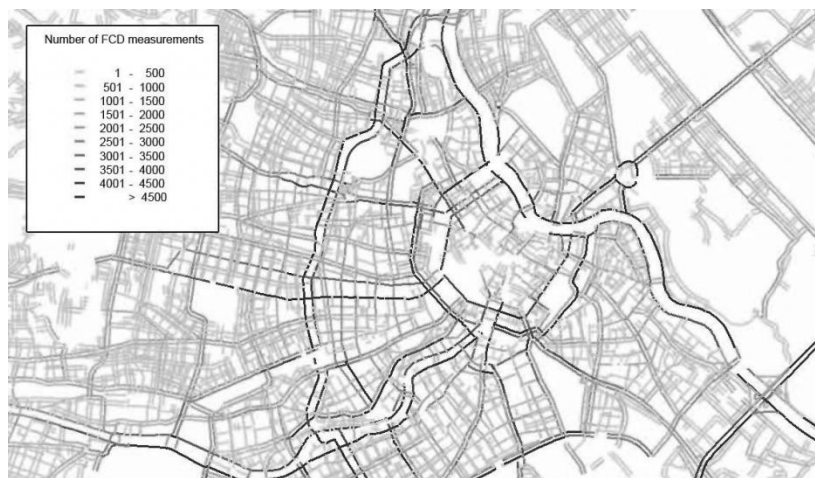| FRC	Number of FCD points	Number of Links	Points per Link
0	798,390	256	3,119
1	34,182	21	1,628
2	6,622,163	2,792	2,372
4	8,157,917	6,382	1,278
6	5	2	3
7	4,235,294	25,485	166
8	1,144	10	114
10	614	33	19
Total	19,849,709	34,981	

Fig. 5:   Density of floating car data points in the central part of Vienna (Source: PARTUSCH 2013, 45)
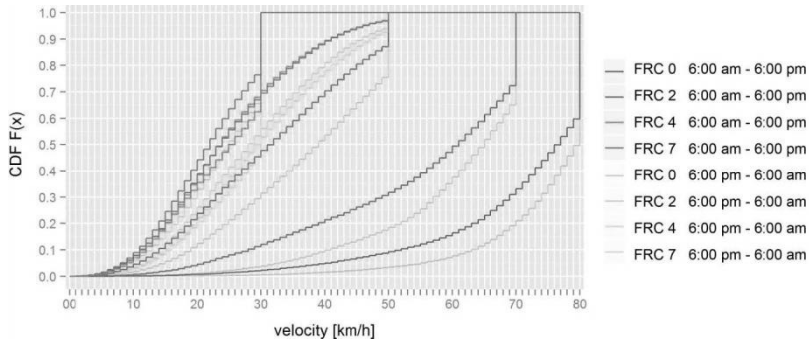
Fig. 6: Density of floating car data points in the central part of Vienna (Source: PARTUSCH 2013, 45)

4   Results

The results show differences depending of the type of path and the time of day. Path 1 is a tangential path and lies in an area with a limited selection of roads to travel. Figure 7 shows the results of the Monte Carlo Simulation. Thickness of the lines corresponds to the number of times the segment was part of the resulting path. In the northern part of the track more than 99% of the paths use the same link. In the southern part the effect of a small variability of driving speed is visible. The paths in Figure 7(a), which shows the results between 6:00AM and 9:00AM, have a maximum frequency of approximately 50%. Thus half the time it would be faster to use one of the other links. During the night hours (Figure 7(b)), however, the spatial variability is much lower.
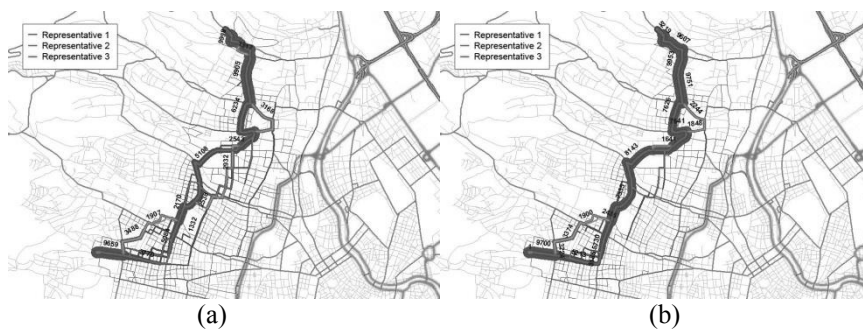

(a)   (b)

Fig. 7: Path plots for path 1 at the time slices (a) 6:00AM to 9:00AM and (b) 1:00AM to 4:00AM (Source: Partusch 2013, 66)

Path 6 starts in the outskirts and is destined for the city centre. The results show two clearly separated alternatives with similar probabilities (see Figure 8). Between 6 AM and 9 AM, the probabilities are approximately 40% for each alternative (Figure 8a), and between 1 AM and 4 AM the probabilities are almost 60% for the northern alternative and almost 40% for the southern alternative (Figure 8b). Between 6 AM and 9 AM using the narrow streets

between the two alternatives could be faster than using one of the main roads but the probability is only around 20% for all alternatives.
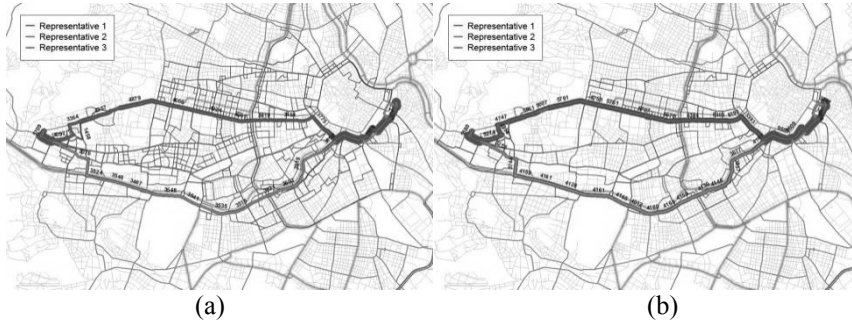

Fig. 8:   Path plots for path 6 at the time slices (a) 6:00AM to 9:00AM and (b) 1:00AM to 4:00AM (Source: PARTUSCH 2013, 76)

Path 7 is a longer path through the city. The plot for the time slice from 6:00 am to 9:00 am shows two completely different paths with higher probability values for this test case (see Figure 9a). The shorter, northern path will be faster in three quarters of the cases. However, there is a 20% chance that the southern path using the highway will be faster. The average distance between these two possibilities is around 3.5 km. The reason for the small number of other alternatives is the limited number of bridges crossing the Danube. During the night hours the difference is even smaller (see Figure 9b)

Fig. 9:   Path plots for path 7 at the time slices (a) 6:00AM to 9:00AM and (b) 1:00AM to 4:00AM (Source: PARTUSCH 2013, 78)

5   Conclusions and Future Work

In this paper we presented a Monte Carlo approach for assessing the effects of the travel speed uncertainty on the optimal path. Floating car data provided the necessary information to create distribution functions. Data from two weeks was used for the study to avoid effects from exceptional events like extreme weather conditions, major events, or a long

weekend. The presented results show that Monte Carlo Simulation is a suitable method to analyse the spatial distribution of shortest paths under uncertain travel speed.

The simulation results show different patterns for different times of the day. Since the speed distribution is based on real data, the observable effects reflect actual patterns. During night hours, driving speed has less variation than during the day. Thus paths with a high probability to be faster during morning traffic may be slower during the night time, because the speed on the major roads is closer to the speed limit. The simulation thus shows that path selection during the night time is simpler. However, although the spatial variation is bigger during the daytime, the use of major roads is typically faster than using narrow streets. Although sometimes paths using narrow streets are faster, the probability for these cases is quite low.

The results show that Monte Carlo Simulation can be used to control the efficiency of traffic management. Traffic in cities should be concentrated along the main roads because this guarantees that narrow streets are quiet and the quality of living increases for the people living there. This is important specifically at night. The simulation shows that this is the case for the test paths in Vienna. The narrow roads may be faster but that probability that this happens is typically below 1%. Thus people will not use the narrow roads.

A problem of the Monte Carlo Simulation as presented here is the time necessary for the computation. On a standard PC the computations take several hours. Thus it is too slow for on-thy-fly computation. Once this problem is solved, the approach can provide alternative path proposals. These proposals (as indicated in the results in Figures 7 – 9) could be presented to the user, who then selects the most appropriate proposal. However, this also requires up-to-date distribution functions for traffic speed and for longer path predictions for the change of the distribution functions. An interesting question would be, for example, how the predicted distribution function differs from the correct distribution function and how large the effect of that difference is.

One aspect has been ignored in this paper: Traffic speed is spatially correlated. A reduction of speed in one road will cause drivers to deviate from their intended path. This causes more traffic in nearby roads and eventually driving speed reduction. The creation of correctly correlated random samples thus requires knowledge on the spatial correlation. This information must then be applied during the creation of the random samples. Other approaches would be the application of ant colony behaviour or agent based modelling.

References

DIJKSTRA, E. W. (1959), A Note on Two Problems in Connexion with Graphs. Numerische Mathematik, 1, 269-271.

GEROE, D. (2010), Development of a Routing Program for Finding the Optimal Route in Motorized Individual Traffic based on the Graph Integration Platform. Master Thesis, Vienna, Austria, University of Applied Sciences Technikum Wien, Intelligent Transportation Systems.

GRASER, A., DRAGASCHNIG, M., PONWEISER, W., KOLLER, H., MARCINEK, M.-S. & WIDHALM, P. (2012a), FCD in the Real World – System Capabilities and Applications. Proc. 19th ITS World Congress, (p. 7). Vienna, Austria.

GRASER, A., WOLFGANG, P., DRAGASCHNIG, M., BRÄNDLE, N. & WIDHALM, P. (2012b), Assessing Traffic Performance using Position Density of Sparse FCD. Proc. 15th International IEEE Conference on Intelligent Transportation Systems. Anchorage, Alaska, IEEE, 1001-1005.

HART, P. E., NILSSON, N. J., & RAPHAEL, B. (1968), A Formal Basis for the Heuristic Determination of Minimum Cost Paths. IEEE Transactions of System Science and Cybernetics, 4 (2), 100-107.

JAIN, A. K. & DUBES, R. C. (1988), Algorithms for Clustering Data. New Jersey, Prentice Hall.

LITZINGER, P., NAVRATIL, G., SIVERTUN, Á. & KNORR, D. (2012), Using Weather Information to Improve Route Planning. Bridging the Geographic Information Sciences (AGILE 2012). Avignon, France, Springer, Lecture Notes in Geoinformation and Cartography, 119-214.

PARTUSCH, A. (2013), Monte Carlo Simulation for the Evaluation of Routing Results Concerning their Spatial Variation and Distribution. Master Thesis. Vienna, Austria: University of Applied Sciences Technikum Wien, Intelligent Transportation Systems.

RICE, J. A. (2007), Mathematical Statistics And Data Analysis (3 ed.). Thomson Books/ Cole.

TAN, P.-N., STEINBACH, M. & KUMAR, V. (2006), Introduction to Data Mining. Pearson Addison Wesley.

WATERS, N. M. (1999), Transportation GIS: GIS-T. In Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. (Eds.), Geographical Information System (2 Ed., Vol. 2). John Wiley & Sons, 827-844.