

Estimation of Travel Times on Signalized Arterials

Ivana Cavar¹, Zvonko Kavran¹ and Rino Bosnjak²

1. Faculty of Transport and Traffic Sciences, University of Zagreb, Zagreb HR-10000, Croatia

2. Faculty of Maritime Studies, University of Split, Split HR-21000, Croatia

Abstract: This paper describes procedure for estimation of travel time on signalized arterial roads based on multiple data sources with application of dimensionality reduction. Travel time estimation approach incorporates forecast of transportation nodes impedance and travel time on network links. Forecasting period is two hours and the estimation is based on historical data and real time data on traffic conditions. Travel time estimation combines multivariate regression, principal component analysis, KNN (k-nearest neighbours), cross validation and EWMA (exponentially weighted moving average) methods. When comparing estimation methodologies, relevantly better results were achieved by KNN method than with EWMA method. This is true for every time interval considered except for evening time interval when signalized arterial roads were uncongested.

Key words: Intelligent transportation systems, travel time estimation, signalised arterial roads, exponentially weighted moving average, k-nearest neighbours.

1. Introduction

Travel time information can be applied in various ITS (Intelligent Transport Systems) applications and services. ITS services such as DRG (dynamic route guidance) and VMS (variable message sign) are providing users with travel times and actual traffic conditions.

Most drivers and travellers find this information useful because it helps them in route selection choice, saves their travel time and when given pre-trip it allows them to choose the time of departure and therefore to mitigate traffic congestion. Research has shown that 77.7% drivers will change their route based on travel time information and even if there is no any alternative route, they will find driving less stressful when knowing what to expect ahead of them [1].

From the traffic managers point of view, travel time information is an important index for traffic system operation and it helps in managing travel demand, making transport greener and achieving more efficient

use of existing transport infrastructure.

Travel time is affected by a range of different factors as weather conditions, driver characteristics, road geometry, traffic flow characteristics, free flow travel speed etc.. Therefore, travel time estimation requires large number of well selected input variables as well as carefully selected estimation methodology.

This paper gives short overview of travel time estimation researches from literature and description of procedure applied for estimation of travel times on signalized arterials in respect to the multiple sources input values and different estimation intervals. Results for estimation based on EWMA (exponentially weighted moving average) and KNN (k-nearest neighbours) method were compared for peak and off-peak time intervals.

2. Literature Review

In literature, it can be found different approaches for the travel time estimation and different road characteristics that travel time estimation is performed for. Best results are achieved for freeways and for free flow traffic conditions. The reason for this lies in the lower number of factors influencing the travel time. Thus understanding those traffic factors affecting

Corresponding author: Ivana Cavar, Ph.D., assistant professor, research fields: intelligent transportation systems, data mining, traffic modeling and simulation, geographic information systems. E-mail: ivana.cavar@fpz.hr.

travel time is essential for improving prediction accuracy.

Free flow travel speed is one of the factors that affect travel time. The travelling speed along an arterial road depends not only on its geometry but also on the traffic flow characteristics and traffic signal coordination [2]. Also, occurrence of incident situations, holidays or other uncommon events influence travel time prediction [3], as well as weather conditions [4, 5], traffic operation and congestion level [6]. The author’s research takes into account these findings by considering working days separately from holidays and uncommon events influence, as well as, including weather conditions data and information on of peak and off-peak time intervals into estimation model.

Another important element of travel time estimation is the forecasting period. The greater the forecasting period, the higher the prediction error is [7].

Different approaches for travel time data collection are also used in literature and can be divided into two main groups [8]:

(1) Logging the passage of vehicles from selected points along a route with loop detectors, transponders, registration plate matching, remote or indirect tracking, radio beacons, video surveillance etc. [9, 10];

(2) Using moving observation platforms travelling in the traffic stream itself and recording information about their progress as probe vehicle, floating car and volunteer driver methods [10-12]. Models proposed by the authors are based on data collection procedure that could be considered to belong to this group.

When it comes to travel time estimation approaches, they can be divided in two basic groups [13]:

(1) Extrapolation models are based mainly on

statistical approaches and historical values as regression models [14, 15], ARIMA (autoregressive integrated moving average) models [15], STARIMA (space-time autoregressive integrated moving average) [16, 17], Kalman filter [18], ANN (artificial neural networks) [19, 20], SVM (support vector machines) [21, 22] and pattern based forecasting [23];

(2) Explanatory models are based mainly on factor or parameters analyses and traffic flow theory as dynamic traffic assignment [24, 25].

Travel time approach suggested by the authors is considered to belong to the extrapolation models group.

3. Travel Time Estimation Approach

Travel time estimation approach is based on combination of data mining methods applied in different steps of estimation procedure. Short overview of procedure is given in Fig. 1.

Data collected from multiple sources are fused and data pre-processing (map matching, interpolation and error reduction) is performed so data could be prepared for dimensionality reduction. Dimensionality reduction aims to reduce the number of random variables under consideration so that future analysis steps would be less compute demanding. For this purpose, multivariate regression and PCA (principle component analysis) were applied. Multivariate regression is used for modelling and analysing variables by focusing on the relationship between a dependent variable and one or more independent variables in order to understand how the typical value of the dependent variable changes when one of the independent variables is varied, while the other independent variables are held fixed. PCA is used as

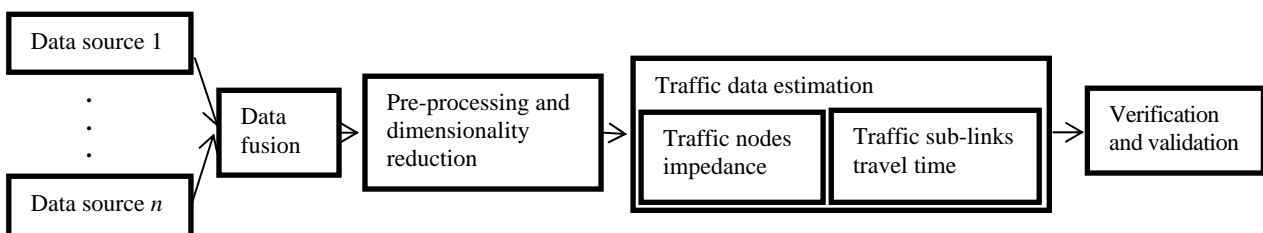


Fig. 1 Travel time estimation procedure.

an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first principal component, the second greatest variance on the second and so on. It converts a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. This helps in reduction of original set of variables (dimensions) into smaller one and eases up the computing process in estimation phase.

For estimation, two sets of variables were used. One is to estimate impedance of the nodes (or delay) in the traffic network (signalized intersections on arterial roads) and one is to estimate travel time on traffic network sub-links. Two approaches were used for estimation, one based on EWMA (exponentially weighted moving average) and one based on KNN (k-nearest neighbours) method. EWMA is a type of infinite impulse response filter that applies weighting factors which decrease exponentially. The weighting for each older data point decreases exponentially, never reaching zero. Generally, series Y may be calculated recursively:

$$\begin{aligned} S_1 &= Y_1, t > 1 \\ S_t &= \alpha \times Y_{t-1} + (1-\alpha) \times S_{t-1} \end{aligned} \quad (1)$$

where,

α is coefficient representing the degree of weighting decrease;

Y_t is the observation at a time period t ;

S_t is the value of the EWMA at any time period t .

KNN is machine learning algorithm that assigns the property value for the object to be the average of the values of its k nearest neighbours where k is a positive integer, in this case, defined by cross validation. Determination of value k has big influence on modelling results therefore should be addressed on adequate level [5]. For this purpose, the authors used ν -fold cross-validation. ν -fold cross-validation uses original sample and randomly portions it into ν subsamples. Of the ν subsamples, a single subsample

is retained as the validation data for testing the model, and the remaining $(\nu-1)$ subsamples are used as training data. This process is then iterated ν times, with each of the ν subsamples used exactly once as the validation data to achieve best value of k . After estimation, the model is verified and validated based on collected data that were not used for modelling estimation.

4. Case Study

4.1 Site Description

Zagreb is the capital and the largest city of the Republic of Croatia situated at elevation of approximately 122 m above sea level in the northwest of the country, along the Sava River, at the southern slopes of the Medvednica Mountain (Fig. 2). The population of the City of Zagreb and the Zagreb County is 1,110,517 [26].

The city is served by the Zagreb Airport, which carries more than 2,150,000 passengers per year [27]. Transport system relies on parallel Zagreb Commuter Rail system and combination of city-managed mass transit (19 inner-city tram lines and 120 bus routes) and individual transportation. It is a regional highway hub with eight highways and expressways radially leading into the city through the Zagreb bypass. Most of the commuter trips are on the east-west relation, causing high traffic on roads like Vukovar Avenue, Dubrovnik Avenue and Zagrebacka Avenue. These three avenues are used for people's travel time estimation test site (Table 1).

4.2 Data Description

Data used for travel time estimation were collected from three sources including GPS (global positioning system) tracks of 297 vehicles, meteorological data from official meteorological and hydrological service and road segments data. Geographical area covered with data collection process was urban area of City of Zagreb, Croatia and time window for data collection

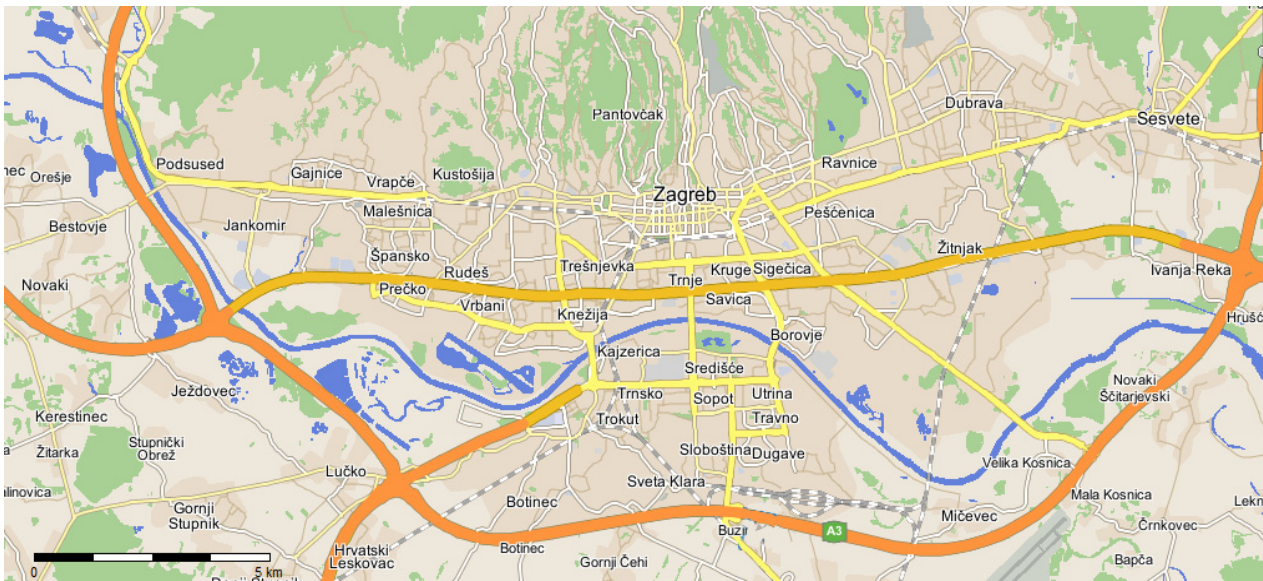


Fig. 2 Site area.

Table 1 Avenues.

Street	Overall length (km)
Vukovar Avenue	3.7
Dubrovnik Avenue	4
Zagrebacka Avenue	5.2

was 13 months.

Traffic condition data are analyzed based on the GPS vehicle tracks. The data were recorded by the GPS device and sent via GPRS (general packet radio service), GSM (global system for mobile communications) or SMS (short message service) data transfer technologies every travelled 100 m (if the vehicle engine was running) or every 5 min (if the vehicle engine was turned off). Data that were sent to the local server and stored in the database include (Table 2):

- Log time—time of recording expressed in UTC (universal time coordinated);
- Vehicle ID—identifier of the vehicle/GPS device;
- X coordinate—X coordinate of the GPS record

(WGS84—World Geodetic System 1984);

- Y coordinate—coordinate of the GPS record (WGS84);
- Speed—vehicle current speed in km/h;
- Course—angle at which the vehicle is travelling with reference to the north;
- GPS status—three values that indicate the accuracy of the record. GPS status 3 indicates that the data have been collected from at least four satellites. GPS status 2 means that the data have been collected from three satellites. GPS status 1 indicates records with very questionable accuracy;
- ES (engine status)—shows whether the vehicle engine was running or was turned off while making the record.

Second data set is retrieved from meteorological and hydrological service and it includes:

- minimum and maximum daily air temperature and their difference;
- daily air temperature (measured at 7 h, 14 h and 21 h and daily average air temperature);

Table 2 Part of traffic conditions data.

Log time	Vehicle ID	X	Y	Speed	Course	GPS status	ES
1127839078	242	16,3344	44,0093	87	267	3	1
1127839078	21	17,1138	45,4316	58	335	2	1
1112335806	3	15,9988	45,7488	19	298	3	1

- wet ground temperature (measured at 7 h, 14 h and 21 h and daily average value);
- air pressure (measured at 7 h, 14 h and 21 h and daily average value);
- sun shining period;
- precipitation;
- snow cover thickness;
- thickness of new snow cover;
- ground condition (measured at 7 h, 14 h and 21 h);
- horizontal visibility (measured at 7 h, 14 h and 21 h);
- air steam pressure (measured at 7 h, 14 h and 21 h and daily average value);
- relative humidity (measured at 7 h, 14 h and 21 h and daily average value);
- clouds with phenomena (measured at 7 h, 14 h and 21 h and daily average value);
- direction and speed of wind (measured at 7 h, 14 h and 21 h and daily average value);
- ground temperature at -2 cm, -5 cm, -10 cm, -20 cm, -30 cm, -50 cm, -100 cm (measured at 7 h, 14 h and 21 h and daily average value).

Observatory diary contains daily data on phenomena (rain, dew, rain shower, mist, fog, frost, strong wind, snow, sleet, thunder, etc.) as well as data on phenomena duration and intensity.

Third data set contains information on road network (Table 3):

- Segment ID—identifier of the road segment;
- Type—numeric code of the road type;
- Direction (D)—code representing road direction;
- Start X — X coordinate of the beginning of segment (WGS84);
- Start Y — Y coordinate of the beginning of segment (WGS84);
- End X — X coordinate of the end of segment

(WGS84);

- End Y — Y coordinate of the end of segment (WGS84);
- Length (L)—length of the segment in m;
- Name—name of road.

Road segment is referred to part of the road between two nodes (intersections) and is further divided to the level of sub-segments. The road sub-segment represents a part of segment that has only one direction of the traffic flow.

When analyzing overall number of road sub-segments, arterial road sub-segments represent 11, 76% of overall road transportation network in the City of Zagreb with a share of 23.4% of overall road transportation network length. Average length of arterial road sub-segment is 221.35 m. On each arterial sub-segment there were averagely 5,010.5 tracks recorded with average vehicle speed of 56 km/h and speed deviation of 15.045.

5. Travel Time Estimation Results

For the purpose of travel time estimation characteristic time intervals are identified based on the analysis of speed records on each sub-segments. Characteristic time intervals are considered to be time periods when the vehicle throughput capacity (numbers of vehicles that pass the sub-segment in time interval) on the observed sub-segments is significant (significantly low—to identify peak periods or significantly high—to identify free flow conditions). A presentation of one such analysis is given in Fig. 3 for sub-segment that corresponds to the Zagrebacka Avenue for the traffic direction from East to the West where the authors can see pattern of the average vehicle speed during the day. The presentation is given for the whole week and the authors can see how traffic conditions depend on the

Table 3 Part of road network data.

Segment ID	Type	D	Start X	Start Y	End X	End Y	L	Name
60201	1030	1	15,95	45,7943	15,9492	45,7943	61	Zagrebacka avenue
60202	1030	2	15,9405	45,7938	15,9409	45,7938	34	Zagrebacka avenue

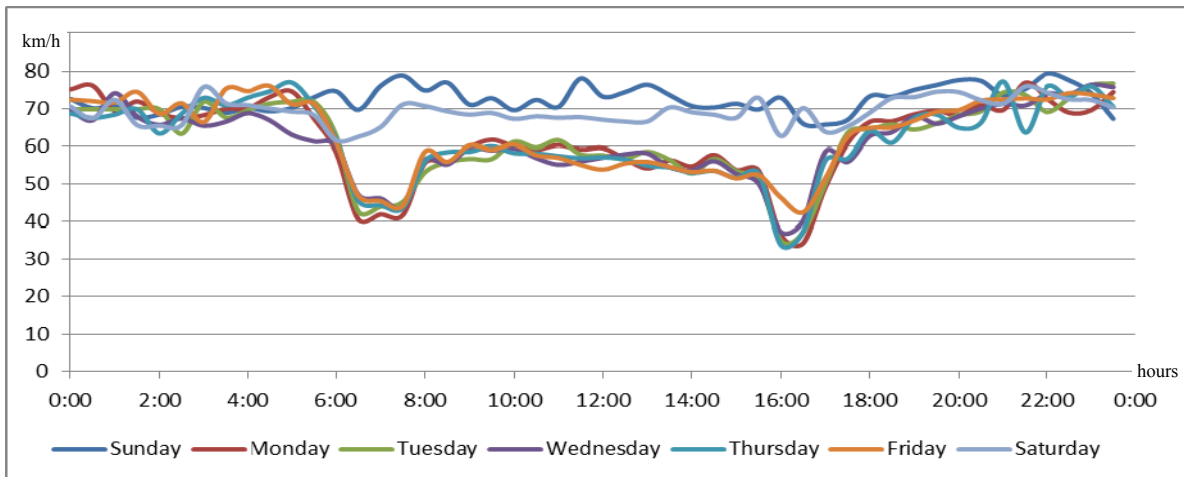


Fig. 3 Weekdays speed pattern for Zagrebicka Avenue.

day of the week because weekend traffic pattern is significantly different than the ones for working days. Working days traffic pattern has clearly visible morning and afternoon congestion periods that are most expressed in the interval between 06:00 h and 08:00 h and 15:30 h-17:30 h. Therefore four time intervals are selected for travel time estimation procedure morning peak interval (06:00-08:00 h), daylight free flow interval (12:00-13:00 h), afternoon peak interval (15:30-17:30 h) and evening free flow interval (21:00 h-22:00 h).

Travel time estimation is performed on sequence of sub-links in average length of 4,403 km, including seven signalized intersections and two signalized on

demand pedestrian crossings. KNN and EWMA method are used to estimate traffic network nodes impendence and sub-link travel time. Results for 10 fold cross validation with Euclidian distances and standardized values are presented in Fig. 4.

Same methodology is applied for estimation based on historical data and for real time estimation for forecasting interval of 2 h. MAPE (mean absolute percentage error) for booth approaches and booth estimation procedures is given on Figs. 5-8.

Based on the figures, it can be seen that for almost every time interval better results are achieved by KNN than EWMA method (except for off peak evening time interval).

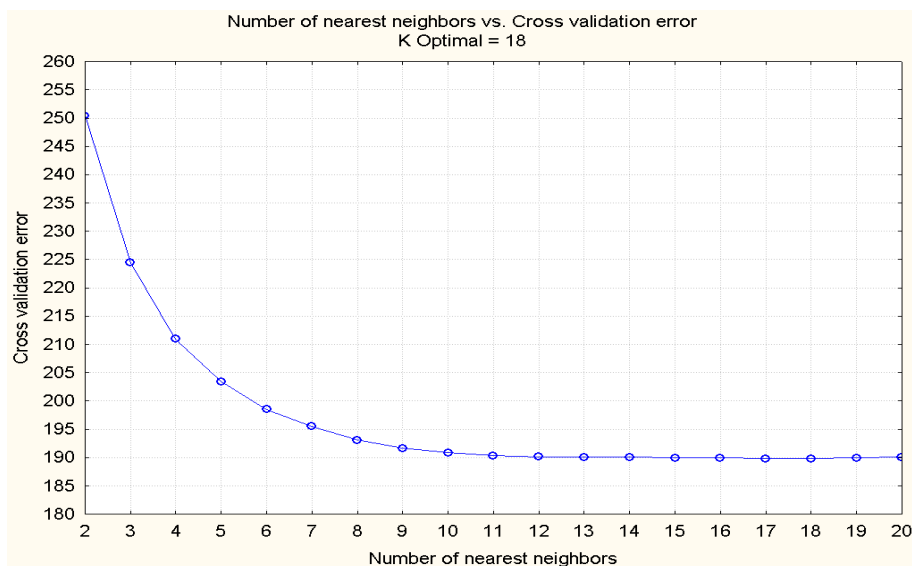


Fig. 4 Cross validation error.

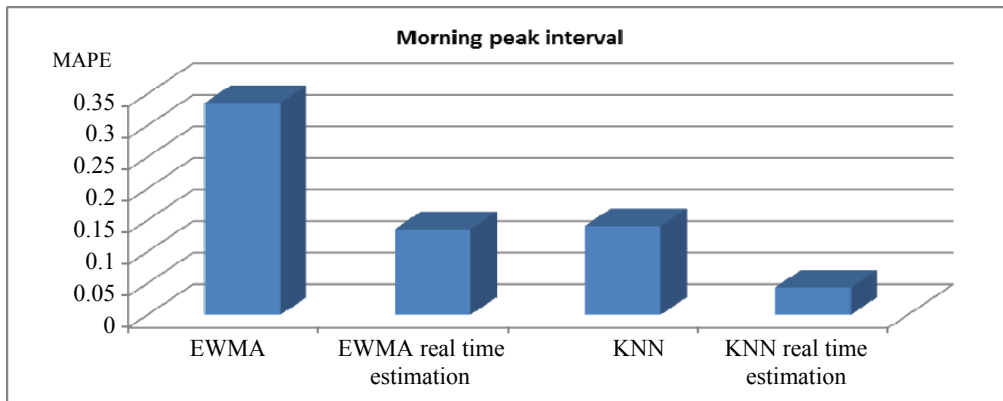


Fig. 5 MAPE for travel time estimation on arterial roads during the morning peak time interval.

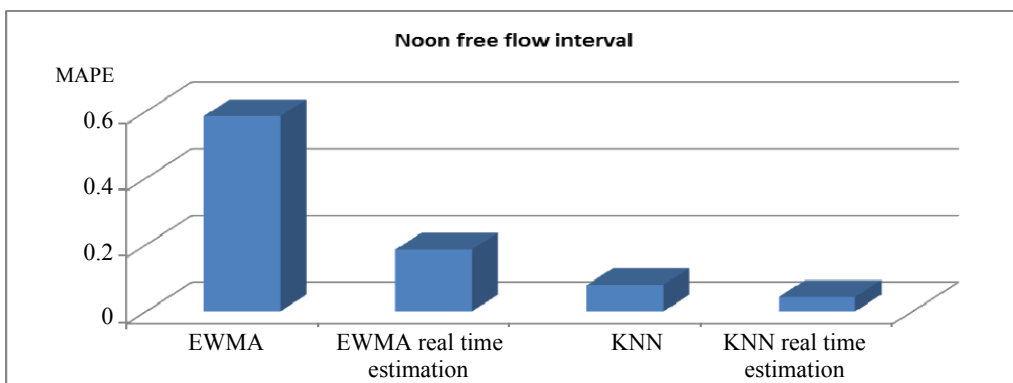


Fig. 6 MAPE for travel time estimation on arterial roads during the noon uncongested time interval.

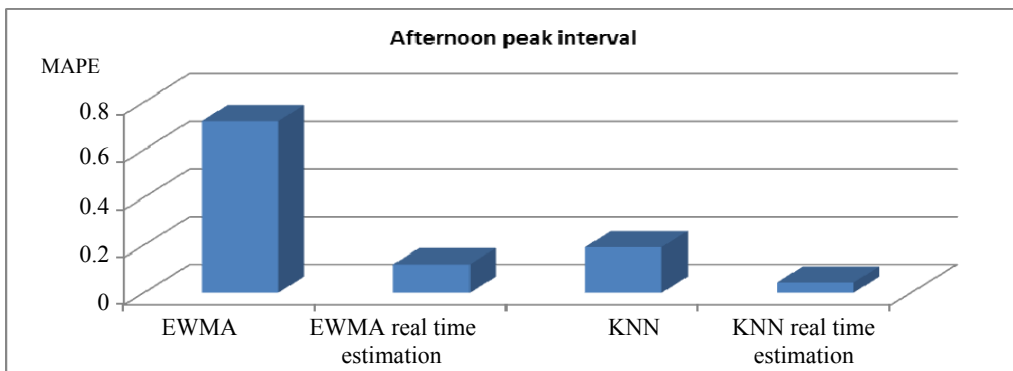


Fig. 7 MAPE for travel time estimation on arterial roads during the afternoon peak time interval.

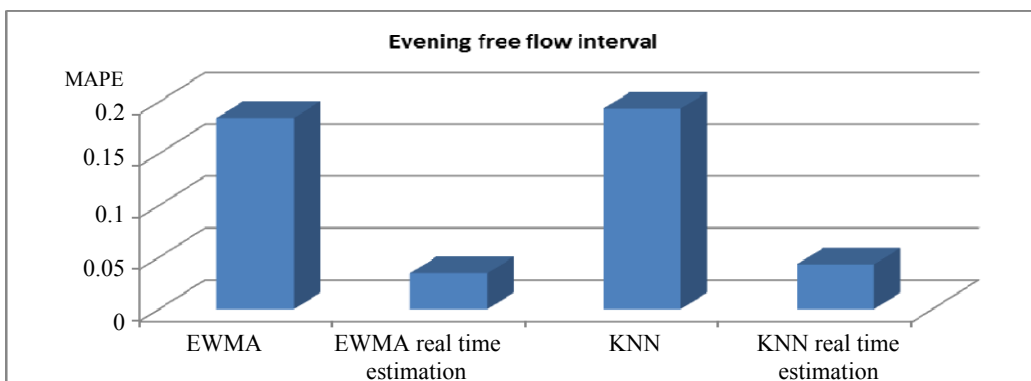


Fig. 8 MAPE for travel time estimation on arterial roads during the evening peak time interval.

6. Conclusions

When comparing estimation methodologies, relevantly better results were achieved by KNN method (18 nearest neighbors) than with EWMA method. This is true for every time interval considered except for evening time interval when signalized arterial roads were uncongested. In this case, results are comparable and MAPE is very low (lower than 0.18) for travel time estimation based only on historical data and lower than 0.03 for estimation with forecasting interval of 2 h.

This way achieved travel time estimations can be used in real life systems as dynamic route guidance and variable message sign information for arterial roads that provides users with travel times and actual traffic conditions. More importantly, this way estimated travel time can be used as pre-trip travel time information allowing users to postpone beginning of their trip or to change route based on time saving estimation for an alternative route.

References

- [1] E. Ramsay, J.E. Catchpole, J.Y.K. Luk, Evaluation of VicRoads Drive-Time System, Research report ARR 305, ARRB Transport Research Ltd., Vermont South, Victoria, Australia, 1997.
- [2] K.M. Lum, H.S.L. Fan, S.H. Lam, P. Olszewski, Speed-flow modeling of arterial roads in Singapore, *Journal of Transportation Engineering* 124 (6) (1998) 213-222.
- [3] J.W.C. van Lint, Reliable Travel Time Prediction for Freeways, TRAIL Research School, Netherlands, 2004.
- [4] S.I.J. Chien, C.M. Kuchipudi, Dynamic travel time prediction with real-time and historic data, *Journal of Transportation Engineering* 129 (6) (2003) 608-616.
- [5] I. Cavar, Z. Kavran, M. Petrovic, Hybrid approach for urban roads classification based on GPS tracks and road subsegments data, *Promet—Traffic & Transportation* 23 (4) (2011) 289-296.
- [6] A.J. Richardson, M.A.P. Taylor, Travel time variability on commuter journeys, *High Speed Ground Transportation Journal* 12 (1) (1978) 77-99.
- [7] L. Kisgyorgy, L.R. Rilett, Travel time prediction by advanced neural network, *Periodica Polytechnica Civil Engineering* 46 (1) (2002) 15-32.
- [8] H.E. Lin, R. Zito, M.A.P. Taylor, A review of travel-time prediction in transport and logistics, *Journal of the Eastern Asia Society for Transportation Studies* 5 (2005) 1433-1448.
- [9] G. Huisken, E. van Berkum, Short-term travel time prediction using a neural network, in: *Proceedings of the 13th Mini-EURO Conference*, Bar, Italy, 2002.
- [10] R.L. Bertini, M. Lasky, C.M. Monsere, Validating predicted rural corridor travel times from an automated license plate recognition system: Oregon's frontier project, in: *Proceedings of 12th World Congress on ITS*, San Francisco, United States, 2005.
- [11] R. Hawkins, P. Stopher, Collecting data with GPS: those who reject, and those who receive, in: *Proceedings of 27th Australian Transport Research Forum*, TSC, Adelaide, Sep. 2004, pp. 29-31.
- [12] I. Cavar, Vehicle routing and travel time prediction, in: *Proceedings of the GIS Research UK, 17th annual conference*, GISRUK, University of Durham, 2009, pp. 235-240.
- [13] H.H. Versteegt, C.M.J. Tampere, Predict Time State of the Art and Functional Architecture, TNO Intro report 2003-07, 2005.
- [14] H. Sun, H.X. Liu, H. Xiao, B. Ran, Short Term Traffic Forecast Using the Local Linear Regression Model, TRB paper no. 03-3580, Transportation Research Board, Jan. 2003.
- [15] B.L. Smith, B.M. Williams, R.K. Oswald, Comparison of parametric and nonparametric models for traffic flow forecasting, *Transportation Research* 10 (2002) 303-321.
- [16] Y. Kamarianakis, P. Prastacos, Space-time modeling of traffic flow, *Methods of spatial analysis—Spatial time series analysis*, in: *European Regional Science Association Proceedings*, Dortmund, 2002.
- [17] Y. Kamarianakis, P. Prastacos, Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches, *Transportation Research Record* 1857 (2003) 74-84.
- [18] C. Zhou, P.C. Nelson, Predicting traffic congestion using recurrent neural networks, in: *9th World Congress on ITS*, Chicago, 2002.
- [19] Y. Tao, B. Ran, An application of neural network in corridor travel time prediction in the presence of traffic incidents, in: *9th International Conference on Applications of Advanced Technology in Transportation*, Chicago, Illinois, USA, 2006.
- [20] A.K. Singh, G. Abu-Lebdeh, Arterial network travel time estimation using conditional independence graphs and state space neural networks, in: *18th IASTED International Conference on Modelling and Simulation*, Montreal, QC, Canada, 2007.

- [21] C.H. Wu, J.M. Ho, D.T. Lee, Travel-time prediction with support vector regression, *IEEE Transactions on ITS* 5 (4) (2004) 276-281.
- [22] L. Vanajakshi, L.R. Rilett, Travel time prediction using support vector machine technique, in: *IEEE Intelligent Vehicles Symposium*, Istanbul, Turkey, 2007.
- [23] D. Wild, Pattern-based forecasting, in: *Proceedings of the 2nd DRIVE-II Workshop on Short-Term Traffic Forecasting*, Delft, 1994.
- [24] M.E. Ben-Akiva, M. Bierlaire, B. Didier, H.N. Koutsopoulos, M. Rabi, Simulated-based tools for dynamic traffic assignment: DynaMIT and applications, in: *ITS America 10th Annual Meeting*, Boston, USA, 2006.
- [25] M.E. Ben-Akiva, M. Bierlaire, B. Didier, H.N. Koutsopoulos, M. Rabi, Real-Time Simulation of Traffic Demand-Supply Interactions within DynaMIT' Transportation and Network Analysis: Current Trends, *Miscellanea in Honor of Michael Florian*, Springer US, Kluwer, 2002, pp. 19-36.
- [26] *Census of Population, Households and Dwellings 2011*, Croatian Bureau of Statistics (1441), June 2011.
- [27] I. Akmadzic, *Official Zagreb Airport Statistical Report*, Statistical report, 2012.