TRANSPORT
ISSN 1648-4142 / eISSN 1648-3480
2017 Volume 32(3): 302–313
doi:10.3846/16484142.2016.1190402

METHOD FOR ANALYSIS AND PREDICTION OF DWELL TIMES AT STOPS IN LOCAL BUS TRANSPORTATION

Csaba Csiszár, Zsolt Sándor

Dept of Transport Technology and Economics, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Hungary

Submitted 15 October 2014; resubmitted 16 June 2015, 17 September 2015; accepted 28 December 2015; first published online 25 May 2016

Abstract. The punctuality of local public bus transportation contributes to the service quality and has impact on mode choice. Appropriate planning and control measures as well as adequate passenger information require efficient analysis and prediction methods. Advanced fleet tracking systems provide enough data for these research purposes. The main factors that cause schedule deviation has been identified through the analysis of the data. As important time elements of the journey time are dwell times at stops, the research focused on it; however, the elaborated database structure is adequate also for analysis of the other time elements, which coincides with our further research intentions too. Innovative methods based on the historical data have been elaborated for the prediction of dwell times. The essence of the method: multi-variate analysis of the dwell times by exploration of the significant influencing factors, and then prediction of times based on the factors describing the certain situations. Soundness of the methods has been verified with examples, and the results well approximated the real values.

Keywords: data analysis; database structure; bus transport; prediction; dwell times.

Introduction

More and more data elements that describe the features of the traffic events are generating in the informatics systems applied in the transportation. These are in most cases used for monitoring, control, account and quality control functions, but this data set is also appropriate for other operative and research purposes. It is quite frequent, that data are available from several sources for the description (mapping) and analysis of a transportation process. The main challenges of the data analysis are:

- various methods of acquiring and recording;
- dissimilar meanings, reliability, data structures and formats.

The research objectives were twofold:

- to reveal the features (regularity) of schedule deviation of vehicles;
- to identify the interrelation for the estimation of dwell times at stops based on the preliminary identification of the significant factors that influence the deviation (it can be used to inform passengers using the real-time and the predicted data).

This topic is actual and relevant because efficient operational management requires the comprehensive analysis of the processes and introduction of new methods.

The data with high quality and reliability have great value either for direct information purposes or for further calculations. Especially the simulation models require detailed and precise mapping of the attributes/properties of the infrastructure and the operations (Čertický *et al.* 2015).

The initial hypotheses of the research were the dwell time at stop:

- depends on the volume of the passenger-traffic at the stop;
- fluctuates according to the time of day (in accordance with the volume of the passenger-traffic);
- is influenced by the weather circumstances;
- depends on the floor height of the vehicle;
- is influenced by the deviation of the schedule of the vehicle.

During the research, data originating from the satellite based fleet management system of a public transport bus company in the city of Győr in Hungary have been processed. After the preliminary specifications, Kisalföld Volán Ltd., who is solely responsible for the urban public transport in Győr, handed over the required

Corresponding author: Csaba Csiszár
E-mail: *csiszar.csaba@mail.bme.hu*

data. Detailed weather information regarding the region of Győr was collected from the Időkép Weather Service Ltd. and National Weather Service. Results of manual traffic counting data provided by the Municipality of Győr have also been taken into consideration.

The effects of the factors have been examined both separately and collectively. Dwell times have been estimated by two different methods and the results have been compared. As an initial step, the most relevant scientific papers available have been overviewed and the results have been summarized.

1. Literature Review – Related Work

Levinson (1983) was the earliest study on the bus dwell time estimation. He formulated the bus dwell time as a function of two primary contribution factors by using the linear regression approach. The factors were:

- number of alighting and boarding passengers;
- the amount of time required for bus doors opening and closing.

Weather circumstances may also affect the transportation in two ways, either indirectly or directly. In the first case, the weather influences the open-air activities and thus the number of the public transport users. In the second case, it has an impact on the travelling and transportation process. Based on Stover and McCormack (2012) it can be declared that rain has the most significant impact on the transportation.

The road traffic characteristics can also be partially estimated applying statistical methods after the analysis of the GPS based tracking data and the separation of the movement and stopping phases (by identification of the inceptive and terminating timestamps at stops). Uno *et al.* (2009) determined the level of service of a certain route by the journey times. Mean values and deviations of the journey times have been considered. It was recognised that effects of the weather conditions on journey time should also be investigated.

Several researches dealt with the examination of planned, realized and perceived time elements of the road passenger transportation. Travellers are especially sensitive to the schedule deviations. 'Gaps' between perceived and planned quality can be mitigated by the provision of reliable information (Watkins *et al.* 2011). Significant development is discernible on this field by the enhancement of the accuracy of prediction processes, by new and more sophisticated approaches, by the incorporation of historical and real-time data into the prediction models (Vu, Khan 2010). The processing of tracking data from fleet management and passenger counting systems is the basis of these procedures. Methods of artificial neural networks, search for the *k*-nearest neighbours and linear regression are commonly used in these algorithms (Yu *et al.* 2011).

Bajwa *et al.* (2008) investigated the mode and departure time choice processes of passengers and devised different model specifications for a combined mode and departure-time choice-model, which takes the delays also into consideration. Zito *et al.* (2011) explored the opportunities of advanced traveller information systems at a public transportation provider company. They made model calibration to determine the potential additional share of demand attracted by the adoption of advanced traveller information systems. They also made a simulation to appraise the uncertainty of some parameters in the calibrated demand model in order to determine the importance of the type of information and its cost, whereas they are less interested in the system that provides the information.

Meng and Qu (2013) analysed the infrastructural and traffic flow parameters of the bus bays and the passenger flow in order to establish a model for the dwell time estimation. They determined that the conventional data-driven regression approaches cannot be used to estimate the dwell times at stops because they are incapable of dealing with the interactions among buses, arrival passengers and traffic on the shoulder lane. There are a lot of influencing factors that cause uncertainty that have to be identified and formulated. Tirachini (2013) has analysed the dwell times at stops in order to determine influence of different payment methods with the assessment of fare collection systems, bus floor level and age of passengers.

Song *et al.* (2013) have investigated the accurate prediction of bus arrival time with the analysis of bus running processes. They have created a self-adaptive exponential smoothing algorithm to predict the bus running speed. The delay caused by the signal control and the acceleration and deceleration were also taken into consideration. Zhang and Teng (2013) created a model for the estimation of dwell times based on the data of Automatic Vehicle Location and Automatic Passengers Counters. The model takes other secondary factors like crowding and fare type into consideration. The results have indicated that the models can be well applied in high demanded urban bus lines, especially in presence of high occupancy of vehicles.

Peña and Moreno (2014) investigated the dwell times in the Transmilenio Transport System, which transports huge amount of people and being operated as a bus rapid transport system. They analysed the dwell times in order to identify them and reduce the delays due to the long dwells at the stops. They established a relationship between the bus stop capacity and dwell times with the use of several static and dynamic infrastructural and vehicle parameters. The calculation was used for capacity determinations.

A proven measurement to facilitate smoothing of bus follow up times is establishment of bus lanes; however, it cannot be applied in every case. Esztergár-Kiss *et al.* (2012) have elaborated a method that supports decision-making regarding realization of a bus lane. By application of this method, several influencing factors and their effects can be calculated in an efficient way.

The basic idea of bus lanes is separation of vehicle types in the road traffic. This idea can be extended towards several directions. One of them is the trip booking or in wider interpretation the infrastructure booking. This 'hard tool' of traffic management offers predictable travel times for the vehicles introducing limitations according to the actual traffic conditions and capacities.

This concept was summarized in the paper by Soltész *et al.* (2011).

During the recent years, researchers have analysed the influence on boarding and alighting times of several vehicle characteristics and passenger related factors, including:

– bus door width and bus floor height (Fernández *et al.* 2010) – wider door can reduce the average alighting time by almost 40% and lower floor height only reduces the average alighting time from 1 to 9%;
– lift operation (Dueker *et al.* 2004) – increases the dwell time by 60–120 s;
– alternative fare payment techniques (Fletcher, El-Geneidy 2013; Tirachini 2013) – how do on-board sale and ticket validation increase the dwell times;
– age of passengers (Tirachini 2013) and crowding and/or friction effects among passengers boarding; friction and/or crowding can be detected, when passengers boarding form two queues through a single door, which increases the boarding time.

Currie *et al.* (2013) found that crowding had a more significant effect on dwell times than the presence of entrance steps.

The referred studies are various and they can be applied for a specific area, with restrictions regarding to the data needs. In these papers, the analysed parameters have been limited to the recorded data by the on-board systems. External data sources were not involved to the research. Effects of weather and vehicle types have not been identified whereas it has a significant impact on the dwell times. Consequently, our results are new, original and unique because such results have not been published where the elaborated multi-variate data analysis model uses several data from different sources simultaneously.

2. Method for Analysis

The elaborated method is based on the historical data recorded by the satellite-based fleet tracking system. It facilitates the examination of punctuality of the public vehicles and drawing conclusions.

2.1. Preliminary Limitations

Spatial and temporal limitations have been introduced:

– typical line types and lines have been selected (Table 1);
– the investigation was limited to certain workdays, which were typical from meteorological aspects (temperature and precipitation) in order to obtain relevant results.

Two-digit identifiers have been assigned to each weather category (Table 2). The following weather attributes were used: rainfall amount, rainfall intensity, minimum and maximum temperatures. The periods (hour intervals) when the weather events influenced the transportation were identified by the radar images. The investigation covered 18 representative irrespective days between 4 July 2012 and 8 May 2013. The identified weather categories were covered multiple times (at least 2 days for each category).

At a certain stop, the dwell time of the previous tours has not been considered. This may only be necessary, when the travel demand exceeds the capacity of the vehicle and passenger lags occur. In the study area, such problems do not occur, or when it still happens, it has not had any effect because of the infrequent headways of the vehicles (headways are more than 10 min, even in the rush hours too).

Table 1. Selected line types and line IDs (*source*: research by the authors)

Line type	Line ID
diametrical	11 (11B)
merging, intersecting	25, 38
suburban	22
short	2
belt-line	7, 17
loop shape	city bus

2.2. Data from the On-Board Units and Fleet Tracking System

The On-Board Units (OBUs) are performing satellite-based positioning and they have duplex data connection with the traffic control centre. OBUs record all events related to the vehicle movements (e.g. door opening, door closing, arrival and departure, etc.). Data (records) are issued in two cases:

– during uneventful periods, in every 5 s (when the vehicle is unmoving and there are no events on board);
– in event driven way, when a parameter changes (e.g. door opening, pressing the stop indicator, etc.).

Table 2. Weather categories according to the temperature and precipitation (*source*: research by the authors)

Temperature	Precipitation			
	no		yes	
cold (below –10 °C)	11.	cold and dry	21.	snowfall
0 °C around freezing point	12.	dry, around freezing point	22.	rain/sleet (drizzle)
mild	13.	uneventful from transportation viewpoint	–	
hot (~30 °C)	14.	hot and dry	24.	thunderstorms, storms (or hail)

The issued data are partly processed by the on-board computer and transmitted to the traffic management centre via GPRS/3G connection, where the data are fully processed. Delays and hurries (the actual deviation from the schedule in seconds) are determined in the centre based on the system time, schedule data and the data from the sections that have already been covered.

Data originating from the vehicle tracking system were handed over in CSV format for the research purposes. The data set included the following selected parameters:

- tracking data (date, time, speed, tour identifier, sequential number of stop, platform identifier, schedule deviations, door status);
- base data of vehicles (manufacturer, type);
- stop names and their locations;
- data regarding the network and routes;
- data regarding the tours.

As foremost step, the physical processes 'behind' the data recording were identified.

Prior to the analysis, the data have been categorized, the unnecessary records and attributes have been filtered and an Access database has been created. Table 3 demonstrates the structure of the data provided by the vehicle tracking system.

One record consists of two types of data elements:

- static 'base' data: e.g. data of vehicles, routes, stops (planned schedule);
- dynamic data (marked with grey background in the table) recorded by the OBU: time stamp, speed, door status, actual deviation from the planned schedule (+/– [s]).

2.3. Database Structure

Certain recorded data elements were irrelevant from the aspect of this research. Therefore, during the development of the database, limitations and simplifications have been carried out. Some attributes have been contracted (e.g. detailed door status) and some have been neglected (e.g. GPS coordinates, speed, stop indicator, etc.). Two records have been created per stop for each tour:

- timestamp of door opening: arrival delay;
- timestamp of door closing: departure delay.

Thus, the identification of the boundaries of the phases in the investigated process is possible. The movement phases (dwelling at stops and movements between stops) have been separated. This process was the key for the success of the analysis and it was the most complex and difficult action during the research. Positive and negative schedule deviations have been analysed, and causes of the extreme values have been revealed. According to the ISO 9001:2009 standard – which is widely used by the public transportation companies – a tour is called 'delayed' when it has more than 300 s deviation (delay) from the schedule. Tours and their schedule deviations were investigated in n×30 s intervals (*n* is an integer).

During the preliminary process of data, the most significant difficulties were the following:

- formation of two records per stop for each tour (unambiguous identification of door openings and closings);
- keeping the sequence;
- ordering data without primary keys by various criteria.

The pre-processed and filtered data have been inserted into the created database structure. Vehicle tracking data were completed with weather data, thus analysis of the weather effects on the traffic flow has become possible. Fig. 1 illustrates the structure of the database.

The ***Dwelling at stops*** table contains the tracking data. Each tour is described by two records per stop (one for the door opening and one for the door closing). These records contain date, licence plate number (*licence plate*), tour identifier (*tour ID*), sequential number and identifier of the stop (*platform*), timestamp for different door status with the schedule deviation and a *weather connecting field* that assigns the weather category to the time.

Based on the ***Dwelling at stops*** table, the ***Vehicle movements between stops*** table has been created. This table contains the data of runs between the stops for each day, vehicle and tour: departure stop (*departure platform*), arrival stop (*arrival platform*), departure and arrival time, sequential number of the stops (*number of departure and arrival stop*), delay at departure and arrival (*deviation from the schedule at departure* and *ar-*

Table 3. Structure of data recorded by the vehicle tracking system

Date	Time	Plate number	Speed	Tour ID	Sequential number of stop	Platform ID	Deviation from schedule	Arrival?	Door 1 open	Door 2 open	Door 3 open	Door 4 open	Stop indicator
4 July 2012	12:56:33	AFF-651	0	16703	2	20758	153	0	0	0	0	0	0
	12:57:20	AFF-651	10.23	16703	3	23039	140	1	0	0	0	0	1
	12:57:23	AFF-651	5.9	16703	3	23039	143	0	0	0	0	0	1
	12:57:29	AFF-651	0	16703	3	23039	149	0	1	1	0	0	0
	12:57:39	AFF-651	0	16703	3	23039	159	0	1	0	0	0	0
	12:57:40	AFF-651	0	16703	3	23039	160	0	0	0	0	0	0
	12:58:14	AFF-651	10.1	16703	4	23009	134	1	0	0	0	0	1
	12:58:17	AFF-651	6.87	16703	4	23009	137	0	0	0	0	0	1
	12:58:23	AFF-651	0	16703	4	23009	143	0	1	1	0	0	0

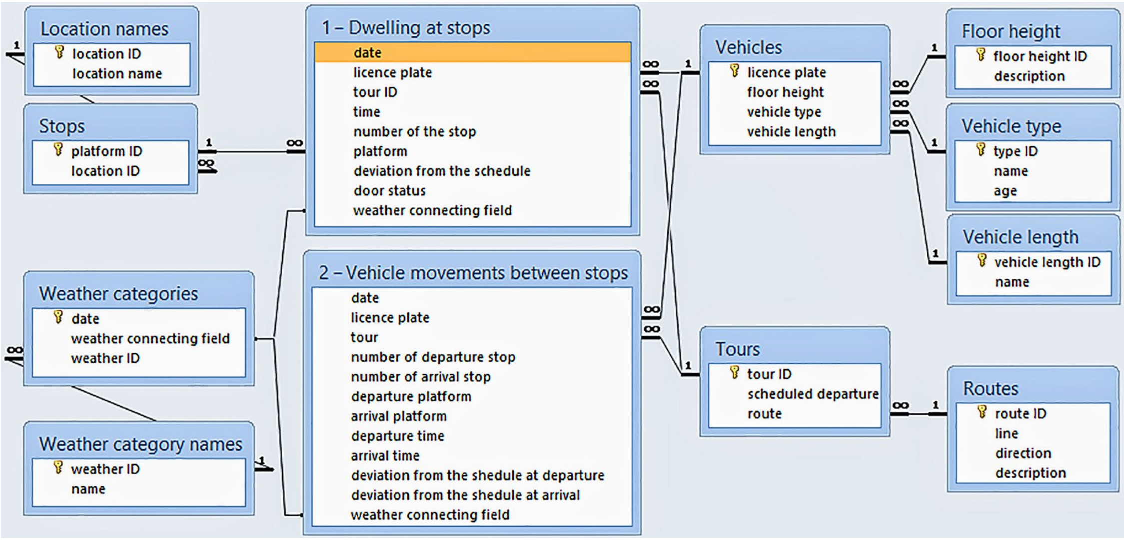

Fig. 1. Relational structure of the database (*source*: research by the authors)

rival) and the *weather connecting field* in order to assign the weather data to the traffic. In this study, data from the ***Vehicle movements between stops*** table has not been used for analysis purposes because this research paper concerns only the dwell time at stops.

The weather data are stored in the ***Weather categories*** table broken down into quarter-hour intervals. It contains the date, the identifier of the weather category (*weather ID*) and the *connecting field* in the following structure: date + hour + quarter (e.g. 2012.07.04. 10 1). In order to distinguish the weather categories, a j index has been introduced. The weather categories are described in the ***Weather category names*** table.

Because one stop name (*location name*) belongs to several geographical places (*platform ID*) – so called platform – determination of the stop name was possible only with the insertion of a cross-reference table. This table is called ***Stops***, which contains the identifier of the name of a geographical place (*location ID*) and the identifier of the stop (platform) (*platform ID*).

The characteristics of the vehicles are stored in several tables. The ***Floor height*** table contains the differentiation according to the floor height. Index k has been introduced in order to distinguish the floor height categories. The ***Vehicle type*** table contains the type and the age; in the ***Vehicle length*** table the articulated and solo buses are distinct.

The ***Tours*** table provides the details for every tour identifier (*tour ID*): scheduled departure time (*scheduled departure*) and route (*route*). The ***Routes*** table describes that a certain tour runs on which line (*line*), in which direction (*direction*) and it may contain some additional text information (*description*).

Analyses by the vehicle attributes (floor height, type, length) and the lines are also possible in the devised structure. The complete database contained 277000 records. The stops were categorized according to data of the passenger-traffic provided by manual data collection (Table 4). 38 stops have been selected on 8 lines. Each line is represented by at least 2 stops.

Table 4. Stop categories according to the volume of passenger-traffic (*source*: research by the authors)

	Category name	Limit values (passenger-traffic) [passengers/day]	Number of selected stops
1	low traffic	below 500	17
2	medium traffic	500–1000	6
3	high traffic	1000–2000	10
4	very significant traffic	over 2000	5

Unfortunately, because of the limited available resources the manual data collection could not have been extended to the assessment of the individual characteristics (age, disabilities, etc.) of the passengers.

The elaborated database structure makes possible several types of analysis. Table 5 summarizes the subjects of the analysis and the aspects. Their coupling is

Table 5. Subject of the analysis – analysis aspects matrix (*source*: research by the authors)

Subject of the analysis	Analysis aspects					
	1 time/day	2 weather category	3 floor height	4 stop category	5 lines	6 section category
dwell time at stops	×	×	×	×	×	
travel time (on the sections)	×	×				×
schedule deviations	×	×	×		×	×
summed up delays	×				×	

also determined. Investigation of the effects of vehicle types and manufacturers have been disregarded because different vehicle types are mixed assigned to the lines and there is no significant difference regarding the driving dynamics of the vehicles.

3. Results of Analysis

In this paper, only the results of analysis regarding the dwell times at stops have been published (first row of Table 5 with grey background). The following tables of the results also contain the introduced notations of the variables that are used for the description of the method for prediction (details in Section 4). The dwell times at stops fluctuate during the day, in accordance with the volume of passenger-traffic (Fig. 2). The dwell times are significantly different considering the stop categories created by the volume of passenger-traffic (*analysis aspects 1 and 4*).

Dwell times at stops are influenced by the weather effects (*analysis aspect 2*) too. It is summarized in the Table 6 – standard deviations are indicated in brackets in each cell. The shortest dwell times have been measured on the days with unfavourable weather conditions (indicated with red colour), whereas the longest dwell times have been measured during more favourable weather conditions (indicated with green colour). Presumably, passengers try to board and alight more quickly during cold and rainy weather circumstances.

Analysing the stop and weather categories at the same time (Fig. 3), it can be claimed that during dry and cold weather (weather category 11) the boarding and alighting processes have short duration irrespectively of the stop category (volume of the passenger-traffic) (*analysis aspects 2 and 4*).

Dwell times at stops are significantly influenced by the floor height. The boarding and alighting processes in the case of a low height floor vehicle are always faster. Table 7 illustrates the dwell times at stops according to floor height and stop categories – standard deviations are indicated in brackets in each cell. The values are represented with consideration only to the stop categories (third column) and the floor height (first and second columns) (*analysis aspects 3 and 4*).

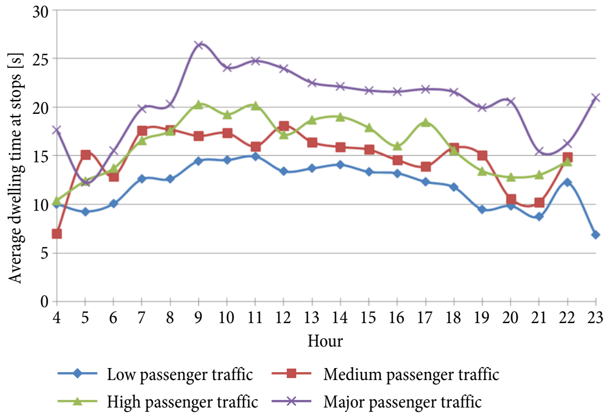

Fig. 2. Mean values of dwell time at stops according to time intervals and volume of passenger-traffic at stops (*source*: research by the authors)

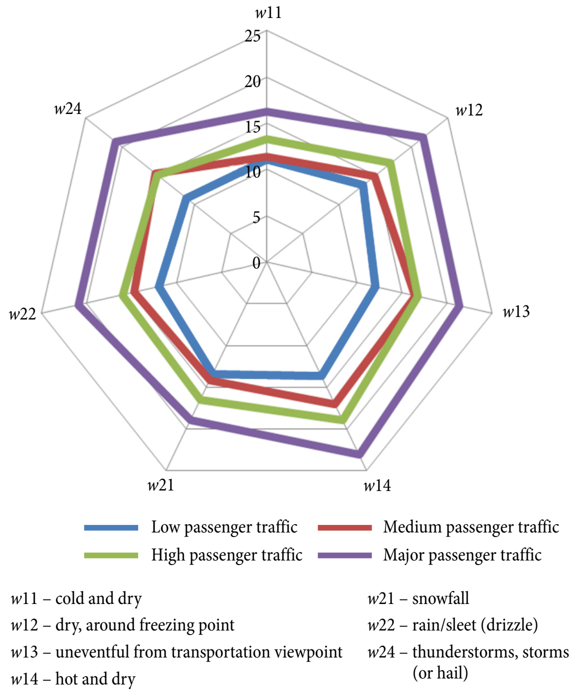

Fig. 3. Mean values of dwell times [s] at stops according to the volume of passenger-traffic and weather categories (*source*: research by the authors)

Table 6. Mean values and standard deviation of dwell times [s] at stops according to weather categories (*source*: research by the authors)

Weather category and its notation w_j	1 low floor f_l	2 high floor f_h	3 mean value of dwell times at stops (altogether) $\overline{\tau}_{w_j}$
w_{11} cold and dry	13.4 (6.6)	13 (4.9)	13.1 (5.5)
w_{24} thunderstorms, storms (or hail)	14.8 (5.4)	16.2 (5.4)	15.4 (5.2)
w_{21} snowfall	14.8 (5.3)	16.6 (4.7)	16.0 (4.9)
w_{22} rain/sleet (drizzle)	14.7 (5.5)	17.1 (4.7)	16.1 (5.1)
w_{13} uneventful from transportation viewpoint	15 (6.3)	17.9 (6.4)	16.7 (6.8)
w_{12} dry, around freezing point	15.3 (6.2)	18 (5.9)	16.9 (6.1)
w_{14} hot and dry	17 (7.6)	19.3 (7.1)	18.3 (7.5)

Table 7. Mean values and standard deviation of dwell times [s] at stops according to floor height and stop categories (*source*: research by the authors)

		1 low floor f_l	2 high floor f_h	3 altogether $\overline{\tau}$	
Mean value of dwell times $\overline{\tau}_{f_k}$		15.2 (6.6)	17.6 (6.2)	16.6 (6.4)	
1	low traffic	11.7 (6.1)	13.5 (4.8)	12.7 (5.3)	distinction by stop categories
2	medium traffic	13.7 (6.5)	17.3 (8.1)	15.6 (7.4)	
3	high traffic	15.6 (6.8)	17.7 (5.7)	16.9 (6.1)	
4	very significant traffic	19.7 (7.5)	22 (6.4)	21.1 (6.5)	

The dwell times at stops are not influenced by the capacity of the vehicle (articulated or solo buses), as the boarding is realized only through the front door. Therefore, dwell times are influenced only by the volume of passengers.

Passengers are permitted to use the public transportation vehicles with tickets and passes purchased in advance. On-board ticket and pass sale is not available. The tickets are validated on-board the vehicle, which does not influence the boarding and alighting processes. Thus, the payment method is irrelevant from the viewpoint of the research.

Dwell times have also been compared by lines (*analysis aspect 5*), but shifts in the trends have not been observed in any of the lines.

Beyond the analysis of the selected stops, general and aggregated indicators have been formed:

– The number of daily tours is different on each day. On weekdays 494 and on weekends 322 tours departed on the selected lines on average.
– Schedule deviation of the vehicles at stops is 80 s on average. The mean value of delays is 150 s, whereas the mean value of hurries is 59 s.
– Dwell time at a stop is also influenced by the schedule deviation. The greater the delay, the greater the dwell time, because the boarding and alighting process requires more time due to a higher volume of passengers. Table 8 illustrates the details according to the floor height and the delays – standard deviations are indicated in brackets in each cell. In the case of hurry, a clear trend cannot be observed; in such cases, vehicle drivers usually wait for the allotted departure time.

Based on the mean values and the low standard deviation values, it is claimed that the mean values well characterize the database. It means that the choice of the indicating factors was correct. It can be stated that the distribution of the dwell times is not a Gaussian (normal) distribution. The precise determination of the distribution needs further research.

4. Method for Prediction of Dwell Times at Stops

The aim was to develop a generally applicable method for prediction of dwell times at stops. Two solutions have been devised:

– *Method I:* historical data are fully available – clustering according to four criteria;
– *Method II:* historical data are not fully available – clustering according to two or three criteria.

Table 8. Mean value and standard deviation of dwell time [s] at stops according to the floor height and the delays (*source*: research by the authors)

Schedule deviation t	Low floor vehicle	High floor vehicle
on time ($-30 < t < 30$)	11.6 (4.7)	13.2 (4.9)
delay 1 min ($30 < t < 90$)	11.6 (3.9)	13.5 (4.1)
delay 2 min ($90 < t < 150$)	11.4 (3.7)	13.6 (4.3)
delay 3 min ($150 < t < 210$)	11.9 (4.2)	14.0 (4.0)
delay 4 min ($210 < t < 270$)	12.5 (4.6)	14.2 (3.8)
delay 5 min ($270 < t < 330$)	12.2 (3.9)	14.2 (3.8)
delay over 5 min ($t > 330$)	12.6 (4.1)	14.3 (3.9)

4.1. Method I: Historical Data are Fully Available – Custering According to Four Criteria

With the complement of tracking data (collected by OBUs) by weather information, a large amount of historical data is available. Four factors have been determined that significantly influence the dwell times at stops.

These are the following:

– stop;
– time period;
– weather;
– floor height of the vehicle.

Dwell times have been analysed according to these factors. The data structure of the database enables the inclusion of additional investigation aspects in the future.

The essence of the method is that the records are grouped by several influencing factors (in this case according to four factors) at the same time. The predicted dwell time of a stop (indicated by *i*) with *Method I*: $\tau^{i'}$ – the mean value of the dwell times [s] in a group of the data records, where the records are grouped by all factors and then queried by the actual circumstances (actual stop, time period, weather circumstances and floor height).

4.2. Method II: Historical Data are Not Fully Available – Clustering According to Two or Three Criteria

Such combination of the influencing factors of dwell times can also occur, which did not come about before. It is common at the beginning of the operation (test period), but it may occur at other times too. For these cases another calculation method has been developed. The dwell times at stops can be predicted by this method. The same influencing factors have been taken into consideration as in *Method I* and the source of the input data is the same database.

Mean values for data record sets grouped by the influencing factors 'stop' and 'time period' can be calculated in all cases, because these historical values are already available after the trial operation of the vehicle tracking system. This is not always met for the other two factors ('weather' and 'floor height'). The factors that are not represented in the database can be substituted by so called correction factors. In accordance with this, the clustering of records is carried out by only two or three aspects, and the remaining factors are considered by one or two correction factors. The method is demonstrated for a situation, when clustering is executed according to the two definitely existing factors and the missing factors are substituted by two correction factors. In this case, the calculation of dwell time at an *i* stop using *Method II* is represented in Eq. (1):

$$\tau^{i''} = \overline{\tau}_t^i \cdot c_{w_j} \cdot c_{f_k}, \quad (1)$$

where: $\tau^{i''}$ – predicted dwell time at one stop (indicated *i*) with *Method II* [s]; $\overline{\tau}_t^i$ – average dwell time at stop *i* in one time period (indicated *t*); the mean values are calculated from record sets grouped by influencing factors 'stop' and 'time period'; c_{w_j} – weather correction factor according to the *j* weather category; c_{f_k} – floor height correction factor according to the *k* floor height category (one value for the low and one value for the high floor vehicles).

The $\overline{\tau}_t^i$ average value is available for each time period after the trial operation. Hourly intervals have been used for the distinction of the time periods, but other intervals can also be applied.

c_{w_j} is a quotient, where the numerator is the average dwell time regarding one weather category $\overline{\tau}_{w_j}$ and the denominator is the mean value of the dwell times at all stops $\overline{\tau}$. This interrelation is represented in Eq. (2):

$$c_{w_j} = \frac{\overline{\tau}_{w_j}}{\overline{\tau}}. \quad (2)$$

c_{f_k} is a quotient, where the numerator is the average dwell time regarding one floor height $\overline{\tau}_{f_k}$ and the denominator is the mean value of the dwell times at all stops $\overline{\tau}$. This interrelation is represented in Eq. (3):

$$c_{f_k} = \frac{\overline{\tau}_{f_k}}{\overline{\tau}}. \quad (3)$$

$\overline{\tau}_{w_j}$ and $\overline{\tau}_{f_k}$ are mean values calculated from record sets grouped by a single aspect; whereas $\overline{\tau}$ is the mean value based on the complete database (without clustering) (see the grey coloured cells in Tables 6–7).

The interrelation in the case of clustering according to three aspects can be formed in a similar way. The advantage of the illustrated method is that the dwell time can be calculated for each situation, even for such combinations of the factors, when the historical data are not available yet.

The relative error *d* between the values determined by the two different methods and the real values can be calculated according to the Eqs (4–6):

$$d_{r-1} = \left|1 - \frac{\tau^{i'}}{\tau_r^i}\right| \cdot 100\left[\%\right]; \quad (4)$$

$$d_{r-2} = \left|1 - \frac{\tau^{i''}}{\tau_r^i}\right| \cdot 100\left[\%\right]; \quad (5)$$

$$d_{1-2} = \left|1 - \frac{\tau^{i''}}{\tau^{i'}}\right| \cdot 100\left[\%\right], \quad (6)$$

where: τ_r^i – recorded (real) dwell time [s]; d_{r-1} – relative error between the recorded dwell time values and calculated values by *Method I* [%]; d_{r-2} – relative error between the recorded dwell time values and calculated values by *Method II* [%]; d_{1-2} – relative error between the calculated dwell time values by *Method I* and *Method II* [%].

The *Method I* and *Method II* is verified, if the differences are small between the real and calculated data. In this way, the calculation of the relative error is part of the verification. As the size of the database increases, the value of *d* is decreasing due to the more detailed data.

5. Verification of the Prediction Method

Soundness of the prediction methods have been verified. The application of the methods is introduced by the following example: the selected stop is *Tihanyi Árpád út kórház* (1st influencing factor). This stop was selected because it is served by several lines, it has average volume of passenger-traffic, and it is located far from the city centre near a traffic generating facility (hospital).

The results calculated for this stop by *Method I* are shown in Table 9. The row headers contain the time intervals (2nd influencing factor), the column headers contain the weather categories (3rd influencing factor) according to the category identifier. The cells are divided into two parts according to the floor height (4th influencing factor) (low floor – blue values, high floor – red values). Some grey cells do not contain any value, because in the selected time periods some weather categories did not occur. The same notations have been used in Tables 11 and 12. Furthermore, due to the irregular assignment of vehicles to the tours, periods may occur when only low or high floor vehicles run. These cases are indicated by empty white cells. Values in cells with orange coloured background (in Tables 9 and 11) are selected to demonstrate the comparison of the two methods.

Before application of the *Method II*, determination of the average dwell times at a given stop (currently for the stop *Tihanyi Árpád út kórház*) (1st influencing factor) for each time interval (2nd influencing factor) is necessary. Table 10 summarizes the results. Furthermore, it is required to determine the correction factors for the prevailing weather situation (3rd influencing fac-

Table 9. Results [s] calculated with *Method I* (part) – clustering according to three criteria at the selected stop (*source*: research by the authors)

Hour \ Weather	w_{11}		w_{12}		w_{13}		w_{14}		w_{21}		w_{22}		w_{24}	
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
15	10	8	13	14	10	13	10	14	10	13			11	9
16	8	9	12	14	12	12	10	9	12	14		9		
17	7	11	10	13	10	13	8	11	7	12	10	13	13	
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

Table 10. Average dwell times at the given stop for different time intervals [s] (part) (*source*: research by the authors)

Hour	Average dwell time $\overline{\tau}_t^i$
…	…
15	12.44
16	11.67
17	11.32
…	…

Table 11. Results [s] calculated with *Method II* (part) (*source*: research by the authors)

Hour \ Weather	w_{11}		w_{12}		w_{13}		w_{14}		w_{21}		w_{22}		w_{24}	
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
15	9	10	12	13	11	13	13	14	11	13	11	13	11	12
16	8	10	11	13	11	12	12	14	10	12	10	12	10	11
17	8	9	11	12	10	12	11	13	10	12	10	12	10	11
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

tor) and for the given floor height (4th influencing factor). Due to space limitations, only one correction factor from each type is calculated in detail, the others can be formed in a similar way.

Weather correction factor for cold and dry condition (weather category 11):

$$c_{w_{11}} = \frac{\overline{\tau}_{w_{11}}}{\overline{\tau}} = \frac{13.1}{16.6} = 0.789.$$

Floor height correction factor for low height:

$$c_{f_l} = \frac{\overline{\tau}_{f_l}}{\overline{\tau}} = \frac{15.2}{16.6} = 0.916.$$

Tables 6 and 7 contain the values for the quotients. After the determination of input parameters, it is possible to calculate the dwell times for the circumstances when *Method I* cannot provide any result due to the lack of data.

At the selected stop (i = *Tihanyi Árpád út kórház*), in the 15–16 h time interval (t = 15), in the case of cold and dry weather category w_{11} and in the case of low floor vehicles (f_l) the calculated dwell time is:

$$\tau'' = \overline{\tau}_{15}^i \cdot c_{w_{11}} \cdot c_{f_l} = 12.44 \cdot 0.789 \cdot 0.916 \approx 9\text{ s}.$$

Values for each time period, weather category and floor height can be calculated in a similar way. Table 11 contains the results calculated by *Method II* (based on the correction factors). Calculation can be performed in the same way for the other stops.

In order to verify the methods, the relative errors d_{r-1}, d_{r-2} and d_{1-2} have been determined for a given stop (i = *Győr, Városháza*) with major passenger traffic by

Table 12. Recorded (real) and calculated dwell times and relative errors (*source*: research by the authors)

Hour	Weather	τ_r^i		$\tau^{i'}$		$\tau^{i''}$		d_{r-1}		d_{r-2}		d_{1-2}	
		f_l	f_h	f_l	f_h	f_l	f_h	f_l	f_h	f_l	f_h	f_l	f_h
4	w_{13}	15	21	14	20	13	21	7	5	15	0	8	5
5	w_{13}	16	22	17	23	18	24	6	4	11	8	6	4
6	w_{13}	18	22	20	22	19	25	10	0	5	12	5	12
7	w_{13}	25	30	23	31	26	33	9	3	4	9	12	6
8	w_{13}	22	25	21	26	25	23	5	4	12	9	16	13
9	w_{14}	17	21	17	22	18	23	0	5	6	9	6	4
10	w_{14}	12	15	11	14	14	16	9	7	14	6	21	13
11	w_{14}	11	14	13	14	10	15	15	0	10	7	30	7
12	w_{14}	15	16	16	17	15	16	6	6	0	0	7	6
13	w_{14}	18	20	18	22	19	20	0	9	5	0	5	10
14	w_{14}	17	21	17	21	18	23	0	0	6	9	6	9
15	w_{14}	25	28	25	30	26	27	0	7	4	4	4	11
16	w_{24}	31	34	32	34	31	36	3	0	0	6	3	6
17	w_{24}	22	24	21	25	20	24	5	4	10	0	5	4
18	w_{13}	19	21	18	21	17	20	6	0	12	5	6	5
19	w_{13}	14	16	14	18	13	15	0	11	8	7	8	20
20	w_{13}	13	13	12	11	13	13	8	18	0	0	8	15
21	w_{13}	11	13	10	13	12	14	10	0	8	7	17	7
22	w_{13}	12	14	12	13	12	15	0	8	0	7	0	13
23	w_{13}	11	13	10	13	11	13	10	0	0	0	9	0

Eqs (4–6). Real values were recorded for a summer day with hot weather and thunderstorms (10 August 2013) and then compared with the calculated values in case of both floor heights. Results are represented in Table 12. Differences have grouped into three categories:

- $d \leq 5\%$ green cells;
- $d > 5$ and $d \leq 15\%$ yellow cells;
- $d > 15\%$ orange cells.

The reasons for the deviation of the results calculated by the two methods are the following:

- only four influencing factors have been considered in the calculations;
- the sample of the database do not cover the whole operational period – due to the limited availability of the historical data.

Prediction becomes more and more accurate during the operation due to the increase in the volume of historical database. The correction factors are also 'dynamically' improving at the same time. More data provide more accurate predictions. Consequently, the relative error between the two methods becomes smaller and smaller. Moreover, predictions can be made by the use of *Method I* when data are available for all influencing factors.

6. Application Potentials – Further Research Directions

The results presented can be applied on one hand directly for engineering/management purposes and on the other hand for scientific purposes.

The explored relationships, the devised method are to be applied in dynamic traffic prediction models, which are used in advanced urban traffic management and control information systems for optimization purposes. The data are integrated into forecast algorithm, which takes several factors into consideration in order to calculate more precise time values. The operation of this information system is based on integrated database with data fusion techniques. As the environment, especially weather has significant impact on mode choice and transportation processes. The integration of detailed weather data requires data modelling/storage and appropriate data processes.

The information system has several functions. Predicted dwell times are to be used for headway optimization in congested areas with high mobility needs. By the information provision for passengers, the perceived quality of the services can be improved. The correction of hurries can be achieved by the introduction of rigorous regulations. When hurries occur quite often and show a tendency, then the modification of the schedule is necessary.

The elaborated and presented dwell time prediction method can also be used at other operators, who can provide the necessary input data for the calculations. Furthermore, other companies could ensure more detailed and sophisticated data, which can be utilized for more precise categorization (more categories at the influencing factors, consideration of more factors, etc.).

This research work made identification of further research potentials possible. Realization of some of them is already in progress.

- Determination of further relationships through the investigation of travel times, schedule deviations and summed up delays according to the stops and schedules (based on Table 5).
- Identification of resistance values of certain objects for network modelling of the public bus transportation. In this way, extension of macromodelling regarding road traffic (Péter, Fazekas 2014) is possible. The stops are modelled in the extended road network as special, additional 'traffic lights', where the 'material flow' is stopped. The dwell time is represented as duration of the 'red light'. In the so-called 'extended' model, all network elements (including stops too) have time and weather dependent properties.
- Involvement of new influencing factors (e.g. age, disabilities, belongings, suitcases of the passengers). As the passengers are the most important components of the transportation system, their behaviour are to be investigated. Behaviour covers both their movements (on foot, by vehicle) and their information management (decision-making) as well as the correspondences. This research direction is overlapped by cognitive sciences.
- Determination of complex correction factors that consider the effects of several 'still missing' influencing factors simultaneously (notation: c_{wfjk}).
- Development of an algorithm for travel time prediction that uses historical and real-time data too (if the OBUs are completed with passenger counters, than these data will also be used). Use of data originating from different sources at the same time: floating car data, data of passenger counting, etc.

Conclusions

The main contributions:

- A database structure has been created for various multi-variate analysis techniques, which also contains weather data beside the fleet tracking data. Detailed and precise mapping of the attributes/properties of the infrastructure and the operations related to the public transportation has been performed by us and Čertický *et al.* (2015) too, but traveller was beyond in our scope.
- The analysis has been performed in the public transportation system of city Győr. Four factors and their effects on the dwell times at stops have been revealed.
- Two methods have been elaborated for the prediction of dwell times at stops. The historical data are the basis of both methods.

Method I can be applied in full if the data sets are complete with all the influencing factors accounted for. In *Method II* the effects of the missing influencing factors are considered with correction factors. The application of the methods has been verified with examples originating from database of city Győr.

The key findings:

- Uno *et al.* (2009) suggested investigation of several factors: weather condition, design of road, land use along road, etc. We have also identified the important influencing factors as preliminary hypotheses, and then revealed their effects. Beside the ordinary and expectable effects of factors like time, volume of passengers, floor height; weather has significant effect on dwell times. Cold and hot weather with precipitation have characteristic effects, which are not obvious.
- Compared to the previous research (Fernández *et al.* 2010), it can be observed that the results of the differences of dwell times between the low and high floor height vary more than 10%, which is a consequence of the current construction properties of the vehicles.
- Observed behaviour of people during extreme and unfavourable weather conditions were beyond the expected patterns. Not only the rain has the most significant effect as Stover and McCormack (2012) claimed. The weather circumstances have to be considered as complex effects of several 'components', therefore, the components are to be investigated both separately and jointly.
- *Method II* provides results with slight (acceptable) deviation.
- Accuracy of both prediction methods can be improved by the increase in the volume of historical database.

Because the trial period of the operation of vehicle tracking systems and the test runs may last for a long time – even for more than a year – it is sufficient to collect large amount of data for the operation of *Method I.*

The lessons learnt:

- creation of a relational database based on the data of a vehicle tracking system and the separation of movement and dwell phases were the most challenging task of the research;
- assignment of independent and heterogeneous data sets to the database was time consuming and required enormous effort;
- data acquiring from different transportation organizations was hardly executable due to their lack of cooperation;
- detailed and accurate investigation of weather influence requires wide data acquiring techniques especially with regards to higher spatial resolution.

Our paper presented that how data originating from different sources (with different availability and reliability) can be used for mapping the same physical processes and how data fusion is applicable for process analysis purposes in transportation engineering.

Acknowledgements

TÁMOP-4.2.2.C-11/1/KONV-2012-0012: 'Smarter Transport' – IT for co-operative transport system. The project is supported by the Hungarian Government and co-financed by the European Social Fund.

References

Bajwa, S.; Bekhor, S.; Kuwahara, M.; Chung, E. 2008. Discrete choice modeling of combined mode and departure time, *Transportmetrica* 4(2): 155–177. http://doi.org/10.1080/18128600808685681

Currie, G.; Delbosc, A.; Harrison, S.; Sarvi, M. 2013. Impact of crowding on streetcar dwell time, *Transportation Research Record: Journal of the Transportation Research Board* 2353: 100–106. http://doi.org/10.3141/2353-10

Čertický, M.; Drchal, J.; Cuchý, M.; Jakob, M. 2015. Fully agent-based simulation model of multimodal mobility in European cities, in *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 3–5 June 2015, Budapest, Hungary, 229–236. http://doi.org/10.1109/MTITS.2015.7223261

Dueker, K. J.; Kimpel, T. J.; Strathman, J. G.; Callas, S. 2004. Determinants of bus dwell time, *Journal of Public Transportation* 7(1): 21–40. http://doi.org/10.5038/2375-0901.7.1.2

Esztergár-Kiss, D.; Kaderják, P.; Csiszár, C. 2012. Elaboration of a method's theory regarding the establishment and revision of bus lanes, *Acta Technica Jaurinensis* 5(3): 271–282.

Fernández, R.; Zegers, P.; Weber, G.; Tyler, N. 2010. Influence of platform height, door width, and fare collection on bus dwell time: laboratory evidence for Santiago de Chile, *Transportation Research Record: Journal of the Transportation Research Board* 2143: 59–66. http://doi.org/10.3141/2143-08

Fletcher, G.; El-Geneidy, A. 2013. Effects of fare payment types and crowding on dwell time: fine-grained analysis, *Transportation Research Record: Journal of the Transportation Research Board* 2351: 124–132. http://doi.org/10.3141/2351-14

ISO 9001:2009. *Quality Management Systems. Requirements.*

Levinson, H. 1983. Transit travel time performance, *Transportation Research Record* 915: 1–6.

Meng, Q.; Qu, X. 2013. Bus dwell time estimation at bus bays: a probabilistic approach, *Transportation Research Part C: Emerging Technologies* 36: 61–71. http://doi.org//10.1016/j.trc.2013.08.007

Peña, C.; Moreno, E. 2014. Delay at bus stops of transmilenio transport system according to parameters measured "in situ". Case study Bogotá – Colombia, *Procedia – Social and Behavioral Sciences* 160: 121–129. http://doi.org/10.1016/j.sbspro.2014.12.123

Péter, T.; Fazekas, S. 2014. Determination of vehicle density of inputs and outputs and model validation for the analysis of network traffic processes, *Periodica Polytechnica: Transportation Engineering* 42(1): 53–61. http://doi.org//10.3311/PPtr.7282

Soltész, T.; Kózel, M.; Csiszár, C.; Centgráf, T.; Benyó, B. 2011. Information system for road infrastructure booking, *Periodica Polytechnica: Transportation Engineering* 39(2): 55–62. http://doi.org/10.3311/pp.tr.2011-2.02

Song, X.; Teng, J.; Chen, G.; Shu, Q. 2013. Predicting bus real-time travel time basing on both GPS and RFID data, *Procedia – Social and Behavioral Sciences* 96: 2287–2299. http://doi.org/10.1016/j.sbspro.2013.08.258

Stover, V. W.; McCormack, E. D. 2012. The impact of weather on bus ridership in Pierce County, Washington, *Journal of Public Transportation* 15(1): 95–110. http://doi.org/10.5038/2375-0901.15.1.6

Tirachini, A. 2013. Bus dwell time: the effect of different fare collection systems, bus floor level and age of passengers, *Transportmetrica A: Transport Science* 9(1): 28–49. http://doi.org/10.1080/18128602.2010.520277

Uno, N.; Kurauchi, F.; Tamura, H.; Iida, Y. 2009. Using bus probe data for analysis of travel time variability, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 13(1): 2–15. http://doi.org//10.1080/15472450802644439

Vu, N. H.; Khan, A. M. 2010. Bus running time prediction using a statistical pattern recognition technique, *Transportation Planning and Technology* 33(7): 625–642. http://doi.org/10.1080/03081060.2010.512225

Watkins, K. E.; Ferris, B.; Borning, A.; Rutherford, G. S.; Layton, D. 2011. Where is my bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders, *Transportation Research Part A: Policy and Practice* 45(8): 839–848. http://doi.org/10.1016/j.tra.2011.06.010

Yu, B.; Lam, W. H. K.; Tam, M. L. 2011. Bus arrival time prediction at bus stop with multiple routes, *Transportation Research Part C: Emerging Technologies* 19(6): 1157–1170. http://doi.org/10.1016/j.trc.2011.01.003

Zhang, C.; Teng, J. 2013. Bus dwell time estimation and prediction: a study case in Shanghai – China, *Procedia – Social and Behavioral Sciences* 96: 1329–1340. http://doi.org/10.1016/j.sbspro.2013.08.151

Zito, P.; Amato, G.; Amoroso, S.; Berrittella, M. 2011. The effect of Advanced Traveller Information Systems on public transport demand and its uncertainty, *Transportmetrica* 7(1): 31–43. http://doi.org/10.1080/18128600903244727