DATA FUSION FOR TRAFFIC AND SAFETY INDICATORS: THE INTELLIGENT ROADS PERSPECTIVES

O. de Mouzon[1*], N.-E. El Faouzi[1], B. Nowotny[2], J.-M. Morin[3], E. Chung[4]
1. LICIT, INRETS-ENTPE, France.
2. Arsenal Research, Austria.
3. ISIS, France.
4. LAVOC-EPFL, Switzerland.

ABSTRACT

This work, part of the European project 'Intelligent Roads', aims to achieve real-time data collection and innovative methods for road traffic monitoring, focusing on some performance indicators: travel time and traffic density.

Data fusion provides a consistent and comprehensive picture of network conditions, based only on available data sources: No sensor expense, but improved data quantity and quality. Two main sources are used: Classical loop detectors and probe vehicle reports (floating car data, toll stamp reports). Data is collected on several European sites.

Very encouraging results are obtained on performance indicators. Data fusion will be applied to safety indicators too.

KEYWORDS
Traffic engineering, ITS, data fusion, travel time, DTA calibration, performance indicators.

INTRODUCTION

Providing real-time traffic information is one of the critical issues for the success of Intelligent Transportation Systems (ITS) programs and becomes a major challenge for public institutions and private companies, [10], [11].

Many ITS applications, such as Advanced Traffic Management Systems (ATMS), have been and are still developed in order to improve traffic performance and safety. They mainly rely on real-time data or, more precisely, indicators that are used to manage traffic flows, react on emergency cases, and inform road users. Nevertheless, most conclusions in the studies and research of this field point out that the weak spot of such applications is the poor quantity and quality of real-time data available for input.

As economical aspects are critical for ITS developments, new sensor development and implementation is more and more limited, particularly for the infrastructure part. Consequently, a better use of current sensor equipment becomes a strategic topic in ITS applications. This can be achieved through data fusion.

[*] Corresponding author. Address: 25 Av. F. Mitterrand F-69675 Bron cedex E-mail: olivier.de-mouzon@inrets.fr

In this context, the European research project INTRO (INTelligent ROads) aims to better use existing technology in an integrated way in order to provide realistic and cost-effective solutions to road safety, infrastructure management and maintenance issues. Traffic and safety monitoring, which is one of INTRO's focuses, aims to develop innovative methods for increased capacity and safety of the road network by combining traffic and weather data from different sources such as loop detectors, floating cars and meteorological stations. INTRO is supported by the European Commission and involves 10 partners from 8 European countries from the ITS and research sectors.

This paper concentrates on data fusion for two traffic performance indicators: Travel time and traffic density. The data fusion processes developed in the INTRO project aim to provide a consistent and comprehensive picture of network conditions.

DATA FUSION FRAMEWORK WITHIN INTRO

Conventional loop detector surveillance systems are expensive, and troublesome to install, to maintain and to operate. Hence more and more network operators and cities are moving towards emerging data collection systems, such as CCTV cameras for traffic detection, as well as newer methods using Automatic Vehicle Identification (AVI), Floating Car Data (FCD) and automatic vehicle classification. Using such data for traffic management purposes is becoming increasingly relevant.

Additional data sources may also be available, such as detectors for monitoring weather conditions and manual data that describes major events. Hence, for providing a consistent and comprehensive picture of network conditions, the complete fusion and integration of complementary multi-sources data has to be performed. The ability to use conventional as well as emerging data collection systems for the calibration and operation of traffic estimation and prediction systems may be an important driving force for these technologies. It may also change the way traffic data is collected and used.

The approach developed within INTRO aims to develop operational methods for road operators. These methods depend on the available sources and are tested for several source configurations for different types of networks. They allow an optimal use of data from all available sources. Namely:

- Traffic data coming from loop detectors, collected in Austria, France, Poland and Sweden.

- Floating car data (FCD) and toll stamps data, collected in Austria, France and Sweden.

- Other types of sources, such as weather or road surface conditions. Available in France, Poland and Switzerland.

As data is collected on diverse sites in several countries, the methods developed on one site will be validated at similar test sites in different EU countries.

The added value of synergistically using and combining different data sources lies not only within shared costs for shared infrastructure components, but also in intelligent services taking into account available data from more than one sensor type. Indeed, combination of different sensor data (infrastructure or in-vehicle based, weather, etc.) enables the estimations of entirely new real-time safety parameters and performance indicators to be used in traffic and safety monitoring systems. These performance indicators will allow users to be warned in real-time of accident risks, bad weather conditions, etc. Such systems will be implemented within INTRO in future work.

A major innovative feature of the current work is that it considers fusion of sensor data from vastly different sources and time lines. This is a challenge and opens new perspectives, considering the amount of data collected each year.

In the following sections, data fusion is applied to travel time and to traffic density estimation, using French and Austrian data respectively. Fusion for road safety monitoring is still under development.

DATA FUSION FOR TRAVEL TIME

Two different sources of data are considered here: Traffic data collected from loop detectors, and floating car data (mainly toll data for motorway networks). This data (traffic and toll data) has been provided by the toll motorway company Autoroutes Paris Rhin Rhône which also sponsored part of this work.

Toll data in particular is a very rich source of data, as each vehicle is a probe: The toll ticket is stamped with the actual time when it is taken at the entrance barrier and when it is given back at the exit barrier. From this source one can get the time spent in motorway sections by all vehicles and, via specifically designed filtering algorithms, one can derive travel time [4]. Although travel times from toll data are only known at the time when vehicles exit the motorway, it can be used in real time to enhance the travel time estimation. It can also be used off-line as a reference for the evaluation of travel time estimation and for motorway operation assessment.

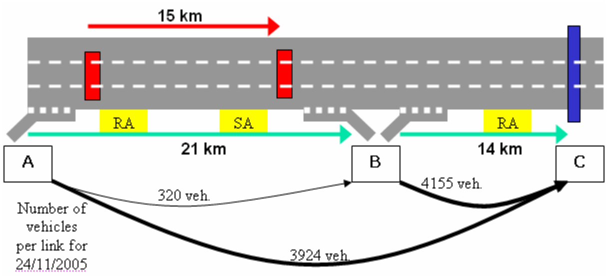

Figure 1 – The network

As shown on Figure 1, A and B are on/off-ramp tolls and C is a main toll gate. Ground based detectors (red rectangles on the road) are also available for A to B link. The fusion process can thus use toll stamps and ground detectors to compute a travel time from A to B. Rest areas (RA) and Service Areas (SA) are also represented. Finally, the number of vehicles is given for each link on November, 24th 2005. This will enable the computation of a reference travel time for A to B from toll stamps.

Reference travel time from toll stamps

Before any fusion strategy for travel time estimation, a reference travel time is necessary in order to compute and to compare performances of travel time estimations. This reference can be computed off-line from toll data collection by suitable filtering and smoothing strategies, [4]. Indeed, there can be misleading data in the toll data collections as the total time is rather time spent on the network (including stops at rest areas) rather than travel time. Hence, vehicles stopping for too long have to be filtered as they do not represent traffic dynamics such as congestion.

The left part of Figure 2 shows the result of the filtering and smoothing procedure on the direct link travel time. As is also shown, the 70 % confidence interval is rather large, especially during congestion. This is due to the fact that few vehicles made this precise link (entry in A and exit in B) during the day: only 320 vehicles.

Yet, many more vehicles went from A to C (3924) and from B to C (4155), see Figure 1. As the reference travel time can be computed off-line, it is possible to apply the smoothing and filtering procedure to the two links above and deduce the travel time from A to B. As shown on the right part of Figure 2, the 70 % confidence interval is much thinner during congestion. Thus, this travel time shape is closer to reality and is used as the reference.

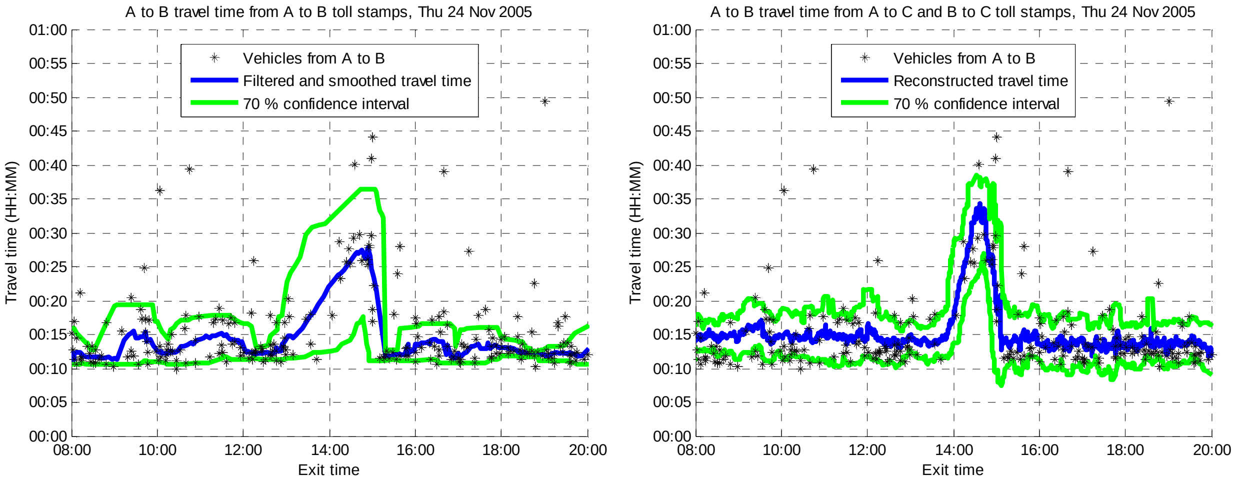

Figure 2 – From toll stamps to a reference travel time

The same process applies to any other toll C' before A, in-between A and B, or after B. Yet it is not useful to try other links, as A to C and B to C are the only links with several thousands of vehicles. However, for other cases, travel time using C and travel time using C' can be fused using their confidence intervals, leading to a more precise reference. More generally, the fusion can use all the C' of the network.

Travel time estimation

Two different kinds of fusion applied to travel time estimation can be developed:

- An *aggregation approach* aims to derive an enhanced estimation from the competing estimation schemes, and

- An *integration approach* which tries to correct bias of one method by integrating output of another method.

Along the first direction, two different approaches can be elaborated: 1) A projective approach which is able to choose at any moment the 'best' of two (or more) estimators and then produce a more robust travel time less dependent on data quality (such as breakdown of data collection stations, insufficient number of toll travel time data) than a single source estimator. Such an approach was applied to travel time estimation problems through a switching procedure using two travel time estimation algorithms and both ISIS and INRETS elaborated switching processes based on traffic patterns and/or quality indices, [5]. ISIS recently implemented its method on a motorway network in France, [15]. 2) An aggregative approach in which constant weight (at least for linear strategy) is assigned to each estimator over the whole range of input space to minimize the overall error. The main advantage of such strategy is that errors from different estimators may cancel out one another, [9].

Within INTRO the stress was therefore put on the integration approach. More specifically, the approach followed here is to correct the stock method (or its hybrid version) through the use of toll travel times.

The travel time estimates are performed using a hybrid version of the stock method which combines the speed method (Equation 1) during free flow conditions and the stock method (Equation 2) during congestion conditions.

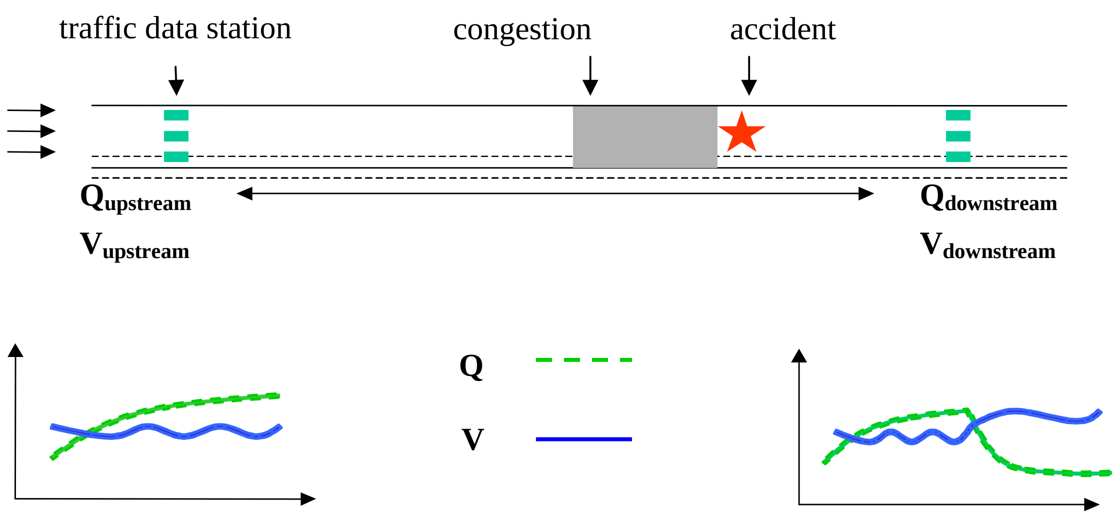

Figure 3 – The mixed method

$$TT_{Speed}(t) = \frac{L}{2} \times \left[\frac{1}{V_{upstream}(t)} + \frac{1}{V_{downstream}(t)} \right] \tag{1}$$

$$Stock(t) = Stock(t_0) + \Delta \times \sum_{k=t_0}^{k=t} \left[Q_{upstream}(k) - Q_{downstream}(k) \right]$$

$$TT_{Stock}(t) = Stock(t) / Q_{downstream}(t) \tag{2}$$

where $Q_{upstream}$, $V_{upstream}$, $Q_{downstream}$, $V_{downstream}$ denote the traffic volume and speed of the upstream and downstream traffic data station respectively, L the length of the section, Δ the elementary time slice (here 6 minutes), Stock(t) and Stock(t_0) the current and initial numbers of vehicles in the section.

To achieve the afore-mentioned switching strategy, several traffic flow pattern detection modules have been developed. Namely, a congestion detection module, a decreasing congestion trend detection module, a free flow detection module and a traffic data collection drifting detection module.

As a matter of fact, although special automatic updating procedures of the stock are already used to prevent long term drifting (the drawback of the stock method because of data collection measurement inaccuracies), there is room for further enhancement through the use of the additional information procured by the travel times of the vehicles currently exiting the motorway.

The idea is to calculate what should have been the value of the stock in the past (had the stock travel time be equal to the toll travel time) and then update the stock from this time in the past onto the current time. Once validated, this correction procedure will be incorporated into a real-time estimation tool called TPI (Temps de Parcours Instantané) designed by ISIS and already deployed on APRR motorway network, [7]. Below is an example of the travel time calculations during an accident which occurred in July ’05 on the A6 motorway.

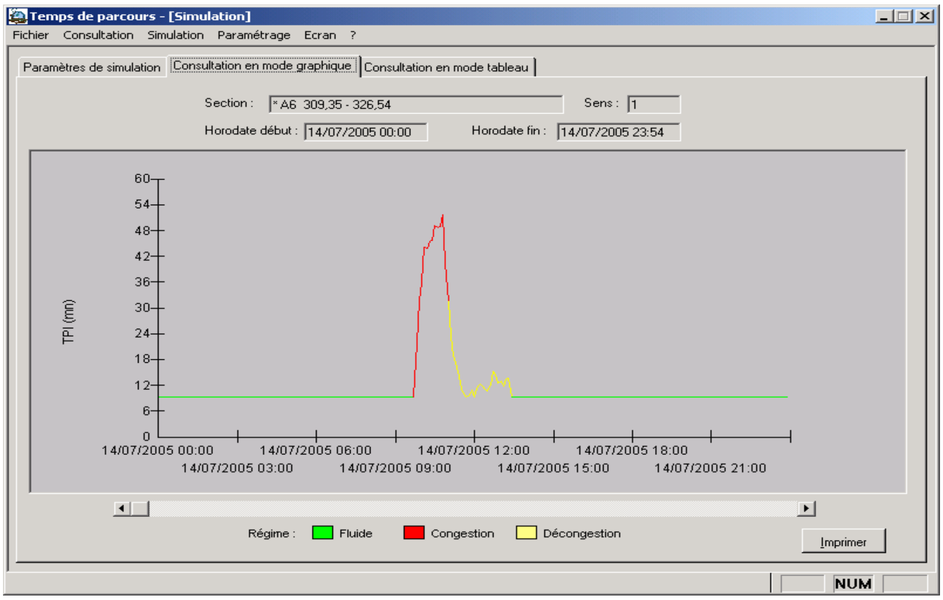

Figure 4 – An example of TPI screen during an accident

To assess the implemented correction approach, several scenarios elaborated from several real situations have been tested. It follows that the correction approach has proved to be very effective.

Figure 5 shows an example of results:

- On-line toll travel times (square dots) and reference toll travel times generated off-line (continuous curve) are in good accordance; significant variations in the reference at the beginning of the day express the stop-and-go traffic patterns, which prevail in the section before the occurrence of the accident.
- The speed method (bottom curve) fails to calculate the travel time: it only reacts when the congestion steps back onto the upstream loops.
- The stock method (curve with triangle dots) shows a rather good fit with the toll travel times, although it suffers from some underestimation and starts slightly late.
- The fusion method (curve with circle dots) clearly shows a better fit with reference toll travel times.

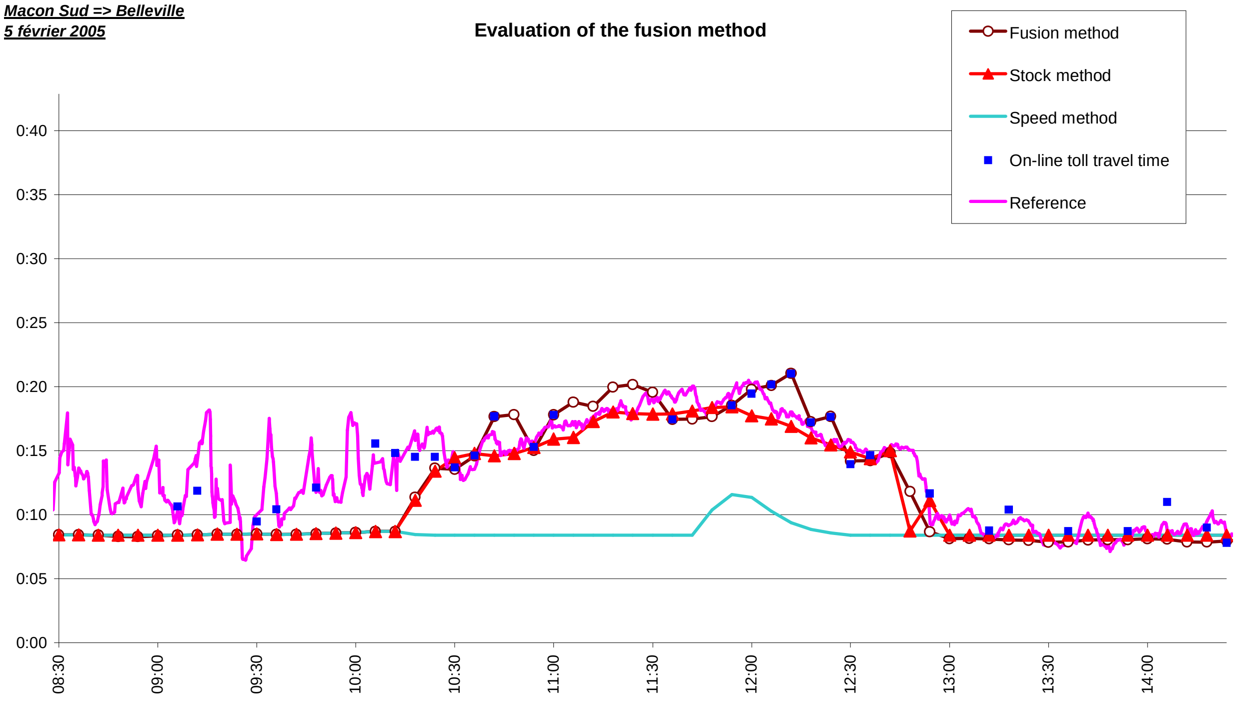

Figure 5 – An example of the results of the integration approach

Additional improvements can still be made. For example, due to the fact that the incident was not very big, the stock method did not detect the incident very quickly (at 10:18 whereas one can see longer toll travel times at 10:06 already). If the method were linked to toll travel times for detection and not only for updating, it would have detected the congestion sooner.

TRAFFIC DENSITY ESTIMATION USING DTA MODEL

Dynamic Traffic Assignment (DTA) systems have been incorporated into several traffic management centers (TMC) and can support both planning and real-time applications. DTA applications which have been implemented and tested in TMCs include DynaMIT [1], DYNASMART [14] and DINO [13]. Proper use of these applications requires calibration of a number of parameters.

Therefore, DTA models provide a flexible methodology that can be used for fusion of available data sources. They enable using Automated Vehicle Identification data (e.g. from floating cars, transponder based systems, or CCTV-based license plate recognition systems), as well as point sensor data (e.g. data from loop detectors or microwave/radar/acoustic sensors), see [2].

The objective of this part is to motivate and demonstrate the use of floating car data for DTA model calibration. One particular aspect of using floating car data is that – due to the inherent sampling of vehicles associated with the data collection technology – it is not possible to obtain direct estimates of link flows. Link flow information is very important in order to be able to generalize the sample information from the equipped vehicles to the entire traffic population. The application that is presented in this work focuses on the estimation of link flows from floating car data (in particular travel time information).

Application setup

The DynaMIT-P model [6] has been used and calibrated using floating car data for a network in the centre of the city of Vienna. The 5.3 km-long inner ring of Vienna has been selected, based on data availability. Most of the ring has three lanes, with a few links having four lanes, and the condition of the road is fairly homogeneous, e.g. in terms of lane widths and surface conditions. There are also several traffic signals. The network has been coded in the DynaMIT-P format using 17 links.

Sequential calibration has been performed, i.e. the supply was calibrated first and then the demand was calibrated using the output of the supply calibration as inputs. Only floating car data was available for the ring, as there were no reliable loop detectors operating in the network area.

Travel time data was available from the FLEET system, which uses a fleet of approximately 500 taxis operating in the entire Vienna area, equipped with GPS units [12]. Data is collected and processed in a central system. The system can generate a large number of outputs based on this information, such as point to point travel time information, turning fractions, and partial OD information. However, at the time of this application only travel time information was available and therefore only this information has been used. Data from 2.5 months of regular week-days, i.e. no holidays, were used. Due to the small sample size in the considered network (average number of equipped vehicles in each 60-min interval was between two and four per link), the link travel times for 15-min intervals showed a high variability. Hourly travel time averages showed more stable patterns and were therefore chosen for this application.

DTA calibration framework and results

The calibration of the supply side involves the estimation of segment capacities and the estimation of the parameters of the speed-density relationships for the various segments. As it was mentioned above, no reliable loop detector data was available for the study network. As a result, calibration of the speed-density relationships was based on data from a nearby link with similar characteristics to the ring. Since in general the same population drives in the two locations and since the two streets have similar characteristics, the driving dynamics experienced at the two locations are expected to be comparable. Capacities were determined using Highway Capacity Manual guidelines and average green times for the ring signals.

Given the lack of detailed information about the driving patterns throughout the network, it was assumed that the traffic dynamics (as far as the speed-density relationship is concerned) is uniform throughout the network. Therefore, a single speed density relationship has been calibrated and applied to the entire network.

Following [3], the calibration problem of a DTA model is formulated as an optimization problem. One important input to the OD estimation problem is a set of starting values for the optimization. In some applications, a planning level OD matrix is available and is used as a starting point. Such a historical OD matrix with an appropriate resolution for the test network in the city of Vienna was not available. Therefore, arbitrary (yet reasonable, given the level of traffic and capacity constraints) starting demand values were selected. As the distribution of traffic flows (i.e. path flows and turning fractions) in the network was unknown, the OD flows from node to node were estimated (i.e., dynamic estimation of unknown link flows).

The problem of OD estimation in this context is essentially an optimization problem where the unknown quantity is the matrix of OD flows that, when loaded into the simulation model, would result in measurements equal to the actual surveillance data. As the only available measurements, travel times were used for the calibration.

Historical speed time-series of each link were calculated using the harmonic average of link speeds for each time interval in each day class. A classification scheme with 4 day-classes was used for grouping. The historical time-series of a working day (outside holiday periods) was used as indirect measurement for the historical OD matrix. Estimated OD flows were used as input for DynaMIT-P in order to simulate travel times for each time interval.

Loading an OD matrix x_h to the simulation model S would then result in a set of simulated travel times TT_h^s for that interval h corrupted by a vector of errors v_h :

$$TT_h^m = TT_h^s + v_h = S(x_h) + v_h \tag{3}$$

where:

- TT_h^m is the (n_l*1) vector of measured travel times
- $TT_h^s = S(x_h)$ is the (n_l*1) vector of travel times simulated by DynaMIT
- and v_h is the (n_l * 1) vector of random error terms.

The estimation of (true) OD flows can be solved as a constrained optimization problem. If all measurements of several time periods are used simultaneously in order to make optimal use of the information contained in the measurements, a simultaneous optimization formulation is obtained. In many practical applications, a sequential optimization approach is preferred. In that case, the OD flows for each interval are estimated once and kept fixed to their estimated values during the optimization of the following intervals. This can be justified as most information about each OD flow is obtained the first time it is estimated. In this application OD flows were estimated for a period of three hours from 6 a.m. to 9 a.m. on a typical working day.

The results suggest that the calibration has indeed captured the prevailing travel times. A typical result of one calibration run is shown in the left part of Figure 6. This graph shows the simulated travel times (obtained from the calibrated DTA model) against the observed travel times for a particular link. A mild downward bias can be detected, which suggests that the calibrated travel times tend to be lower than the respective measurements.

The fit was quantified using the Normalized Root Mean Square Error (RMSN) between the observed and simulated values. The obtained RMSN values for the time periods 6am-7am, 7am-8am, and 8am-9am were 0.166, 0.155, and 0.129 respectively.

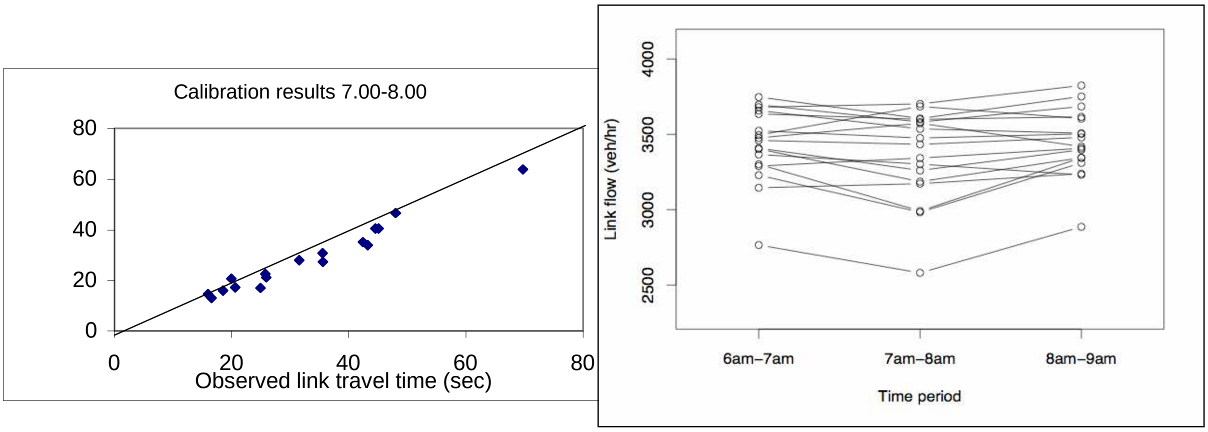

Figure 6 – Calibration results

The calibrated flows for each of the 17 links (all lanes) and each time period are shown in the right part of Figure 6. The estimated flows are reasonable and within the capacity of the network.

While no measurements of the link flows are available to compare with the estimated flows, the results are encouraging and demonstrate that the calibrated model accurately matches the historical travel times. This is an interesting finding in terms of traffic management applications, as travel times are one of the most important outputs of DTA and traffic estimation and prediction models.

Further steps include the incorporation of extra surveillance information as it becomes available. In particular, additional data can be inferred from floating car data, such as turning fractions, partial OD flows and sub-path flows. Furthermore, data from point sensors can also be incorporated to the framework, thus providing direct measurements of the speed-densityrelation, link flows, time-mean speeds and densities. This information is particularly interesting in the case of floating car data, as it can be used to generalize the information collected from the equipped vehicles (usually a small sample of the traffic) to the entire population. Applications of this methodology to extended networks (both in Vienna as well as other cities with similar data) will be valuable in strengthening the impact of these findings.

TRAFFIC SAFETY

One component of INTRO deals with safety monitoring of traffic using existing sensors data. Different traffic condition causes drivers to driver in a different manner, for example drivers slow down during inclement weather or change lane when following slower vehicle in front. Certain traffic conditions could be perceived by drivers as safe and therefore drivers are willing to take greater risk which may or may not be safe. The objective of safety monitoring is to scan for traffic induced risks in real time such as vehicles travelling at unacceptable small headways for a given weather and pavement condition. However, the objective is not to monitor all risks on the motorway such as single vehicle accidents due to drivers driving under influence of alcohol. Safety indicators are designed to capture traffic condition with respect to potential risk of accident using disaggregated (individual) traffic data. Some safety indicators are designed for individual risks while others are for platoon risks i.e. risk of a platoon of vehicles.

The sensitivity of each safety indicator may vary according to the traffic scenario i.e. some safety indicators may be more responsive for a specific scenario than other such as situations where there are lots of traffic bunching (platooning). In this aspect, fusion of all safety indicators could potentially deliver a single index that best characterise traffic safety for a complete range of traffic conditions and scenarios. Work on fusion of safety indicators is still in progress and therefore no results can be shown yet. A more robust outcome is expected from the fusion of traffic performance indicators and additional data sources as e.g. weather data and road surface data.

CONCLUSION

In this paper, data fusion proves to increase the accuracy of performance indicators, such as travel time, and to enable the estimation of performance indicators, for instance through DTA calibration.

Very encouraging results are obtained for travel time estimation by the fusion of loop and toll travel times. These results must still be confirmed and evaluated on a larger sample of real situations including various distributions of available toll travel times. It is also shown that the combination of toll travel times of different links can result in a reliable reference travel time. At this stage, the algorithms prove to be effective although further improvements, already identified, can still be and will be made during the rest of INTRO project.

Travel time and link flows fusion, under heterogeneous data source configurations, are also examined. It is found that, the outlined combining strategies improve the quality of the estimation and provide a comprehensive traffic estimate using floating car data as the primary source of information for a DTA model.

Hence, data fusion has proven very useful for traffic indicators. Building on this success, data fusion will be applied within INTRO to safety perspectives as well. The aim is to fuse already operational safety indicators in order to achieve a robust and comprehensive characterisation of traffic situation from a safety point of view.

REFERENCES

[1] Abe A. (1997). The correction of the forecasting travel time by using AVI data. In Proceedings *I.V.H.S.*

[2] Antoniou C., Ben-Akiva M., Koutsopoulos H.N. (2004). Incorporating automated vehicle identification data into origin-destination estimation. In Proceedings *10th World Congress on Transport Research*, Istanbul, Turkey.

[3] Balakrishna, R., H. N. Koutsopoulos and M. Ben-Akiva (2005). Calibration and Validation of Dynamic Traffic Assignment Systems. In Proceedings *16th International Symposium on Transportation and Traffic Theory*, Maryland, pp. 407-426.

[4] Ben Aissa A., El Faouzi N.-E, de Mouzon O., Sau J. (2005). Highway travel time estimation and prediction from toll collection system. INRETS-LICIT technical report.

[5] Ben Aissa A., El Faouzi N.-E, de Mouzon O., Sau J. (2006). Motorway travel time estimation by switching algorithms and its assessment using toll data. In Proceedings *1st International Symposium on Freeway & Tollway Operations*, Athens.

[6] Ben-Akiva, M., Bierlaire, M., Burton, D., Koutsopoulos, H. and Mishalani, R. (2001). Network state estimation and prediction for real-time transportation management applications. *Networks and Spatial Economics*, vol.1(3).

[7] Chevreuil M., Buvat J. (2001). Travel Time Calculation on the SAPRR Network: the Stock Method and Future Prospects. In Proceedings *SERTI Meeting*. Avignon.

[8] El Faouzi N.-E. (1997). Heterogeneous data source fusion for impedance indicators. In Proceedings *IFAC Symposium on Transportation Systems*, Greece, vol.3, pp.1375-1380

[9] El Faouzi N.-E. (2004). Aggregative Schemes for Estimation fusion: a road traffic application. In Proceeding *SPIE conference on Multisensor, Multisource Information Fusion, Architectures, Algorithms and Applications*, Orlando, Florida, USA.

[10] Federal Highway Administration (1996). ITS Data Fusion. Report N°WA-RD, 4010.1.

[11] Federal Highway Administration (2003). Data fusion for Delivering Advanced Traveller Information Services. In Proceedings *US DoT and ITS joint Program*, pp.247-254.

[12] Laborczi P., Nowotny B. (2004). Optimal Route Guidance on the Basis of Floating Car Data. In Proceedings *10th World Conference on Transport Research*, Istanbul, Turkey.

[13] Logi F., Ullrich M., Keller H. (2001). Traffic Estimation in Munich: Practical Problems and Pragmatic Solutions. In Proceedings *IEEE Intelligent Transportation Systems Conference*, Oakland (CA), USA.

[14] Mahmassani, H. S. (2001). Dynamic network traffic assignment and simulation methodology for advanced system management applications. *Networks and Spatial Economics*, vol.1(3), pp.267–292.

[15] Morin J.-M., Morello S. (2006). Tools to Facilitate Operator' Decision in Freeway and Tollway Traffic Control Centres. In Proceedings *1st International Symposium on Freeway and Tollway Operations*, Athens.