Data Fusion Technique in the Context of Traffic State Estimation
Irina Matschke, Bernhard Friedrich, Karsten Heinig
Institute of Transport, Road Engineering and Planning
University of Hannover, Appelstr. 9A, 30167 Hannover, Germany
matschke@ivh.uni-hannover.de; friedrich@ivh.uni-hannover.de;
heinig@ivh.uni-hannover.de;

1 Introduction
Real time traffic management in urban road networks requires information on the present and future traffic state. This information needs to be as complete and precise as possible. Today, traffic information comes from a variety of sources such as inductive loops, video observation and floating car data (FCD). To obtain a truly intelligent transportation management it is essential to have real time information. However, to get comprehensive information of nearly all lanes and turning movements based on measurement equipment is not realistic. Such supply coverage would be too expensive for most public budgets.
Therefore, the focus is to improve traffic demand estimation using sparse detector data including FCD as input combined with algorithms based on integrated data fusion techniques. These techniques contribute additional information on turning movements, flow, delay and routes and in this way enhance data coverage and data quality. The benefit of such a system will be the improved accuracy of the estimation by using the existing detection infrastructure.

2 State of the Art
Traffic state estimation needs information from different sources to provide consistent flows and travel time. These sources can be divided in measurements (data collected by detectors or coming from traffic light timing, etc.) and additional information gained by using the measurements in estimation algorithms or similar methods. Those additional information can be the estimated flows within intersections (CREMER and KELLER 1981, MAHER 1984, MIRCHANDANI ET AL. 2002), the estimated movements at intersections with traffic lights (MATSCHKE and FRIEDRICH 2001, 2002), propagated link flow counts (VORTISCH and HENGST 2002) or estimated queue lengths and waiting times at intersections with traffic lights (KIMBER and HOLLIS 1981). Data deriving from floating cars (floating car data, FCD) have recently been described as valuable supplementary information (KÜHNE ET AL. 2003). At present those data sources exist parallel and are barely used corporately.

3 Methodology of the Approach
The concept of a new method for online traffic state estimation presented in this paper is to split the system into a network level and an intersection level. On the intersection level data fusion techniques are applied in real time to combine detected flow data and information on the signal timing. The respective algorithms generate information on turning movements, queue lengths, delay and flow. On network level the enhanced data then is basis for the determination of consistent flows and travel times.
Figure 1 gives an overview on the concept and on the data, data flows and models:
- Based on the fusion of traffic counts and traffic light timings, i.e. data which is available to a certain degree for all networks, the volumes for all movements within an intersection are calculated.
- The volumes calculated in that way then are propagated to the circumjacent sectors while accuracy is considered. This procedure offers the opportunity to obtain more exact data at each detected intersection and also to fill data gaps on links where no detector is available.
Figure 1: Data fusion technique in the context of traffic state estimation

- Queue lengths are estimated using again data fusion technique by combining traffic counts and traffic light timing. Floating car data (FCD) may be used to compare and calibrate the estimation.

- Given the information on link flows (and in particular on turning movements) and using a guess of the route choice a first OD matrix can be estimated.

- A traffic assignment then uses the processed data on volumes, queue lengths and OD-relations and results in consistent flows and travel times. Based on this data a new iteration of OD-estimation and assignment is performed. Floating car data again can be used to compare and calibrate travel times as well as an additional information (weight) for the OD-estimation.

Intersection Movements

to get more and primarily precisely data a model is applied which only uses common upstream detector loops as data source. The method combines signal phase timing information and detected flow data by probability functions for each particular movement over time. Disaggregated traffic flow information collected by the upstream detectors is related to the green times of the signal timing. Using the temporary dependence of detection time and signal phasing of the different streams, estimates of the fractions could be made, because not all of the possible streams which pass the same detector could be green lighted at the same time. This approach was already implemented and tested by the authors (see Matschke and Friedrich 2002 and Matschke and Friedrich 2003) and resulted in a high performance of the estimation of turning flows at an intersection (determination coefficient of estimated compared to real turning flows > 0.9).

Flow Propagation

The volumes detected or calculated before are now propagated to the next links upstream and downstream. The approach is here to allocate as much as possible links with flow information because OD estimation results get the better the more information is made available to the estimation algorithms.
In Figure 2 the concept of link flow propagation is illustrated. The information of $q_{12}$ is calculated by propagating the detected links flows of detectors $D24$, $D26$ and $D28$ using the information of intersection movements described above. Therefore the data gap can be filled out by calculating

$$q_{12} = D24 + D26 + D28.$$ 

An upper limit for $q_{12}$ is provided by the addition of $D13$ and $D11$:

$$q_{12} \leq D11 + D13.$$ 

The smaller of both values is then selected.

Every flow value $q_{ij}$ on link $ij$ is then provided with a confidence value $c_{ij}$. This confidence value is $c_{ij} = 1$, if the flow originates from a detector and $c_{ij} = 0$, when the flow was calculated by link flow propagation.

These confidence values of link flows are later used to determine a confidence value of every OD-pair. The procedure is described later.

### Queue estimation

To involve also other additional information apart from link flows and turning movements, an approach is implemented which determines queue lengths on the basis of vehicle counts from detectors located close to the stop line and information of signal timings. This model was firstly introduced by MÜCK 2002 and the authors (see FRIEDRICH ET AL. 2003) have shown, that a high performance gain could be achieved employing this estimation module as a quasi measurement with Kalman filtering technique in queuing theory models (e.g. Markovian chains, Kimber-Hollis). Using this feed back procedure the problem of systematic instability of queue length and delay estimation could be improved especially in cases of high and varying degree of saturation.

### Introducing supplementary information to OD estimation

In recent years a large number of models have been developed to estimate an origin-destination matrix from link traffic count data (see VAN ZUYLEN and WILLUMSEN 1980), CASCETTA 1984, BELL 1983 among the most cited papers). Typically, the entropy maximising, information minimizing, and least squares estimators have been proposed and applied. The concept of the models is to update or improve an old OD matrix provided so that the estimated link volumes are consistent with the measured ones. VAN ZUYLEN, for example, has used the principle of minimum information to define the most likely OD matrix as

$$T_{ij} = t_{ij} \cdot X_{ij} \cdot \prod_a \left( \frac{\phi_a}{\phi_a} \right)^{\frac{\sigma_a}{\sigma_a}} \quad \left( g_{ij} = \sum_a p_{ij}^a \right),$$

where $t_{ij}$ is a priori guess of the OD matrix, $X_{ij} = \frac{T}{\sum_i \sum_j t_{ij}}$ a factor to include information on the total number of trips and $X^{a+1}_a = X^a_a \cdot \frac{q^{\text{real}}_a}{q^{\text{est}}_a}$ a factors for the iterative solution of the

---

**Figure 2:** Propagation of link flows
equation using observed and estimated traffic volumes $q_{a}^{real}$ and $q_{a}^{est}$ respectively for the adjustment.

Apart from historic information included in $t_{ij}$, the assignment process which converts the knowledge of the trip matrix into link flows has a high influence on the quality of OD estimation. The accuracy of the assignment information $p_{ij}^{a}$, the fraction of trips from origin $i$ to destination $j$ that passes over link $a$, is important in the matrix estimation process.

The additional information which is generated by the aforementioned approaches is made available to the OD estimation algorithms described above. It is expected, that the calculated OD matrix gets therefore a higher quality. To achieve an even higher improvement the additional information is not only used in the OD estimation process but also in the assignment. This is done by using the respective data as reference value in the cost function for calculating the link travelling cost. Furthermore, the estimated delay and queue lengths on the intersection level could be used to precise the link costs which are needed in the assignment process. And finally, a better assignment improves the OD estimation.

The aforementioned confidence values of the link flows $c_{ij}$ are now used to determine a confidence matrix. For these purposes the quantity of links for all routes of every OD pair with detected flows is correlated to the quantity of links with calculated flows. The quotient is then used as confidence value within the confidence matrix. Thus the trustfulness of an estimated OD matrix can be evaluated.

4 Computational Results

To examine the effects of the new approach on OD estimation, the network displayed in Figure 3 was designed. The reference OD matrix which is given in Figure 3 and a specific OD route fraction and OD flow was assumed. Based on these assumptions the “true” link traffic counts could be determined.

The net wide OD estimation was accomplished by the aforementioned model of VAN ZUYLEN. For our test scenarios the priori guess of the OD matrix was the unit matrix to determine the estimation quality based only on the influence of additional information and to avoid any other influence. The information on turning flows and their propagation to the circumjacent links was used as quasi traffic counts and thus the set of input data could be extended. Besides the set of counted links also the quality precision of the assignment information $p_{ij}^{a}$ has an impact of the OD estimation. The impact was analysed by comparing the quality of the OD matrix using an exact assignment, a nearly exact assignment, an assignment with some different OD flow proportioning and an all-or-nothing assignment. The exact assignment was used as the reference.

Further different scenarios had been chosen to indicate the different influences of incomplete sets of link counts. Seven scenarios with a different number of link counts were generated. Scenario 1 made up of all 38 links (14 links between intersections and 24 turning links) was used as the reference. In scenario 2 and 3 the OD matrix is estimated based only on the links between the intersections respectively on the turning links. Scenario 4 and 5 add successively turning links to the set of links to test which benefit could be gained by adding this additional information. In scenario 6 and 7 link information between intersections is removed and replaced by turning movements. An overview of the different scenarios and the corresponding sets of links is given in table 1.
Table 1: Scenarios and corresponding sets of links

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Set of links</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>all links</td>
</tr>
<tr>
<td>2</td>
<td>0, 6, 1, 34, 2, 24, 3, 28, 12, 16, 9, 31, 20, 37</td>
</tr>
<tr>
<td>3</td>
<td>4, 5, 7, 8, 10, 11, 13, 14, 15, 17, 18, 19, 21, 22, 23, 25, 26, 27, 29, 30, 32, 33, 35, 36</td>
</tr>
<tr>
<td>4</td>
<td>0, 6, 1, 34, 2, 24, 3, 28, 12, 16, 9, 31, 20, 37, 19</td>
</tr>
<tr>
<td>5</td>
<td>0, 6, 1, 34, 2, 24, 3, 28, 12, 16, 9, 31, 20, 37, 17, 21, 25</td>
</tr>
<tr>
<td>6</td>
<td>0, 6, 1, 34, 2, 24, 3, 28, 12, 16, 9, 31, 20</td>
</tr>
<tr>
<td>7</td>
<td>0, 6, 1, 34, 2, 24, 3, 28, 12, 16, 9, 31, 20, 19</td>
</tr>
</tbody>
</table>

Table 2: Incorrect additional information

<table>
<thead>
<tr>
<th>link</th>
<th>correct turning flows</th>
<th>version 1</th>
<th>version 2</th>
<th>version 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>veh</td>
<td>veh</td>
<td>veh</td>
<td>veh</td>
</tr>
<tr>
<td>17</td>
<td>20</td>
<td>15</td>
<td>100</td>
<td>15</td>
</tr>
<tr>
<td>21</td>
<td>200</td>
<td>180</td>
<td>180</td>
<td>150</td>
</tr>
<tr>
<td>25</td>
<td>60</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 3: RMSE of the test scenarios

<table>
<thead>
<tr>
<th>scenario</th>
<th>exact</th>
<th>nearly exact</th>
<th>some discrepancies</th>
<th>all-or-nothing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>veh</td>
<td>veh</td>
<td>veh</td>
<td>veh</td>
</tr>
<tr>
<td>1</td>
<td>0.19</td>
<td>4.82</td>
<td>11.58</td>
<td>40.07</td>
</tr>
<tr>
<td>2</td>
<td>19.77</td>
<td>20.02</td>
<td>29.51</td>
<td>52.53</td>
</tr>
<tr>
<td>3</td>
<td>0.30</td>
<td>8.92</td>
<td>15.88</td>
<td>34.10</td>
</tr>
<tr>
<td>4</td>
<td>14.77</td>
<td>16.13</td>
<td>23.80</td>
<td>50.33</td>
</tr>
<tr>
<td>5</td>
<td>11.97</td>
<td>13.01</td>
<td>23.97</td>
<td>45.40</td>
</tr>
<tr>
<td>6</td>
<td>27.70</td>
<td>23.98</td>
<td>33.94</td>
<td>52.62</td>
</tr>
<tr>
<td>7</td>
<td>16.31</td>
<td>19.86</td>
<td>24.98</td>
<td>49.06</td>
</tr>
</tbody>
</table>

Table 4: RMSE of scenario 5

<table>
<thead>
<tr>
<th>without additional information</th>
<th>correct turning flows</th>
<th>version 1</th>
<th>version 2</th>
<th>version 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>veh</td>
<td>veh</td>
<td>veh</td>
<td>veh</td>
</tr>
<tr>
<td>17</td>
<td>29.51</td>
<td>23.97</td>
<td>26.80</td>
<td>28.30</td>
</tr>
</tbody>
</table>

The scenarios are evaluated using the average estimation error (Route Mean Square Error - RMSE). Therefore, the estimated OD matrix is compared to the reference OD matrix and the

Figure 3: test network
average RMSE of the OD pairs is determined (the diagonal elements are not considered). A comparison of the results of the different scenarios is shown in Table 3.

The results show that the inclusion of the additional turning flows improves the OD estimation. Especially scenario 2 compared to scenario 3 shows that the OD-estimation benefits particularly when turning flows are known. Comparing scenario 6 and 7 the benefit of propagation of data could be analysed. These scenarios differ in the inclusion of the links 37 and 19 in the set of links. In case of estimating the OD matrix by including the turning link 19 an improvement could be made and completing the link set (plus link 37, scenario 4) via propagation the estimation error is reduced furthermore.

But what would happen if the determination of the turning flows were not exact and therefore the additional information were incorrect? This case is analysed in scenario 5 by falsifying the turning flows of link 17, 21 and 25 are as shown in Table 2. These first results show that the OD estimation could be improved through additional information on the turning flows even in cases of incorrect flows.

Further on, testing which impact the inclusion of delay and queue length information in the assignment will have is under process and will be finished within the next months. Based on the fist results it is expected that the OD estimation can be improved even more.

5 References


CASCETTA, E., 1984, Estimation of trip matrices from traffic counts and survey data: a generalised least squares approach estimator, Transportation Research B, 18 (4/5), 289-299


MAHER, M., 1984, Estimating the turning flows at a junction: a comparison of three models, Traffic Engineering and Control, 25, 19-22


VAN ZYULEN H., WILLUMSEN, L. G., 1980, The most likely trip matrix estimated from traffic counts”, Transportation Research B, 14, 291-293