On the short-term prediction of traffic state: an application on urban freeways in Rome

L. Mannini a *, S. Carrese a, E. Cipriani a, U. Crisalli b

Abstract

This paper explores the traffic state estimation on freeways in urban areas combining point-based and route-based data in order to properly feed a second order traffic flow model, recursively corrected by an Extended Kalman Filter. In order to overcome the possible lack of real-time information, authors propose to use simulation-based data in order to improve the accuracy of the traffic state estimation. This model was tested on a urban freeway stretch in Rome, for which a set of real-time data during the morning of a typical workday was available. Results of the application point out the benefits of the proposed approach in predicting the traffic state, as shown by GEH, RMSE and RME values similar to those presented in the literature.

© 2015 The Authors. Published by Elsevier B.V.
Selection and peer-review under responsibility of Delft University of Technology.

Keywords: freeway; traffic state estimation; extended kalman filter; traffic flow model

1. Introduction

Today, large urban areas are affected by high road traffic congestion also in off-peak periods. Therefore, the capability to forecast such traffic conditions, particularly travel times, is of utmost importance in management applications aiming at relieving negative social, environmental and economic impacts for people. Nowadays different advanced monitoring systems providing new different types of information are available, and several
models and methods have been studied and implemented in order to improve both traffic state estimation in off-line applications and travel time prediction in on-line ones. Vlahogianni et al. (2014) provide a comprehensive and updated literature review for short term traffic forecasting and underline that the combination of recent huge availability of data is moving the ITS research area from macroscopic and microscopic models based on traffic flow theories towards data driven procedures such as, for instance, methods for the travel time estimation that highly depends on available data sources, as in Schrader et al. (2004) where authors use historical data with respect to the time horizon. Differently, the spatial dimension is analyzed along links in Xie et al. (2004) and in Yang et al. (2004), or along routes in Chakroborty and Kikuchi (2004) and in Ni and Wang (2008) where trajectories are obtained.

As for data fusion techniques, many authors study and propose methods in order to use the different current data provided by the several types of detectors (Zhu and Chen, 2012; Hwang and Wu, 2012; Chu et al., 2005) or by combining real-time observations with historical ones (Wu and Shen, 2009). In Lee et al. (2009) authors observe that the combined model that assimilates historical and current data has better accuracy than the two single models. A combined use of probe and point data is also reported in Liu et al. (2010).

As for estimation issues, the most applied models are simulation models, learning models (e.g. neural networks and support vector regression based models) and statistical models (e.g. linear regression and time series, Bayesian models and Kalman Filter algorithms (KF). The Kalman Filter (KF) has been studied by many authors considering a first order traffic model, as in Yuan et al (2011), and in Zuurbier and Knoop (2006). Work and Bayen (2008) applied the Ensemble Kalman Filter (EnKF). Differently, Wang et al. (2008) and Nanthawichit et al. (2007) apply the Extended Kalman Filter (EKF) to a second order traffic model: the former reports a data testing of a real-time freeway traffic state estimator; the latter integrates probe data into the observation equation of the EKF.

In Chen and Chien (2001), authors observe that direct measurements of route based travel time rather than link based one could generate a more accurate prediction. They applied the KF both for path-based and for link-based methods. KF is also applied in Kuchipudi and Chien (2003) and the experimental results reveal that in peak hours the travel times predicted with the path-based model are more accurate than those predicted with the link-based model. Differently, Van Lint and Hoogendoorn (2010) propose a data fusion algorithm based on the Extended Generalized Treiber-Helbing filter.

Many authors combine two or more methods in order to improve the estimation of travel time. Yang et al. (2004) present a comparison between KF and adaptive recursive least-square methods, while a state-space neural network and the EKF algorithm is combined in Liu et al. (2006). In Van Lint (2008) a delayed EKF algorithm is applied and the weights of the state-space neural network model are trained with the realized travel times replying to the changes in traffic flow conditions. In Kwon and Petty (2005) the section travel times are forecasted through current probe data in order to predict the long route travel time made of the sections sequence. Zhang and Rice (2003) propose a model that updates its coefficients through historical data at each time interval, while a Bayesian dynamic linear model based on loop detector data is formulated in Fei et al. (2011), where the freeway travel time is the sum of the median values of historical travel times, taking into account the travel time random variations and a model evolution error.

Bayesian forecasting is a learning process that sequentially reviews the state of the travel time a priori knowledge based on new available data. A Bayesian estimator and an expanded neural network model are merged in Park and Lee (2004) using roadside and in-vehicle sensors; in the Bayesian procedure the forecasted travel time is an a posteriori distribution estimated on the basis of an a priori travel time distribution and current traffic measurements. Van Hinsbergen et al. (2008) combine a linear regression model and a locally weighted linear regression model in a Bayesian framework.

Rahmani et al. (2014) proposes a non parametric method for travel time estimation based on Floating Car Data (FCD) and camera data overlapping, even partially, on specific routes. A multiple linear regression model and an artificial neural network are merged in Ramakrishna et al. (2006).

This paper aims at improving the accuracy of travel time prediction when real-time point-based measurements, provided by loop detectors, are combined with route-based data, produced by an automatic number plate recognition system (ANPR), and link-based data represented by historical FCD data. The prediction model is based on a second order macroscopic traffic flow model (Wang and Papageorgiou, 2005) recursively corrected by an Extended Kalman Filter (EKF) properly fed by a measurement data fusion technique (Cipriani et al., 2012). Moreover, aiming at
overcoming the possible lack of real-time information, this paper proposes to use simulation-based data obtained through a traffic assignment model.

Specifically, in a first step a Bayesian procedure is applied to distribute the detected ANPR route travel time among links belonging to the route on the basis of: historical FCD speeds, length of the links and existence of signalized intersections; in a second step, the link apportioned travel times are combined with the real-time loop detector data, through the measurement data fusion technique; finally, EKF correction is applied to the traffic model in order to forecast the travel time on a real stretch of an urban freeway. Moreover, in order to better represent traffic dynamics by entering and exiting flows on ramps, when such data are missing, they are replaced by flows obtained through a traffic assignment model.

In the following, section 2 presents the prediction models and methods, while section 3 describes some experimental evidences obtained by an application on an urban freeway stretch of about 7 km in Rome. Finally, section 4 reports conclusions and further developments of this research.

2. Prediction models and methods

This section describes models and methods that characterize the prediction tool, whose logical architecture is reported in Fig. 1.

Given the generic time $t$ of the simulation period, for which traffic data both historical (FCD) and real-time (loop detectors and route travel times) are available, link travel performances (times and speeds) are estimated as described in section 2.1. In order to integrate data coming from different sources, a data fusion technique (see section 2.3) is applied by using the above estimated link travel speeds as well as the speeds available from loop detectors.

Fused link travel speeds, estimated link travel times, loop detector flows, as well as simulated ramp flows (when real-time data are not available), feed the macroscopic traffic flow model (see section 2.2) that allow us to obtain an a-priori estimation of link performances (i.e. link speed and density) at time $t+1$ on the basis of the traffic state at time $t$, $\hat{x}(t+1|t)$. Such an estimation is the object of a correction by an EKF to obtain the prediction of the traffic state at time $t+1$, $\hat{x}(t+1)$.

Due to possible measurement inaccuracies, traffic data are subject to a data processing, which also includes the filtering of data errors. In particular, FCD data are verified in the map-matching phase by checking data distortion on link basis, as the control between FCD speeds and link free-flow ones. Moreover, as route travel times are detected by an ANPR (Automatic Number Plate Recognition) system, the validation of route travel time data is made on the basis of the Overtaking Rule proposed by Robinson and Polak (2006), which allows to filter detected time outliers due to vehicles temporarily stopping within the monitoring area or vehicles not subject to traffic regulation (e.g. emergency vehicles).

Models and methods here briefly mentioned are detailed in the following sections.

2.1. Link travel time estimation

The procedure reported in this study, that distributes the urban route travel times on links belonging the route, has been implemented for the Rome Mobility Agency and has been tested on some routes of the city of Rome. This procedure is based on the historical information of link travel times, derived from the FCD historical speed, expressed in an a priori probability distribution from which, taking into account the detected route travel time as the conditional probability, the a posteriori probability distribution of the travel time on each link is computed using the Bayes theorem. The three probability distributions, such as a priori, conditional and a posteriori, are described in detail below.

A priori probability distribution

The a priori probability distribution has been assumed as a Gaussian distribution $N(\mu_0, \sigma_0)$, estimated on the basis of the FCD historical speed analysis on each link ($j$) belonging to the considered route, in each time interval ($5$ min, $i=1,\ldots,288$), in each daily time interval ($k=0,\ldots,6$). The mean of the distribution ($\mu_0$) is given by the length ($L_j$) of the
link divided by the average speed observed ($v_{ij}$) and the standard deviation is also provided from FCD historical deviation.

**Conditional probability distribution**

The conditional probability is defined through the detected route travel time and historical speed of FCD. This is the probability of a link travel time, based on the historical values of the FCD speed on that link, conditioned to the detected route travel time in the considered time interval. The conditional probability is assumed to be a Gaussian distribution $N(\mu, \sigma)$, where the mean ($\mu$) is the detected route travel time distributed among the links proportionally to the historical link travel times of FCD in the same time interval ($i_i$). The standard deviation should be given by the standard deviation of the route travel time of the $n$ vehicles detected and validated in the output route section, even if this study considers the historical standard deviation of the route travel time in the same time interval.

**A posteriori probability distribution**

Given the a priori probability distribution, based on historical information of FCD on each link, and the conditional probability distribution, based on the route information detected at the current time interval, the a posteriori probability of link travel times and their confidence interval are computed as follows. The a posteriori
probability distribution is assumed to be a Gaussian distribution $N(\mu_n, \sigma_n)$, where the mean ($\mu_n$) and the standard deviation ($\sigma_n$) are computed as reported below:

$$\mu_n = \frac{\mu_0 + \frac{n \mu_0}{\sigma_0^2}}{1 + \frac{n}{\sigma_0^2}} \quad \quad \sigma_n = \frac{1}{\sqrt{1 + \frac{n}{\sigma_0^2}}} \quad \quad (1)$$

**Adjusted a posteriori probability distribution**

After computing the a posteriori link travel times, the sum of all link travel times belonging to the route is compared with the detected route travel time. The difference between the observed travel time and the modeled one is distributed among upstream links to signalized intersections, proportionally to the standard deviation of FCD speed weighted by the counts of vehicles. If the route has no signalized intersection, the time difference is distributed on each link, as for upstream links to signalized intersections.

### 2.2. Macroscopic traffic flow model, measurement and EKF

The second step aims at forecasting the travel time both on links and on route; it consists in the application of Papageorgiou’s second order macroscopic traffic flow model formulation (Papageorgiou, 1987), introducing the relaxation and diffusion terms taking into account the behavioural assumption that drivers cannot adapt their speed instantaneously. Authors test an urban freeway stretch in Rome, where the length of each link is $L_j$ and the number of lanes ($n_l$) is equal to 2 for all the stretch that is 6.5 km long.

The stationary speed depends on the free-flow speed $v_f$, the critical density $\rho_{cr}$ and the exponential parameter $a$.

Moreover, the capacity of each link is given by the critical speed multiplied by the critical density. In order to model the traffic state and to forecast density on each link, authors apply both the fundamental diagram equation and the conservation equation. The dynamic of the system has been modeled as follows:

$$v_j(t+1) = v_j(t) + \frac{T_v}{\tau} \left[ \gamma_j(t) - v_j(t) \right] + \frac{T_v}{L_j} v_j(t) \left[ \gamma_j(t) - v_j(t) \right] - \frac{\eta \rho_j(t) - \rho_0(t)}{\rho_j(t) + \kappa} \cdot \frac{\rho_j(t) - \rho_0(t)}{\rho_j(t) + \kappa} + \xi_j(t) \quad \quad (2)$$

where $\xi_j(t)$ is the model noise with respect to speed, represented by a normal distribution characterized by a zero mean and $Q$ covariance; $\tau$, $\kappa$, $\eta$, are model parameters. Then, the forecasted flow is computed as the forecasted density times the number of lanes plus the model noise with respect to flow, $\xi_j(t)$.

The state equations of EKF are represented by the continuity and the second order equations that constitute a system of two partial derivative differential equations, $x(t+1) = [\rho(t+1); v(t+1)]$. Observation equations describe the relationship between traffic measurement data and state variables. They refer to loop detectors measurements (i.e. flow and speed) on link $j$ at time $t$ (Wang and Papageorgiou, 2005) as follows:

$$y^{(1)}(t) = v_j(t) \gamma_j(t) + \gamma^{(1)}_j(t) \quad \quad (3)$$

$$y^{(2)}(t) = v_j(t) + \gamma^{(2)}_j(t) \quad \quad (4)$$

where $\gamma_j$ is the measurement noise with respect to the flow ($\gamma_j^q$) or to the speed ($\gamma_j^v$), with a normal distribution characterized by a zero mean and $R$ covariance, and $y_j^{(1)}$ and $y_j^{(2)}$ are respectively the detected flow and speed in $j$ link.

The detected route travel time is distributed on each link and, taking into account link lengths, link speeds are computed and the second measurement equation is applied.

The *a posteriori* forecast of speed and density, $\hat{x}(t+1|t)$, is computed through the correction of a priori estimation, $\hat{x}(t+1|t)$, according to the correction term. The latter is the difference between the *a posteriori* forecast in current time step, $\hat{y}(t) = [\hat{y}^{(1)}(t); \hat{y}^{(2)}(t)]$, and the measurement, $y(t) = [y^{(1)}(t); y^{(2)}(t)]$, in the same time step weighted with the Kalman gain matrix (Kalman, 1960; Wang and Papageorgiou, 2005):

$$\hat{x}(t+1) = \hat{x}(t+1|t) + K_j (y(t) - \hat{y}(t)) \quad \quad (5)$$
2.3. Data fusion

The data fusion technique, which combines preprocessed data or integrates detected data by different sensors, is applied for the speed detected by loops, and the speed estimated by detected route travel time. In Cipriani et al. (2012), two different techniques are applied: state-vector fusion and measurement data fusion; in the former the EKF is used to estimate the state based on any single observation (one at a time); then, estimated states are combined; in the latter EFK is applied as a data fuser, combing data provided by different types of sensors and one state vector is estimated. Test results show that the latter technique is more accurate, so in this paper authors apply only the measurement data fusion.

The data fusion is computed before the EKF correction. Specifically, the measurement equation relative to the same variable (e.g. the speed) is split into two different equations relative to the two different sources:

\[ y_j^{(1)}(t) = C_{j}^{(1)}(t)x_j(t) + y_j^{(1)}(t) \] (6)

\[ y_j^{(2)}(t) = C_{j}^{(2)}(t)x_j(t) + y_j^{(2)}(t) \] (7)

Then, the speeds are merged on the basis of the variances of individual measurements typologies, as reported below:

\[ y_j = \frac{\frac{1}{\sigma_{j1}^2}y_j^{(1)} + \frac{1}{\sigma_{j2}^2}y_j^{(2)}}{\frac{1}{\sigma_{j1}^2} + \frac{1}{\sigma_{j2}^2}} \] (8)

where \( \sigma_{j1,2}^2 \) are the measurements variances.

Finally, the EKF is applied with the classical formulation, where the \( C \) matrix, that connects the state to measurements, \( C = \frac{\partial y(t)}{\partial x} [\tilde{x}(t|t-1)] \), and the covariance noise measurement matrix \( R \) are computed as follows:

\[ C(t) = \begin{bmatrix} C_{1}^{(1)}(t) \\ C_{1}^{(2)}(t) \end{bmatrix}, \quad R(t) = \begin{bmatrix} R_{1}^{(1)}(t) & 0 \\ 0 & R_{1}^{(2)}(t) \end{bmatrix} \] (9)

3. Experimental evidences

In order to assess the accuracy improvement in travel time prediction when different data sources are combined among them, even if this model is conceived to be applied on-line, experimental evidences are carried out by using an off-line approach, because today detected data are available only off-line.

The test concerns an urban freeway stretch in Rome of about 6.5 kilometers characterized by 6 off-ramps and 7 on-ramps, without available traffic data. The stretch consists of 32 sections of 200 meters.

The simulation period lasts 6 hours, specifically from 7 a.m. to 1 p.m. of a specific Friday; the procedure time step is 5 seconds long.

Four loop detectors are located in the 3\(^{\text{rd}}\), 9\(^{\text{th}}\), 29\(^{\text{th}}\) and 31\(^{\text{th}}\) sections. Both loop detector data and route travel time are available every 5 minutes. Data consistency is assured by the application of a linear function, which splits the available 5-min data into the 5-secs simulation procedure time steps.

The free-flow speed \( v_f \), the critical density \( \rho_{cr} \), and the exponential parameter \( a \) are set equal to 100 km/h, 35 veh/km and 2.7, respectively, while model parameters \( k, \eta, \delta, \tau \) are set equal to 13 veh/km/lane, 38 km²/h, 1.1 and 10 sec, respectively. Such parameters are obtained by an a priori historical analysis on data available for this stretch, which are similar to those presented in the literature.

Due to the lack of data, on-ramp and off-ramp flows have been estimated through an assignment model. Specifically, the 5 minutes O-D matrices have been derived on the basis of available loop detector data, considering on-ramps as origins and off-ramps as destinations; then an assignment model has been used and the obtained link out-flows have been assumed as ramp flows.
Accuracy of the estimation has been assessed through Measure of Effectiveness (MoE) statistics, which compare forecasted and detected values of route travel times and link speeds by using $GEH$, $RMSE$ and $RME$ indicators calculated as follows:

\[
GEH = \sqrt{\frac{\sum_{i,j} (x^*_i - x_j)^2}{\sum_{i,j} (x^*_i + x_j)^2}} \quad \text{RMSE} = \sqrt{\frac{\sum_{i,j} (x^*_i - x_j)^2}{n}} \quad \text{RME} = \frac{\sum_{i,j} |x^*_i - x_j|}{\sum_{i,j} x_j}
\]

where: $x^*$ is the forecasted value, $x$ is the detected value, $i$ is the time step, $j$ is the road section and $n$ is the number of measures.

As stated in the literature (U.K. Highways Agency, 1996), $GEH$ is calculated for pairs of detected and forecasted variables for different time intervals and sections. Then, the goodness-of-fit for this indicator is expressed in terms of a $GEH$ less than 5 ($GEH < 5$) calculated at least for the 85% of cases. On the other hand, $RMSE$ and $RME$ indicators take into account of large errors; therefore their values should be as small as possible.

Fig. 2 shows predicted route travel time and detected time. The reader should note that the x-axis pictures 5- min time intervals. It can be observed that the congestion starts at time interval 18 (i.e. 8:30) and lasts up to time interval 50 (i.e. 11:10). This figure also demonstrates that during two hours of congestion, the procedure is able to predict the increase of total travel time. Moreover, Fig. 2 also shows the problem in the accuracy prediction when the noisy of the detected travel time is amplified, as it is possible to see at time interval 34.

Overall, in 6 hours of simulation the results seem to be robust, as confirmed by $GEH$ of 100%, $RMSE$ of 2.5 and $RME$ of 0.17, which are reported in Table 1.

![Fig. 2. Predicted and detected route travel time comparison](image)

<table>
<thead>
<tr>
<th>MoE</th>
<th>$GEH &lt; 5$</th>
<th>$RMSE$</th>
<th>$RME$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel Time</td>
<td>100%</td>
<td>2.5</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table 1. $GEH$, $RMSE$ and $RME$ compared with route detected travel time

Deepening on results in terms of link speed, the comparison between predicted and fused detected speeds is presented through values of Table 2. As the acceptance threshold is 85%, all sections present fully acceptable $GEH$ values, even if section 31 (IV loop) presents a border value of 87%, which is due to inaccuracies of the estimated speeds in the initial time intervals, as it is possible to see from the underestimated forecasted speeds of Fig. 5 with respect to those reported in Fig.3, 4 (see the blue circles within figures), and from $RMSE$ and $RME$ values of 22.9 and 0.38, respectively.
Table 2. GEH, RMSE and RME computed with respect to measurement fusion speeds

<table>
<thead>
<tr>
<th>Section with loop detector data</th>
<th>GEH &lt; 5</th>
<th>RMSE</th>
<th>RME</th>
</tr>
</thead>
<tbody>
<tr>
<td>I loop (section 3)</td>
<td>100%</td>
<td>16.6</td>
<td>0.19</td>
</tr>
<tr>
<td>II loop (section 9)</td>
<td>100%</td>
<td>13.6</td>
<td>0.09</td>
</tr>
<tr>
<td>III loop (section 29)</td>
<td>100%</td>
<td>11.7</td>
<td>0.17</td>
</tr>
<tr>
<td>IV loop (section 31)</td>
<td>87%</td>
<td>22.9</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Finally, aggregate results in both space and time domains are summarized in Fig. 5, which shows the predicted speed along the whole stretch within the 6 hours of the simulation aggregated in 5-min simulation time intervals. It can be seen that congestion occurs in last sections and it spills back up to section 20.

Fig. 5 also shows the trend of link speeds, which is coherent with loop detector data (Fig. 3), as well as the location of congestion. Such congestion is expected in section 9, where there is a speed underestimation due to an overestimation of on-ramps flows.
4. Conclusions

This paper presented a procedure to forecast the route travel time based on different traffic data sources, both historical and real-time, provided by advanced monitoring systems. It can be used to improve the accuracy of travel time prediction when real time point-based measurements are combined with route-based and link-based data. The prediction model is based on a second order macroscopic traffic flow model recursively corrected by an Extended Kalman Filter (EKF) properly fed by a measurement data fusion technique. In order to overcome the possible lack of real-time information, the use of simulation-based data obtained through a traffic assignment model is proposed.

In order to show the goodness of the proposed approach, an off-line application example on an urban freeway stretch in Rome has been presented. The test field was characterized by traffic measurements coming from loop detectors, floating car data and historical travel times, which have been integrated with inflow and outflow data on ramps estimated through a traffic assignment model, when real-time data were not available.

Results of the application highlight the ability to predict the traffic state, especially in terms of route travel time, when real-time detected data are coherent and the input data, such as flows on ramps, are consistent. Finally, the RME and RMSE calculated through the comparison between estimated link speeds and fused detected speeds report values similar to those presented in the literature.

Further developments of this research mainly concern the on-line application, by including the investigation of the problem of spatial and temporal alignment of data, as well as the use of different Kalman filters, such as Unscented, Ensemble, Switching, and others.

Acknowledgements

Authors would like to thank the Mobility Agency of Rome that, within the project “Development of a Decision Support System for monitoring of traffic, environmental and accident”, has made available the traffic data needed to develop this study.

References


Vlahogianni E.I., Karlaftis M.G., Golas J.C. (2014). “Short-term traffic forecasting: where we are and where we’re going”. In Transportation Research Part C 43, pp 3-19.


