MOTORWAY TRAFFIC STATE ESTIMATION BASED ON EXTENDED KALMAN FILTER

Y. Wang*, M. Papageorgiou*, A. Messmer†

*Dynamic Systems and Simulation Laboratory, Technical University of Crete, 73100 Chania, Greece
†Groebenseeweg 2, D-82402, Seeshaupt, Germany

Tel: +30-28210-37289, E-mail: {ywang, markos}@dssl.tuc.gr
Tel: +49-8801-95101, E-mail: Albert.Messmer@t-online.de

Keywords: Motorway, traffic state estimation, stochastic macroscopic traffic flow model, extended Kalman filter.

Abstract: This paper addresses the design of traffic state estimator for traffic flow in motorway stretches. Motorway traffic dynamics is modelled by a stochastic macroscopic traffic flow model, while a traffic state estimator is designed based on the extended-Kalman-filtering methodology. Besides the estimation of normal traffic variables, the developed traffic state estimator deals with the on-line estimation of unknown model parameters. Simulation investigations were conducted to test the traffic state estimator under various traffic conditions in a motorway stretch.

I Introduction

Traffic state estimation had been identified as an important task within a traffic control loop already in the 1970s [13]. Traffic state estimation for a motorway network refers to estimating all traffic variables of the network at the current time instant based on available real-time traffic measurements. More precisely, based on a limited amount of available measurement data from traffic detectors, the estimation algorithms should deliver a complete image of the network’s traffic state at the current time. It should be emphasized that the number of traffic variables to be estimated may be much larger than the number of traffic variables that are directly measured, and this is in fact the essential contribution of the motorway traffic state estimation task.

For traffic state estimation, a limited number of research works produced and proposed corresponding estimation algorithms that were almost exclusively based on the seminal methodology of (extended) Kalman filtering. First applications of traffic state estimation were reported in traffic surveillance systems for short inter-detector distances [6,8,11,12,16]. In these and later investigations, e.g. [1,15], only short sections with a length below 2 km were considered, see a review in [5]. The applied models were relatively simple (due to the short inter-detector distances). Later approaches [2,3] started using more comprehensive dynamic traffic flow models, which opened the way to the consideration of longer motorway stretches (2–4 km), while more recent investigations [9,10] elaborated on some technical details based on previously proposed basic ideas. A noteworthy feature of some of these approaches is the real-time estimation not only of normal traffic variables, but also of some important traffic model parameters (e.g. the free speed, the critical density, etc.), which may change due to environmental impact (weather conditions, darkness) or due to changing traffic composition (e.g. percentage of trucks), see [4,7,16].

Based on the extended Kalman filter, this paper follows a similar avenue to traffic state estimation and exploits to the extent possible the previous approaches and investigations. The added value and innovative aspects of this paper as compared to previous investigations include:

- Development of a general approach to real-time complete traffic state estimation for motorway stretches.
- Design and application of a traffic state estimator to motorway stretches with on-ramps and off-ramps, which were not considered in previous works.
- On-line estimation of unknown important model parameters such as free speed, critical density, and exponent.
- Test of the designed traffic state estimator under various traffic conditions on a motorway stretch with on/off-ramps and a long inter-detector distance (5 km) in simulation.

The rest of the paper is organized as follows: Section 2 presents a general stochastic macroscopic model for traffic flow in motorway stretches as well as a real-time traffic measurement model, which are both organized in a state-space form as required for the design of the traffic state estimator. While constructing the state-space model, special attention is paid to the handling of the boundary conditions and unknown parameters of the macroscopic traffic flow model. In Section 3, a number of simulations are conducted to test the performance of the designed traffic state estimator and investigate the significance of on-line model parameter estimation. Finally, some conclusions are given in Section 4.

2 Modelling and methodologies

2.1 Macroscopic traffic flow model of a motorway stretch
A second-order validated macroscopic traffic flow model [14] is employed to describe dynamic traffic flow behaviour along a motorway stretch in terms of appropriate aggregate traffic variables. For the convenience of computation, a considered motorway stretch is sub-divided into a number $N$ of segments with lengths $\Delta_i$, $i = 1, \ldots, N$, while the time is discretized based on a time step $T$ and the time indices $k = 0, 1, 2, \ldots$. The aggregated traffic variables are expressed in this discrete space-time frame as follows:

- **Traffic density** $\rho_i(k)$ (in veh/km/lane) is the number of vehicles in segment $i$ at time $kT$, divided successively by the segment length $\Delta_i$ and lane number $\lambda_i$.
- **Space mean speed** $v_i(k)$ (in km/h) is the average speed of all vehicles included in segment $i$ at time $kT$.
- **Traffic flow** $q_i(k)$ (in veh/h) is the number of vehicles leaving segment $i$ during the time period $[kT, (k+1)T]$, divided by $T$.
- **Inflow** $r_i(k)$ (in veh/h) at the on ramp and outflow $s_i(k)$ (in veh/h) at the off-ramp in segment $i$ (if any).

The macroscopic model was shown to work pretty accurately with segment lengths in the order of 500 m (or less) [14]. While subdividing a motorway stretch into segments, care should be taken that any geometric inhomogeneities or installed traffic detectors along the motorway stretch are located at the boundaries of the segments. Moreover, each segment is allowed to have at most one on-ramp and one off-ramp, both preferably at the upstream boundary of the segment.

For a segment $i$, the macroscopic model equations are

$$
\rho_i(k+1) = \rho_i(k) + \frac{T}{\Delta_i} [q_{i-1}(k) - q_i(k) + r_i(k) - s_i(k)] \quad (1)
$$

$$
s_i(k) = \beta_i(k) \cdot q_i(k) \quad (2)
$$

$$
v_i(k+1) = v_i(k) + \frac{T}{T} [V(\rho_i(k)) - v_i(k)]
\quad (3)
$$

$$
V(\rho) = v_f \exp \left[ -\frac{1}{a} \left( \frac{\rho}{\rho_{\text{cr}}} \right)^\gamma \right] \quad (4)
$$

$$
q_i(k) = \rho_i(k) \cdot v_i(k) \cdot \lambda_i + \xi_i^f(k) \quad (5)
$$

where $\beta_i(k)$ (dimensionless) denotes the exiting rate at the off-ramp in segment $i$ (if any); $\tau, \nu, \kappa, v_f, \rho_{\text{cr}},$ and $a$ are model parameters which are given the same values for all segments; in particular, $v_f$ denotes the free speed, $\rho_{\text{cr}}$ the critical density, and $a$ the exponent of the stationary speed equation; $\xi_i^f(k)$ and $\xi_i^\rho(k)$ denote zero-mean Gaussian white noise acting on the empirical speed equation and the approximate flow equation, respectively, to emulate the modelling inaccuracies. Note that Equation (1) is not corrupted by noise as it reflects the conservation of vehicles, which holds strictly in any case. The model parameters are normally unknown. However, it is reported in [14] that the model results are most sensitive to variations of the free speed, critical density, and exponent. In addition, these three model parameters may vary with daytime, weather, and other external conditions. Based on the fundamental diagram $Q(\rho) = \rho \cdot V(\rho)$, the capacity of a stretch (per lane) may be deduced as

$$
q_{\text{cap}}(v_f, \rho_{\text{cr}}, a) = v_f \cdot \rho_{\text{cr}} \cdot \exp\left[ -\frac{1}{a} \right]. \quad (6)
$$

According to this segment traffic flow model, $\rho_i(k)$ and $v_i(k)$ may be viewed as (independent) segment variables of segment $i$ ($q_i(k)$ can be calculated from $\rho_i(k)$ and $v_i(k)$).

On the other hand, for each time instant $k$, traffic variables $q_{i-1}(k), v_{i-1}(k)$ and $\rho_{i+1}(k)$ as well as $r_i(k)$ and $\beta_i(k)$ (if any) are needed for calculating $\rho_i(k+1)$ and $v_i(k+1)$. These variables are boundary variables of segment $i$, incorporating the impact of the adjacent segments on the traffic dynamics of segment $i$. Clearly, the complete macroscopic model of a motorway stretch can be built upon a chain of segment models interconnected via their respective boundary variables. More precisely, the stretch model of $N$ segments consists of $2N$ equations with $2N$ independent segment variables $\rho_1, v_1, \rho_2, v_2, \ldots, \rho_N, v_N$ and a number of boundary variables: (a) flow at the stretch origin $q_0$, (b) speed at the stretch origin $v_0$, (c) density at the stretch end $\rho_N$, (d) on-ramp inflows $r_i$ (if any), and off-ramp exiting rates $\beta_i$ (if any).

### 2.2 Model of traffic measurements

Traffic detectors are installed along motorway stretches at a separation of several kilometres as a main tool for obtaining real-time traffic measurements. This paper only deals with the measurements of flow and mean speed, see also [17] for occupancy measurements.

Consider a traffic detector installed at the boundary of two adjacent segments $i$ and $i+1$. For the flow measurement, we have

$$
m_i^f(k) = q_i(k) + \gamma_i^f(k) \quad (7)
$$

where $m_i^f(k)$ denotes the flow measurement during the time period $[(k-1)T, kT]$, and $\gamma_i^f(k)$ the flow measurement noise. Except for the measurement of $q_0$, we have by (5)

$$
m_i^f(k) = \rho_i(k) \cdot v_i(k) \cdot \lambda_i + \xi_i^f(k) + \gamma_i^f(k). \quad (8)
$$

For the mean speed measurement, we have

$$
m_i^v(k) = v_i(k) + \gamma_i^v(k) \quad (9)
$$

where $m_i^v(k)$ denotes the speed measurement during the time period $[(k-1)T, kT]$ and $\gamma_i^v(k)$ the speed measurement noise. Regarding on/off-ramps, only flow measurements are
of interest. The on-ramp and off-ramp flow measurements \( m_1^r(k) \) and \( m_1^l(k) \) (if any) are modelled, respectively, as

\[
m_1^r(k) = r_1(k) + \gamma_1^r(k) \\
m_1^l(k) = s_1(k) + \gamma_1^l(k)
\]

where \( \gamma_1^r(k) \) resp. \( \gamma_1^l(k) \) denote the on-ramp resp. off-ramp flow measurement noise. All measurement noise involved in (7)–(11) is assumed zero-mean Gaussian white.

### 2.3 State-space model

Define vectors \( z \), \( d \), \( p \), \( \xi_1 \) as follows:

\[
z = \begin{bmatrix} \rho_1 & v_1 & \rho_2 & v_2 & \cdots & \rho_N & v_N \end{bmatrix}^T
\]

\[
d = [q_0 \ 0 \ \rho_{N+1} \ r_1 \ \cdots \ r_N \ \beta_1 \ \cdots \ \beta_N]^T
\]

\[
p = [v_f \ \rho_f \ \alpha]^T
\]

\[
\xi_1 = \begin{bmatrix} \rho_1 & \varepsilon_1 & \cdots & \rho_N & \varepsilon_N \end{bmatrix}^T
\]

Note that if some segments do not include an on-ramp or off-ramp, then vector \( d \) is reduced accordingly. With the aid of these vectors, the macroscopic traffic flow model of a motorway stretch can be expressed in a state-space form

\[
z(k+1) = h(z(k), d(k), p(k), \xi_1(k))
\]

where \( h \) is a nonlinear vector function corresponding to the \( 2N \) model equations. The utilization of (12) requires the real-time availability of boundary variables \( d(k) \) and the determination of the model parameters \( p(k) \). However, some elements of \( d(k) \) may not be measured or even not measurable while \( p(k) \) are normally unknown (or partially unknown). More specifically, for \( d(k) \),

- Upstream boundary flow \( q_0 \) and speed \( v_0 \) are usually measured by a traffic detector installed at the upstream boundary of the motorway stretch, while downstream boundary density \( \rho_{N+1} \) may not be directly measurable.
- On-ramp inflows \( r_i \) are sometimes not measured when traffic detectors are not installed at the corresponding on-ramps.
- Direct measurements of exiting rates \( \beta_i \) are usually not available although off-ramp outflows \( s_j \) may be measured.

In order to overcome the obstacle due to partially missing boundary measurements and unknown model parameters, (12) has to be reformed. The main idea is to “eliminate” \( d(k) \) and \( p(k) \) from the model by converting them into state variables. To this end, an auxiliary equation is introduced for \( d \) as follows

\[
d(k+1) = d(k) + \xi_2(k),
\]

where \( \xi_2(k) \) represents a vector of zero-mean Gaussian white noise. Similarly, unknown model parameters are modelled via

\[
p(k+1) = p(k) + \xi_3(k),
\]

where \( \xi_3(k) \) represents a vector of zero-mean Gaussian white noise. The covariance matrices of \( \xi_2(k) \) and \( \xi_3(k) \) must be chosen so as to reflect the typical time variations of the boundary variables and model parameters to be tracked.

Combining (12), (13) and (14) leads to the following augmented state-space model

\[
x(k+1) = f[x(k), \xi_1(k)],
\]

where \( x = [z^T \ d^T \ p^T]^T \) and \( \xi_1 = [\xi_1^T \ \xi_2^T \ \xi_3^T]^T \), and the nonlinear differentiable vector function \( f \) can be determined accordingly. As a result, all boundary variables and unknown model parameters are transformed into a part of the augmented model state \( x \) (to be estimated).

Consider a motorway stretch with traffic detectors installed at its uppermost and lowermost boundaries, at some on/off-ramps, and perhaps also at some internal segment boundaries. In terms of (15), the traffic measurements formulated by (7)-(11) can be rewritten in a compact form

\[
y(k) = g(x(k), \eta(k))
\]

where \( y(k) \) consists of all measurements of flow and mean speed; \( g \) is a nonlinear differentiable vector function; vector \( \eta \) is a function of vector \( \xi \) and vector \( \gamma \) (consisting of all measurement noise in (7)-(11)). Equations (15) and (16) constitute a complete motorway traffic dynamic system

\[
\Sigma(x, y, \xi, \eta).
\]

### 2.4 Traffic state estimator design based on the extended Kalman filter

Consider the dynamic system model \( \Sigma(x, y, \xi, \eta) \). Assume that \( \xi \), \( \eta \), and the system’s initial state satisfy some basic conditions [17]. At any time instant \( k \), given \( y(k) \) (and its values at all previous time instants), it is the goal of the traffic state estimator to deliver state estimates \( \hat{x}(k+1/k) \) so as to minimize the covariance of the estimation error

\[
J(k) = E[(x(k) - \hat{x}(k+1/k))^T \cdot (x(k) - \hat{x}(k+1/k))]
\]

where \( \hat{x}(k+1/k) \) denotes the state estimate for time instant \( k+1 \) based on measurements available up to the \( k \)-th instant. Based on the extended-Kalman-filtering method, the traffic state estimator can be designed as follows:

\[
\hat{x}(k+1/k+1) = f[\hat{x}(k+1/k), 0] + K(k)(y(k) - g(\hat{x}(k+1/k), 0)]
\]

where \( K(k) \) is the gain matrix, which is calculated on-line based on the linear Taylor-expansion of \( f \) and \( g \) at \((\hat{x}(k+1/k), 0)\) for each \( k \). Although some other canonical forms can also be constructed based on the traffic flow model and the measurement model [17], \( \Sigma(x, y, \xi, \eta) \) leads to a straightforward, general, and unique formulation of the traffic state estimator for any motorway network of any topology, size, characteristics, and with any number and locations of sensors. Note that the extended Kalman filter represents a suboptimal solution for this problem (minimization of (17)).
as an optimal filter for nonlinear systems would need infinite dimensions.

3 Simulation investigations

3.1 Simulation setup

3.1.1 Description of a motorway stretch example

A motorway stretch of 9 km is considered for the simulation investigations. The whole stretch is divided into 18 segments, each with a length of 500 m, while 10 segments in the middle of the stretch (4th to 13th; totally 5 km long) are considered for the traffic state estimation. Fig. 1 displays these 10 segments (numbered from 1 to 10) and several adjacent upstream and downstream segments. An on-ramp is located in segment 7, while an off-ramp is located in segment 9. The whole motorway stretch has three lanes. Each lane upstream of segment 12 has a capacity of 2000 veh/h, while each lane downstream of segment 12 (including segment 12) has a capacity of 1500 veh/h. The free speed, critical density, and exponential parameter for the stretch upstream of segment 12, are 120 km/h, 33.5 veh/km/lane and 1.4324, respectively. Figure 2 shows the mean trajectories of the utilized mainstream and on-ramp traffic demands, while the off-ramp exiting rate varies randomly between 4% and 18%. Traffic dynamics in the motorway stretch are simulated by use of the macroscopic traffic flow model (12), while the augmented model (15) (but without noise involved) is employed by the traffic state estimator. Since the traffic state estimator has no prior knowledge on the real model parameter values, the following assumed values are used as the initial values of the model parameters in the traffic state estimator model:

\[ q_{cap}, \beta(v_f 0, \rho_{cr0}, a_0) = 0.85 \cdot q_{cap} \quad \text{with} \quad v_f 0 = 108 \text{ km/h}, \]
\[ \rho_{cr0} = 36.85 \text{ veh/km/lane}, \quad a_0 = 1.1779. \]

The utilized time steps for the model emulating the reality and the model employed in the traffic state estimator are both 10 s. The detector configuration is displayed in Fig. 1, where the grey bars represents detectors.

For this motorway stretch, the internal segment state vector \( \mathbf{z} \) consists of the traffic densities \( \rho_i \) and space mean speeds \( v_i \) in the segments 1 to 10, while the boundary variables are upstream boundary flow \( q_0 \), upstream boundary speed \( v_0 \), downstream boundary density \( \rho_{11}, \) on-ramp inflow \( r_1 \), and off-ramp exiting rate \( \beta_0 \). Thus, we have

\[
\mathbf{x} = \begin{bmatrix} \rho_1 v_1 & \cdots & \rho_{10} v_{10} & q_0 & \rho_{11} r_1 & \beta_0 & v_f & \rho_{cr} & a \end{bmatrix}^\top,
\]
\[
\mathbf{y} = \begin{bmatrix} m_0^x & m_0^y & m_0^z & m_0^v & m_0^\rho & m_0^r & m_0^\beta \end{bmatrix}^\top.
\]

All noise considered in the investigations is zero-mean Gaussian white.

3.1.2 Traffic scenario

The traffic scenario is designed as follows:

- The mainstream traffic demand is quite fluctuated during the whole simulation horizon (Fig. 2).
- A traffic incident occurs in segment 12 at about 7:33 AM and lasts 33 minutes, during which two lanes in segment 12 are blocked. As a result, a serious traffic congestion forms and spreads rapidly upstream into the estimated part of the motorway stretch.
- The total in-coming flow of segment 12 exceeds the capacity during the peak period (10:30 AM–12:00 AM); the formed congestion propagates upstream into the estimated motorway stretch. Both congestions are built up outside the estimated part of the motorway stretch, but we expect its impact on the stretch’s traffic conditions to be appropriately tracked by the estimator based on the real-time traffic measurements.

3.2 Simulation results

A number of simulation investigations have been conducted to test the performance of the designed traffic state estimator. The length of the simulation horizon is 6 hours. A small part of the testing results are presented in Figs. 3 ~ 8. The reader may refer to [17] for more results. The unit of y-axis in each figure is “km/h” for speed, “veh/km/lane” for density, and dimensionless for the exiting rate and exponent. The estimation of three segment variables is presented in Figs. 3 ~ 5. In each of these figures six trajectories are displayed, which may be classified into two categories: (1) speed and (2) density. Each category includes three trajectories: (a) real trajectory (emulated reality), (b) estimation trajectory with on-line model parameter estimation (estimation 1), and (c) estimation trajectory without on-line model parameter estimation (estimation 2). The estimation of the exiting rate is shown in Fig. 6, while Figs. 7 and 8 display the estimates of the free speed and critical density.

The first investigation is conducted with the on-line model parameter estimation under the initial condition \( q_{cap, 0} = 0.85 \cdot q_{cap} \) (as aforementioned). Let us focus on the trajectories of the emulated reality and “estimation 1” in Fig. 3 for segment 9. From 7:00 AM to 7:30 AM, segment 9 is in the free-flow condition with a speed of about 100 km/h. At 7:33 AM an incident occurs in segment 12 and the resulting congestion spills back rapidly and reaches segment 9. As displayed, within half an hour (7:30 AM–8:00 AM) the speed drops steeply to about 15 km/h, while the density trajectory suddenly rises to 60 veh/h/lane. At about 8:07 AM, the incident is cleared, following which the speed and density values return to their former free-flow levels. From 9:00 AM to 10:30 AM, the mean value of the main demand rises from 1000 veh/h to 4800 veh/h (see Fig. 2) and maintains this level until 11:00 AM. It is depicted in Fig. 3 that the segment speed is steadily decreasing since 9:00 AM while the density increases continuously. Similar observations apply to the other segments, and indeed a second traffic congestion appears first in segment 12 at about 10:30 AM since its incoming flow exceeds the capacity, which persists until 12:00 AM; this traffic congestion reaches segment 9 and persists during 10:45 AM–11:45 AM. Obviously, segment 9 experiences various traffic conditions over the simulation horizon, and the estimated trajectories resulting when on-line model parameter estimation is activated (estimation 1) are seen to track the real trajectories under all traffic conditions. Similarly good estimation results are achieved for the other
segments (except in an initial warm-up period needed for the on-line parameter estimation [17]), see Figs. 4 and 5 for segments 7 and 3 as two examples. Note in particular the correct estimation of the space-time extent of both congestions within the freeway stretch.

Although only the estimation of $\beta_0$ is shown in Fig. 6, it should be mentioned that the boundary variables $q_0$, $v_0$, $r_f$, and $\rho_{11}$ are also estimated similarly accurately. The estimates of $v_f$, $\rho_{cr}$ and $a$ approach and fluctuate around the real values (see e.g. Figs. 7 and 8 for $v_f$ and $\rho_{cr}$), which appears to be sufficient for the traffic state estimator to deliver proper estimation results of the segment variables.

The second investigation addresses the significance of the on-line model parameter estimation, which is conducted under the same initial condition, but with the on-line model parameter estimation switched off, i.e. keeping the model parameters at their respective initial values over the whole simulation horizon. Comparison between trajectories “estimation 2” and “estimation 1” in Figs. 3 ~ 5 indicates that without the on-line parameter estimation, an estimation bias is created for each segment variable.

Some investigations were also conducted under condition $q_{cap,0} = 1.15 \cdot q_{cap}$. The conclusions already drawn from the case $q_{cap,0} = 0.85 \cdot q_{cap}$ were found to apply also to this case. In addition, some issues such as sensitivity of the traffic state estimator to initial model parameter values and dynamic tracking of time-varying model parameters are also investigated, see [17] for more details.

4 Conclusions

The investigation results provide sufficient evidence for the following conclusions:

(1) As compared to the previous investigations, a simple, general, and complete approach has been developed for traffic state estimation in motorway stretches.

(2) Satisfactory estimation results have been achieved by applying the designed traffic state estimator to a motorway stretch with on/off-ramps and a long inter-detector distance (5 km) under various traffic situations.

(3) On-line model parameter estimation is indispensable to a proper traffic state estimation, particularly in case of poor prior parameter calibration.

The development of a generic traffic state estimator for motorway networks and the performance evaluation based on real traffic data are under way and will be reported later.

Acknowledgement

This research was partly supported by the European Commission’s IST (Information Society Technologies) Program under the project RHYTHM (IST-2000-29427).

References


Fig. 1. A test motorway example.

Fig. 2. Utilized traffic demands.

Fig. 3. Segment 9.

Fig. 4. Segment 7.

Fig. 5. Segment 3.

Fig. 6. Off-ramp exiting rate.

Fig. 7. Free speed.

Fig. 8. Critical density.