Volume 41B, issue 8, October 2007 ISSN 0191-2615

TRANSPORTATION RESEARCH

AN INTERNATIONAL JOURNAL

Part B: Methodological

Editor-in-Chief: Fred L. Mannering

Available online at www.sciencedirect.com
ScienceDirect
Transportation Research Part B 41 (2007) 823–840
TRANSPORTATION RESEARCH PART B
www.elsevier.com/locate/trb

A structural state space model for real-time traffic origin–destination demand estimation and prediction in a day-to-day learning framework

Xuesong Zhou [a], Hani S. Mahmassani [b,*]

[a] Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, UT 84112, USA
[b] Department of Civil and Environmental Engineering, University of Maryland, College Park, MD 20742, United States

Received 28 November 2005; received in revised form 14 February 2007; accepted 15 February 2007

Abstract

Dynamic origin–destination (OD) estimation and prediction is an essential support function for real-time dynamic traffic assignment model systems for ITS applications. This paper presents a structural state space model to systematically incorporate regular demand pattern information, structural deviations and random fluctuations. By considering demand deviations from the a priori estimate of the regular pattern as a time-varying process with smooth trend, a polynomial trend filter is developed to capture possible structural deviations in real-time demand. Based on a Kalman filtering framework, an optimal adaptive procedure is further proposed to capture day-to-day demand evolution, and update the a priori regular demand pattern estimate using new real-time estimates and observations obtained every day. These models can be naturally integrated into a real-time dynamic traffic assignment system and provide an effective and efficient approach to utilize the real-time traffic data continuously in operational settings. A case study based on the Irvine test bed network is conducted to illustrate the proposed methodology.

Keywords: Real-time traffic estimation and prediction; Dynamic OD estimation and prediction; Kalman filter; Traffic system management

1. Introduction

The premise of Intelligent Transportation Systems (ITS) is the ability to sense prevailing conditions and rapidly devise actions to optimize system performance in real-time. Because the dynamics of traffic systems are complex, as they depend on the interaction of many independent agents (drivers) acting non-cooperatively in a spatially connected network, many situations call for strategies that anticipate unfolding conditions

* Corresponding author. Tel.: +1 301 405 0221; fax: +1 301 405 2585.
E-mail addresses: zhou@eng.utah.edu (X. Zhou), masmah@umd.edu (H.S. Mahmassani).

0191-2615/$ - see front matter
doi:10.1016/j.trb.2007.02.004

instead of adopting a purely reactive approach. Real-time simulation of the traffic network forms the basis of a state prediction capability that fuses historical data with sensor information, and uses a description of how traffic behaves in networks to predict future conditions, and accordingly develop control measures.

Dynamic origin–destination demand estimation and prediction is an important capability in its own right, and an essential support function for real-time dynamic traffic assignment model systems for ITS applications. The dynamic OD demand estimation and prediction problem seeks to estimate time-dependent OD trip demand patterns at the current stage, and predict demand volumes over the near and medium terms in a general network, given historical demand information and real-world traffic measurements from various surveillance devices (e.g. occupancy and volume observations from loop detectors on specific links).

Substantial research efforts have been devoted to dynamic demand estimation and prediction problems over the past 20 years. Existing models can be grouped into two classes: DTA-based vs. non-DTA-based, depending on whether a DTA component is incorporated into the estimation process (Kang, 1999; Chang and Tao, 1999; Peeta and Ziliaskopoulos, 2001). In this paper, existing models are categorized according to the underlying assumptions in representing dynamic demand processes. Assuming that the deviations of flow (demand) from historical averages define a stationary time series, the first group applies autoregressive (AR) models to the recursive estimation and prediction process. In the Kalman filtering formulation proposed by Okutani and Stephanedes (1984), the original data is first detrended from historical observations, then an AR model is used to estimate and forecast time-varying traffic flows on a single link. Along the same line, Ashok and Ben-Akiva (1993, 2000) formulated deviations of OD demand from historical averages as AR processes, and further developed a Kalman filter for real-time OD demand estimation and prediction, in which a 4th-order AR model is adopted based on several data sets. In general, an autoregressive model is suitable to describe a stationary random process with constant mean and variance. On the other hand, if the prevailing OD demand is structurally different from the regular demand pattern, demand deviations will not satisfy the fundamental stationarity assumption for AR processes, and such non-stationarity could seriously degrade the overall prediction performance. In addition, an AR type model with high-order terms requires extensive off-line calibration effort for the autocorrelation coefficients, and the corresponding augmented state space also dramatically increases the on-line computational burden, especially for large-scale network applications.

Alternatively, without requiring prior demand information, a simple random walk model can be relatively easily built for short-term demand prediction, corresponding to an AR(1) model with autocorrelation coefficient of 1. Cremer and Keller (1981, 1987), as well as Chang and Wu (1994) applied the random walk model to predict dynamic OD flow split parameters, by directly extending the latest estimates as the future forecasts. Although this model is effective for a slowly changing process, it might not be rich enough to capture non-linear trends in time-varying OD flows, especially for medium term prediction. In order to describe the non-linearity in dynamic OD demand, Kang (1999) and Mahmassani et al. (1998) proposed a polynomial trend filter to estimate time-dependent OD flows on a general network. This model used historical information, instead of on-line traffic measurements, to calibrate demand evolution processes. The filter was applied to the OD demand values directly rather than to deviations of these from a priori best-estimates, unlike the approach devised in the present paper.

In a closely related problem area, approaches for off-line time-varying OD demand estimation have also been proposed in the past decade, mostly for operational planning applications. Using a simplified assignment model, Cascetta et al. (1993) presented a generalized least-squares framework for estimating time-varying demand in a network. A bi-level DTA-based time-varying demand estimation formulation was introduced by Tavana (2001) and further extended by Zhou et al. (2003) to utilize multi-day link counts. In contrast, little attention has been given to procedures for effectively and systematically updating the historical demand information for on-line estimation and prediction purposes. Ashok (1996) suggested several heuristic approaches to update the historical demand estimate with recent estimates obtained in real-time, but no optimal updating formulation was given.

In general, regular OD trip desires can be viewed as a repeated process with similar within-day dynamic patterns. By utilizing knowledge from household interview surveys and off-line estimation results on multiple days, historical demand data represents the a priori estimate of the regular OD demand pattern. In particular, in the context of long-range demand prediction, reliable historical data can serve as an informative source under normal conditions. On the other hand, it is necessary to recognize the possible existence of structural deviations of real-time OD demand from the average pattern; these might be caused by severe weather conditions, special events, as well as the responses of travelers to information and/or other system management measures. The first two factors have been well recognized as critical determinants in the effectiveness of travel demand management systems. With increasing availability, traveler information, particularly, pre-trip information, is expected to play a more active role in gradually changing day-to-day trip-making decisions and the resulting temporal distributions of OD demand. In addition, random fluctuations would still account for the effect of other unobserved factors and the inherent stochastic nature of daily time-varying demand.

In the early deployment of real-time OD estimation and prediction, a common issue is that only unreliable historical demand data with significant uncertainty is available, often consisting of out-of-date survey data and limited surveillance data. In this case, as the prior estimate cannot adequately describe the average conditions, the real-time estimate becomes more informative in the sense that it captures the prevailing demand pattern and encapsulates up-to-date demand information.

To provide accurate and robust demand estimation and prediction for real-time dynamic traffic assignment in operational settings, the following primary functional requirements need to be satisfied: (1) incorporate regular demand information into the real-time demand prediction process; (2) recognize and capture possible structural changes in demand patterns under various conditions; and (3) optimally update the a priori estimate of the regular pattern using new real-time estimation results and traffic observations.

In this work, actual dynamic OD demand is decomposed to three meaningful components in a structural state space model, namely,

$$\text{true demand} = \text{regular pattern} + \text{structural deviations} + \text{random fluctuations}.$$

The next section first describes a rolling horizon execution framework for real-time OD estimation and prediction in connection with real-time DTA simulators, followed by the introduction of a structural state space model for real-time OD estimation and prediction. By considering demand deviations from the a priori estimate of the regular pattern as a time-varying process with smooth trend, a polynomial trend filter is developed as the core model to capture possible structural deviations in real-time demand. An optimal adaptive procedure is further proposed to capture day-to-day demand evolution, and a case study based on the Irvine test bed network is conducted at the end to illustrate the proposed methodology. While the use of Kalman filter approaches in the context of real-time simulation-based DTA in a rolling horizon framework essentially defines the current state of the art, the present work is the first to (1) incorporate and model systematic structural deviations from a "regular" historical pattern, as well as (2) propose and implement an adaptive day-to-day updating procedure that is internally consistent with the real-time Kalman filter model.

2. Framework of real-time DTA-based OD estimation and prediction

The rolling horizon framework in this paper follows the system design of a real-time dynamic traffic assignment system (Mahmassani et al., 1994; Mahmassani, 2001; Tavana and Mahmassani, 2001). The scheme entails sequential execution of the OD estimator and predictor, in conjunction with real-time DTA simulators. As shown in Fig. 1, the prediction (or planning) horizon represents the time length for which forecasted OD demand should be available for the DTA simulator. The prediction horizon starts at the end of a roll period, which is the time shift between the respective beginning of consecutive prediction horizons. Predictions for a given period are based on the estimation results obtained during the roll period, using observations streaming in real-time over a certain observation period.

In the above approach, a thorny modeling issue in OD estimation is how to handle lagged OD demand on current link observations. This issue arises because each traveler takes a certain time to complete his/her trip in a large city network, and the resulting travel time can be very long depending on trip length and prevailing traffic conditions. Failure to recognize the existence of lagged demand would attribute all current flows to demands departing during the current estimation stage, potentially leading to serious bias in estimation results. One possible solution is to extend the dimension of the state variable vector so as to include all the lagged OD demand variables in the current estimation stage (Okutani and Stephanedes, 1984), but the resulting expanded state space could significantly increase the computational complexity. The proposed polynomial trend model offers a compact representation of lagged demands, as described in a later section.

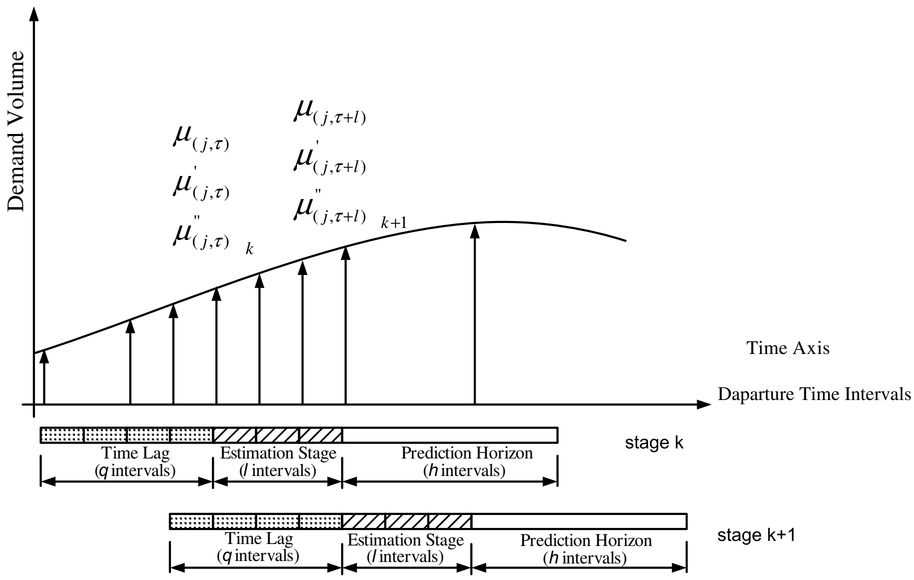

Fig. 1. Illustration of rolling horizon implementation.

The rolling horizon implementation of real-time OD estimation and prediction is stated as Algorithm 1. The discussion here is intended to explain the role of the real-time OD estimation and prediction capability and its connection to real-time traffic simulator in the context of the DTA system under consideration. Fig. 2 depicts the demand data flow in a real-time DTA system. Specifically, the OD estimation module estimates the OD demand at the current stage based on real-time traffic measurements, historical OD demand pattern information, as well as link proportions obtained from the DTA simulator. The estimated demand pattern is fed into the OD demand predictor to provide dynamic OD demand input for the traffic assignment and simulation modules in future stages. To ensure the consistency between simulated states and actual conditions, the OD demand predicts are regulated by the OD demand consistency checking module (see Zhou and Mahmassani, 2005) before being loaded into the real-time network state estimator (RT-DYNA). Finally, real-time OD demand estimates obtained on a new day are extracted to update the regular demand pattern database, which in turn provides the a priori estimate of the regular demand pattern for real-time demand estimation on the next day.

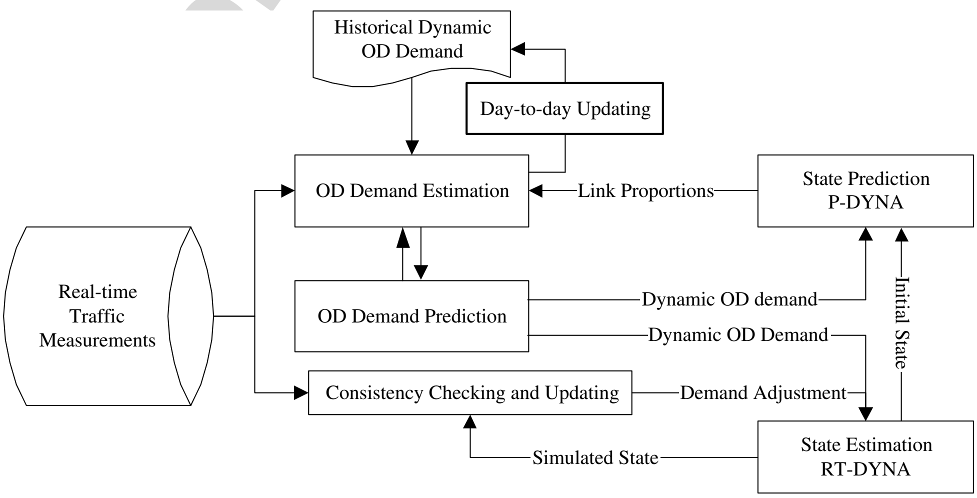

Fig. 2. Demand data flow in a real-time DTA system.

Algorithm 1. Rolling horizon implementation for real-time OD estimation and prediction

Step 1: Receive real-time traffic measurements from surveillance system.
Step 2: Fetch link proportion data for the current estimation stage from the DTA simulator.
Step 3: (OD estimation) Estimate time-varying OD demand matrices involved in the current estimation stage using the Kalman filtering method.
Step 4: (OD prediction) Predict OD demand over next future horizon.
Step 5: Advance roll period forward, and then go back to Step 1.

3. Structural model for real-time OD estimation and prediction

For convenient reference, the notation used in the real-time OD estimation and prediction model is first presented, as follows:

i index for links with traffic measurements, $i = 1, \ldots, N_{\text{obs}}$
j index for origin–destination pairs, $j = 1, \ldots, N_{\text{od}}$
τ index for aggregated departure time intervals, $\tau = 1, 2, \ldots$
t index for observation time interval, i.e. sampling time interval, $t = 1, 2, \ldots$
k index for stage period, $k = 1, 2, 3, \ldots$
ζ small time shift related to a future time point at $\tau + \zeta$
n number of observation intervals per departure time interval
l number of departure time intervals per roll period
h prediction horizon in numbers of departure time intervals
q maximum lag length in numbers of departure time intervals, i.e. the traffic flow at the current departure time interval τ can include traffic demand departing from interval $\tau, \tau - 1, \tau - 2, \ldots, \tau - q$
$c_{(i,t)}$ number of vehicles measured on link i, during observation interval t
$D_{(j,\tau)}$ demand volume from origin–destination pair j during departure time interval τ
$LP_{(i,t),(j,\tau)}$ link proportions, that is the proportion of vehicles on link i at observation time t (coming from OD pair j at departure time τ) to the total demand of OD pair j at departure time τ
$D^r_{(j,\tau)}$ demand volume in regular demand pattern for origin–destination pair j during departure time interval τ
$\widetilde{D}^r_{(j,\tau)}$ a priori estimate of regular demand volume for origin–destination pair j during departure time interval τ
$\mu_{(j,\tau)}$ structural demand deviation of from a priori estimate $\widetilde{D}^r_{(j,\tau)}$ for OD pair j with departure time τ
$\varepsilon_{(j,\tau)}$ error term in approximating true demand for OD pair j with departure time τ
$\mu'_{(j,\tau)}, \mu''_{(j,\tau)}, \mu'''_{(j,\tau)}$ first, second and third-order derivatives of demand deviation $\mu_{(j,\tau)}$, respectively
p order index of a polynomial model
$\mu^{(p)}_{(j,\tau)}$ pth-order derivative of demand deviation $\mu_{(j,\tau)}$
m maximum order of a polynomial model
$w^{(p)}_{(j,\tau)}$ evolution noise for pth-order derivative of demand deviation $\mu_{(j,\tau)}$
$u_{(i,t)}$ combined error term in the estimation of link observation $c_{(i,t)}$ due to inconsistencies in assumptions about traffic assignment, traffic control and flow propagation, as well as measurement noise
$v_{(i,t)}$ combined error term due to $u_{(i,t)}$ and $\varepsilon_{(j,\tau)}$ for link observation $c_{(i,t)}$
$\hat{D}_{(j,\tau)}$ estimated mean value of $D_{(j,\tau)}$
$\hat{\mu}^{(p)}_{(j,\tau)}$ estimated mean value of $\mu^{(p)}_{(j,\tau)}$
Z_k state variable vector at stage k
Y_k measurement vector at stage k
H_k measurement matrix, relating measurement Y_k and state Z_k
w_k process noise at stage k
v_k measurement noise at stage k
$\hat{Z}_{k,k-1}$ prediction of Z_k using observations up to stage $k-1$, i.e. $E(Z_k | Y_1, Y_2, \ldots, Y_{k-1}.)$
$\hat{Z}_{k,k}$ estimation of Z_k using observations up to stage k, i.e. $E(Z_k | Y_1, Y_2, \ldots, Y_k.)$

$P_{k,k-1}$ predicted state covariance matrix of Z_k at stage $k-1$, i.e. $\mathrm{Var}(Z_k - \hat{Z}_{k,k-1})$
$P_{k,k}$ estimated state covariance matrix of $Z_k$at stage k, i.e. $\mathrm{Var}(Z_k - \hat{Z}_{k,k})$

3.1. Transition equation

The objective of the dynamic OD demand estimation and prediction problem is to find the time-dependent demand $D_{(j,\tau)}$ for origin–destination pair j at departure time interval τ. As discussed previously, the true demand $D_{(j,\tau)}$ can be partitioned into three components, namely, the regular pattern, structural deviations and random fluctuations. Theoretically, only the a priori estimate $\widetilde{D}^r_{(j,\tau)}$ of the regular demand, reflecting prior survey data and surveillance information up to the previous day, is available before performing real-time estimation on the current day. For this reason, the true demand $D_{(j,\tau)}$ in the following study is modeled as a linear combination of the a priori estimate, structural deviation and random disturbance:

$$D_{(j,\tau)} = \widetilde{D}^r_{(j,\tau)} + \mu_{(j,\tau)} + \varepsilon_{(j,\tau)}, \tag{1}$$

where the random disturbance term is assumed to follow a Normal distribution with zero mean. Moreover, a polynomial trend model is introduced to describe the structural deviations based on the following assumption.

Assumption 1 (*Polynomial trend*). Deviation at time $\tau + \zeta$ can be adequately represented *locally* by an mth-order polynomial function as Eq. (2) near time τ for a small value of ζ, while derivatives of higher orders are assumed to be zero: $\mu^{(p)}_{(j,\tau)} = 0$ for $p > m$

$$\mu_{(j,\tau+\zeta)} = b_0 + b_1\zeta + b_2\zeta^2 + \cdots + b_p\zeta^p + \cdots + b_m\zeta^m. \tag{2}$$

The above assumption implies that the derivatives of the polynomial model at time τ only provide an accurate approximation for short-term demand prediction at time $\tau + \zeta$. The trends of demand patterns could evolve dramatically when the length of the prediction horizon increases. As a result, the local trend estimates at time τ might not be adequate to describe and predict the actual long-term demand deviation pattern.

From Taylor's theorem, the smooth function of $\mu_{(j,\tau+\zeta)}$ can be expanded about the point $\mu_{(j,\tau)}$ as

$$\mu_{(j,\tau+\zeta)} = \mu_{(j,\tau)} + \zeta\mu'_{(j,\tau)} + \frac{\zeta^2}{2!}\mu''_{(j,\tau)} + \cdots + \frac{\zeta^p}{(p)!}\mu^{(p)}_{(j,\tau)} + \cdots + \frac{\zeta^m}{(m)!}\mu^{(m)}_{(j,\tau)}. \tag{3}$$

A comparison of Eqs. (2) and (3) indicates that the polynomial coefficients in the original functional form can be obtained directly from

$$b_p = \frac{\mu^{(p)}_{(j,\tau)}}{p!}. \tag{4}$$

A more compact form for the pth-order derivative of a polynomial can be generalized as

$$\mu^{(p)}_{(j,\tau+\zeta)} = \sum_{s=p}^{m} \frac{\zeta^{(s-p)}}{(s-p)!}\mu^{(s)}_{(j,\tau)}. \tag{5}$$

The corresponding matrix representation for a third-order polynomial model can be expressed as

$$\begin{pmatrix} \mu_{(j,\tau+\zeta)} \\ \mu'_{(j,\tau+\zeta)} \\ \mu''_{(j,\tau+\zeta)} \\ \mu'''_{(j,\tau+\zeta)} \end{pmatrix} = \begin{bmatrix} 1 & \zeta & \zeta^2/2! & \zeta^3/3! \\ & 1 & \zeta & \zeta^2/2! \\ & & 1 & \zeta \\ & & & 1 \end{bmatrix} \begin{pmatrix} \mu_{(j,\tau)} \\ \mu'_{(j,\tau)} \\ \mu''_{(j,\tau)} \\ \mu'''_{(j,\tau)} \end{pmatrix}. \tag{6}$$

The second assumption is required to allow time-varying trends to evolve stochastically between time stages.

Assumption 2 (*Evolution process*). From stage k to stage $k+1$, the change of derivative $\mu^{(p)}_{(j,\tau)}$ can be described as

$$\mu^{(p)}_{(j,\tau+l)} = \sum_{s=p}^{m} \frac{l^{(s-p)}}{(s-p)!}\mu^{(s)}_{(j,\tau)} + w^{(p)}_{(j,\tau)}, \tag{7}$$

where departure time index$\tau = kl$, and $w^{(p)}_{(j,\tau)} \sim N[0, W^{(p)}_{(j,\tau)}]$.

Taking a third-order polynomial trend model for OD pair j as an example, the corresponding transition equation in the Kalman filtering formulation is given in (8). More precisely, the state vector consists of the zeroth to mth-order derivatives of demand structural deviations from the a priori regular demand pattern estimate for OD pair j. Note that the transition matrix is independent of the current stage k and related departure time interval τ

$$\begin{pmatrix} \mu_{(j,\tau+l)} \\ \mu'_{(j,\tau+l)} \\ \mu''_{(j,\tau+l)} \\ \mu'''_{(j,\tau+l)} \end{pmatrix} = \begin{bmatrix} 1 & l & l^2/2! & l^3/3! \\ & 1 & l & l^2/2! \\ & & 1 & l \\ & & & 1 \end{bmatrix} \begin{pmatrix} \mu_{(j,\tau)} \\ \mu'_{(j,\tau)} \\ \mu''_{(j,\tau)} \\ \mu'''_{(j,\tau)} \end{pmatrix} + \begin{pmatrix} w_{(j,\tau)} \\ w'_{(j,\tau)} \\ w''_{(j,\tau)} \\ w'''_{(j,\tau)} \end{pmatrix}. \tag{8}$$

Consequently, the above single OD-pair model can be easily extended to consider all the OD pairs in a network. Consider a third-order polynomial filter with departure time $\tau = kl$ at stage k, we can define the state vector as

$$Z_k = (\mu_{(1,\tau)}, \mu'_{(1,\tau)}, \mu''_{(1,\tau)}, \mu'''_{(1,\tau)}, \mu_{(2,\tau)}, \mu'_{(2,\tau)}, \mu''_{(2,\tau)}, \mu'''_{(2,\tau)}, \ldots, \mu_{(N_{\mathrm{od}},\tau)}, \mu'_{(N_{\mathrm{od}},\tau)}, \mu''_{(N_{\mathrm{od}},\tau)}, \mu'''_{(N_{\mathrm{od}},\tau)})^{\mathrm{T}} \tag{9}$$

and the transition matrix as

$$A_k = \mathrm{Diag}(A_k^1, A_k^2, \ldots, A_k^j, \ldots, A_k^{N_{\mathrm{od}}}), \tag{10}$$

$$\text{where} \quad A_k^j = \begin{bmatrix} 1 & l & l^2/2! & l^3/3! \\ & 1 & l & l^2/2! \\ & & 1 & l \\ & & & 1 \end{bmatrix} \tag{11}$$

for $j = 1, 2, \ldots, N_{\mathrm{od}}$.

By assuming the evolution noise w_k as

$$w_k = (w_{(1,\tau)}, w'_{(1,\tau)}, w''_{(1,\tau)}, w'''_{(1,\tau)}, w_{(2,\tau)}, w'_{(2,\tau)}, w''_{(2,\tau)}, w'''_{(2,\tau)}, \ldots, w_{(N_{\mathrm{od}},\tau)}, w'_{(N_{\mathrm{od}},\tau)}, w''_{(N_{\mathrm{od}},\tau)}, w'''_{(N_{\mathrm{od}},\tau)})^{\mathrm{T}}. \tag{12}$$

The complete transition equation in real-time OD estimation and prediction can be written as

$$Z_{k+1} = A_k Z_k + w_k. \tag{13}$$

To obtain the future demand level with prediction horizon h, we need to first predict the demand deviation at time $\tau + h$ based on estimated derivatives at the current stage, and then substitute the predicted demand deviation and the a priori estimate of the regular demand pattern into Eq. (1). Thus,

$$E[D_{(j,\tau+h)} | \mu_{(j,\tau)}] = \widetilde{D}^r_{(j,\tau+h)} + E[\mu_{(j,\tau+h)} | \mu_{(j,\tau)}] = \widetilde{D}^r_{(j,\tau+h)} + \sum_{s=0}^{m} \frac{h^s}{s!} \hat{\mu}^{(s)}_{(j,\tau)}, \tag{14}$$

$$\mathrm{Var}[D_{(j,\tau+h)} | \mu_{(j,\tau)}] = \sum_{s=0}^{m} \sum_{s'=0}^{m} \left[\frac{h^{(s)}}{s!} \frac{h^{(s')}}{s'!} \mathrm{Cov}\left(\mu^{(s)}_{(j,\tau)}, \mu^{(s')}_{(j,\tau)} \right) \right], \tag{15}$$

where $\tau = kl$.

Eq. (15) indicates that prediction errors of the proposed model are dependent on the order of model m and the length of prediction horizon h. Essentially, a complicated high-order polynomial model describes the dynamic demand process in greater detail than a low-order model, but it might lead to potential large prediction errors.

By incorporating the a priori estimate of the regular demand pattern in the proposed structural model, one computational advantage is the reduction in the dimension of the state variable vector. To show this, suppose the original demand can be adequately fitted by an sth-order polynomial model as

$$D_{(j,\tau+\zeta)} = a_0 + a_1\zeta + a_2\zeta^2 + \cdots + a_m\zeta^m + \cdots + a_s\zeta^s + \varepsilon_{(j,\tau+\zeta)}. \tag{16}$$

If $\widetilde{D}^r_{(j,\tau+\zeta)}$ is a good approximation to $D_{(j,\tau+\zeta)}$, it can be further assumed that $\widetilde{D}^r_{(j,\tau+\zeta)}$also corresponds to an sth-order polynomial model satisfying $(a_p - \tilde{a}^r_p) = 0$ for $p > m$, as

$$\widetilde{D}^r_{(j,\tau+\zeta)} = \tilde{a}^r_0 + \tilde{a}^r_1\zeta + \tilde{a}^r_2\zeta^2 + \cdots + \tilde{a}^r_m\zeta^m + \cdots + \tilde{a}^r_s\zeta^s. \tag{17}$$

Then, ignoring the higher-order terms from m+1 to s, we have

$$\begin{aligned}\mu_{(j,\tau+\zeta)} &= D_{(j,\tau+\zeta)} - \widetilde{D}^r_{(j,\tau+\zeta)} - \varepsilon_{(j,\tau+\zeta)}\\ &= (a_0 - \tilde{a}^r_0) + (a_1 - \tilde{a}^r_1)\zeta + \cdots + (a_m - \tilde{a}^r_m)\zeta^m + \cdots + (a_s - \tilde{a}^r_s)\zeta^s - \varepsilon_{(j,\tau+\zeta)}\\ &= (a_0 - \tilde{a}^r_0) + (a_1 - \tilde{a}^r_1)\zeta + \cdots + (a_m - \tilde{a}^r_m)\zeta^m - \varepsilon_{(j,\tau+\zeta)}.\end{aligned} \tag{18}$$

and the resulting order of the polynomial model will be reduced from s to m. Since the computational complexity of the Kalman filter is on the order of $O(N^3)$, where $N = N_{\rm od} \times m$ in our case, the reduction of the model order from s to m dramatically decreases the size of the state vector, and therefore improves computational time efficiency. On the other hand, if the real-time pattern is drastically different from the historical pattern, then a high-order polynomial model may be still needed to approximate the demand deviation. As an extreme example, suppose the historical pattern is unavailable, i.e. $\widetilde{D}^r_{(j,\tau+\zeta)} = 0$, then a full sth-order polynomial model has to be used to approximate the live traffic. In this case, the model reduction and the resulting computational saving are not possible.

In addition, incorporating a reliable estimate of the regular demand pattern is always beneficial for improving estimation and prediction quality. From the linear regression standpoint, the regular daily pattern can be viewed as a good explanatory regressor that absorbs a considerable amount of variation in the independent variable (i.e. true dynamic demand). Thus, compared to a pure polynomial model, the proposed structural model with the regular pattern component leads to smaller regression residual errors, that is, smaller estimation and prediction errors. Morrison (1969) and Brookner (1998) provide in-depth discussions on the trade-offs associated with incorporating the regular pattern component in a polynomial trend filter.

3.2. Measurement equations

In general, the measurement equation connects the link observations and OD demands through a link proportion matrix, as in Eq. (19). Specifically, the link proportions map all the lagged and prevailing demands at the current stage k to $n * l$ measurements for each link with available observations

$$c_{(i,t)} = \sum_{\zeta=-q}^{l-1}\sum_{j=1}^{N_{\rm od}}\left(LP_{(i,t),(j,\tau+\zeta)} \times D_{(j,\tau+\zeta)}\right) + u_{(i,t)}, \tag{19}$$

where $\tau = kl$, $i = 1, 2, \ldots, N_{\rm obs}$, and $knl \leqslant t \leqslant (k+1)nl - 1$.

To relate link measurements to the state variables constructed previously, substitute Eqs. (1) and (3), the above equation becomes

$$c_{(i,t)} = \sum_{\zeta=-q}^{l-1}\sum_{j=1}^{N_{\rm od}}\left[LP_{(i,t),(j,\tau+\zeta)} \times \left(\sum_{s=0}^{m}\frac{\zeta^s}{s!}\mu^{(s)}_{(j,\tau)} + \widetilde{D}^r_{(j,\tau+\zeta)} + \varepsilon_{(j,\tau+\zeta)}\right)\right] + u_{(i,t)}, \tag{20}$$

and it can be further transformed to

$$\begin{aligned}&c_{(i,t)} - \sum_{\zeta=-q}^{l-1}\sum_{j=1}^{N_{\rm od}}\left(LP_{(i,t),(j,\tau+\zeta)} \times \widetilde{D}^r_{(j,\tau+\zeta)}\right)\\ &\quad = \sum_{\zeta=-q}^{l-1}\sum_{j=1}^{N_{\rm od}}\left(LP_{(i,t),(j,\tau+\zeta)} \times \sum_{s=0}^{m}\frac{\zeta^s}{s!}\mu^{(s)}_{(j,\tau)}\right) + \sum_{\zeta=-q}^{l-1}\sum_{j=1}^{N_{\rm od}}\left(LP_{(i,t),(j,\tau+\zeta)} \times \varepsilon_{(i,t+\zeta)}\right) + u_{(i,t)}.\end{aligned} \tag{21}$$

Consequently, we can define the observation vector and measurement error in the Kalman formulation as follows:

$$Y_k = H_k Z_k + v_k, \tag{22}$$

$$Y_k = \left(y_{(1,knl)}, y_{(1,knl+1)}, \ldots, y_{(1,(k+1)nl-1)}, \ldots, y_{(N_{\mathrm{obs}},knl)}, y_{(N_{\mathrm{obs}},knl+1)}, \ldots, y_{(N_{\mathrm{obs}},(k+1)nl-1)}\right)^{\mathrm{T}}, \tag{23}$$

$$\text{where} \quad y_{(i,t)} = c_{(i,t)} - \sum_{\zeta=-q}^{l-1} \sum_{j=1}^{N_{\mathrm{od}}} \left(LP_{(i,t),(j,\tau+\zeta)} \times \widetilde{D}^r_{(j,\tau+\zeta)} \right), \tag{24}$$

$$v_k = \left(v_{(1,knl)}, v_{(1,knl+1)}, \ldots, v_{(1,(k+1)nl-1)}, \ldots, v_{(N_{\mathrm{obs}},knl)}, v_{(N_{\mathrm{obs}},knl+1)}, \ldots, v_{(N_{\mathrm{obs}},(k+1)nl-1)}\right)^{\mathrm{T}}, \tag{25}$$

$$\text{where} \quad v_{(i,t)} = \sum_{\zeta=-q}^{l-1} \sum_{j=1}^{N_{\mathrm{od}}} \left(LP_{(i,t),(j,\tau+\zeta)} \times \varepsilon_{(j,\tau+\zeta)} \right) + u_{(i,t)}.$$

The final measurement error term in the transition equation combines the random noise in OD demand, other errors associated with link proportions, as well as the sensor errors in traffic measurements.

The dimension of measurement matrix H_k is $(N_{\mathrm{obs}} \times nl, N_{\mathrm{od}} \times (m+1))$, and its (i,t)th, $(j,(p+1))$th element is

$$H_{(i,t),(j,p)} = \sum_{\zeta=-q}^{l-1} \left(LP_{(i,t),(j,\tau+\zeta)} \times \frac{\zeta^p}{p!} \right), \tag{27}$$

where $\tau = kl$.

By applying a polynomial approximation for OD demands during departure time intervals from $kl - q$ to $(k+1)l - 1$, the polynomial trend filter neatly incorporates the lagged demands into the estimation procedure for the current stage, leading to an efficient state space representation, desirable for large-scale network applications.

Assumption 3. w_k and v_k are white noise terms uncorrelated with the initial state Z_0 and with each other, where $w_k \sim N[0, W_k]$ and $v_k \sim N(0, V_k)$.

From Assumption 3, the Kalman filtering algorithm is ready to be integrated into the following recursive estimation and prediction algorithm.

Algorithm 2. Real-time dynamic demand OD estimation and prediction

Step 1: (Initialization) Set up initial estimates $P_{0,0} = \mathrm{Var}(Z_0)$ and $\hat{Z}_{0,0} = E(Z_0)$. Let $k = 1$.

Step 2: (Prediction) Propagate the mean and covariance estimates from $k-1$ to k.

$$\hat{Z}_{k,k-1} = A_k \hat{Z}_{k-1,k-1}, \tag{28}$$

$$P_{k,k-1} = A_k P_{k-1,k-1} A_k^{\mathrm{T}} + W_k. \tag{29}$$

Step 3: (Estimation of state variable) After receiving new link proportions and link observations, calculate the weighting matrix as

$$K_k = P_{k,k-1} H_k^{\mathrm{T}} (H_k P_{k,k-1} H_k^{\mathrm{T}} + V_k)^{-1} \tag{30}$$

and then update the a posteriori mean and covariance estimates

$$\hat{Z}_{k,k} = \hat{Z}_{k,k-1} + K_k (Y_k - H_k \hat{Z}_{k,k-1}), \tag{31}$$

$$P_{k,k} = (I - K_k H_k) P_{k,k-1}. \tag{32}$$

Step 4: (Estimation of real-time demand) Calculate the estimation of real-time demand using new estimates $\hat{\mu}_{(j,\tau)}$

$$\hat{D}_{(j,\tau)} = E(\widetilde{D}^r_{(j,\tau)} + \mu_{(j,\tau)} + \varepsilon_{(j,\tau)}) = \widetilde{D}^r_{(j,\tau)} + \hat{\mu}_{(j,\tau)}, \tag{33}$$

where $\tau = kl, kl+1, \ldots, (k+1)l - 1$.

Step 5: Advance roll period forward from k to $k+1$, and then go back to Step 2.

Further, if independence of measurement errors is assumed, that is,

$$v_k \sim N(0, \mathrm{diag}[V_{(1,knl)}, V_{(1,knl+1)}, \ldots, V_{(N_{\mathrm{obs}},(k+1)nl-1)}]), \tag{34}$$

we can apply the scalar updating scheme (see Chui and Chen, 1991; Ashok, 1996) in order to avoid complicated matrix inversion in a real-time setting.

In the context of short-term economic forecasting, the zeroth, first and second-order polynomial models can be viewed as the local level model, local linear trend model and local quadratic growth model, respectively. Regarding connections between the polynomial trend models and other time-series ARIMA models, West and Harrison (1997) demonstrated that, if restrictions are imposed on the autocorrelation structure, the limiting case of an $(m+1)$th polynomial trend model is equivalent to an ARIMA $(0,m,m)$ model. According to the generalized state space architecture proposed by Harvey (1989), an auto-regression (AR) term can be also incorporated into the state variable vector to model the autocorrelation structure in the random disturbance. It is worth remarking that, even if the underlying trends for demand structural deviations are negligible, it is advisable to embed a polynomial trend component in the space state representation so as to monitor and identify possible changes in the process structure.

We use the following example to highlight the importance of combining both the polynomial trend model and historical demand information into the prediction of real-time traffic demand flows. In Fig. 3a and b, solid lines represent true demand in morning peak hours; dotted lines represent available historical demand data that describe a regular demand pattern. Dashed lines in Fig. 3c–j correspond to predicted demand flows from various state space models in two different situations. The plots on the left hand side represent a case in which the realized true demand is quite similar to the regular demand pattern. The plots on the right hand side represent a case in which the actual demand pattern is significantly lower than the regular demand level as a result of severe weather conditions.

Fig. 3c–f illustrate the performance of the local level and linear trend models under both small and large deviations between the actual and regular demand patterns. Compared to the local level model, the linear trend is more suitable to describe upward and downward patterns in the peak and off-peak periods, respectively. However, the linear trend model dramatically overestimates the OD demand during the peak time point, because all real-time measurements before the turning point indicate an upward tendency.

Fig. 3g–j show the benefit of incorporating the historical demand information component under both regular and irregular conditions. Under the regular condition, the "historical demand + autoregressive error" model in Fig. 3g successfully detrends the real-time demand time series and produces accurate prediction results. In the presence of structural changes, however, deviations between the actual and regular demand have non-zero means, leading to systematic biases for models that rely on the stationarity assumption. In contrast, the "historical demand + linear trend" model utilizes the polynomial filter to absorb possible structural difference, producing robust predicts under both regular and irregular conditions. Comparing the combined "historical demand + linear trend" model with the pure "linear trend" model, one can also verify that the inclusion of the historical demand information component considerably decreases the slope of the linear trend model and the resulting prediction errors. Thus, a robust demand predictor should incorporate both regular pattern and trend components to ensure prediction quality under a wide range of conditions. Even though the underlying trend of the structural deviations might be negligible in some cases, it is still advisable to embed a simple trend component in the space state representation so that possible future changes in the process structure can be systematically monitored and identified.

4. Adaptive day-to-day updating of regular demand pattern information

As discussed earlier, the initial estimate for the regular demand pattern could be unreliable due to limited sample size, and the normal daily pattern could evolve smoothly due to day-to-day demand dynamics. Hence, it is necessary to update the a priori estimate using the new demand estimate and new observations. A desirable updating formulation should be able to adaptively recognize and capture the systematic day-to-day evolution, and also maintain robustness under disruptions due to special events. An updating formulation based on a Kalman filter framework is proposed.

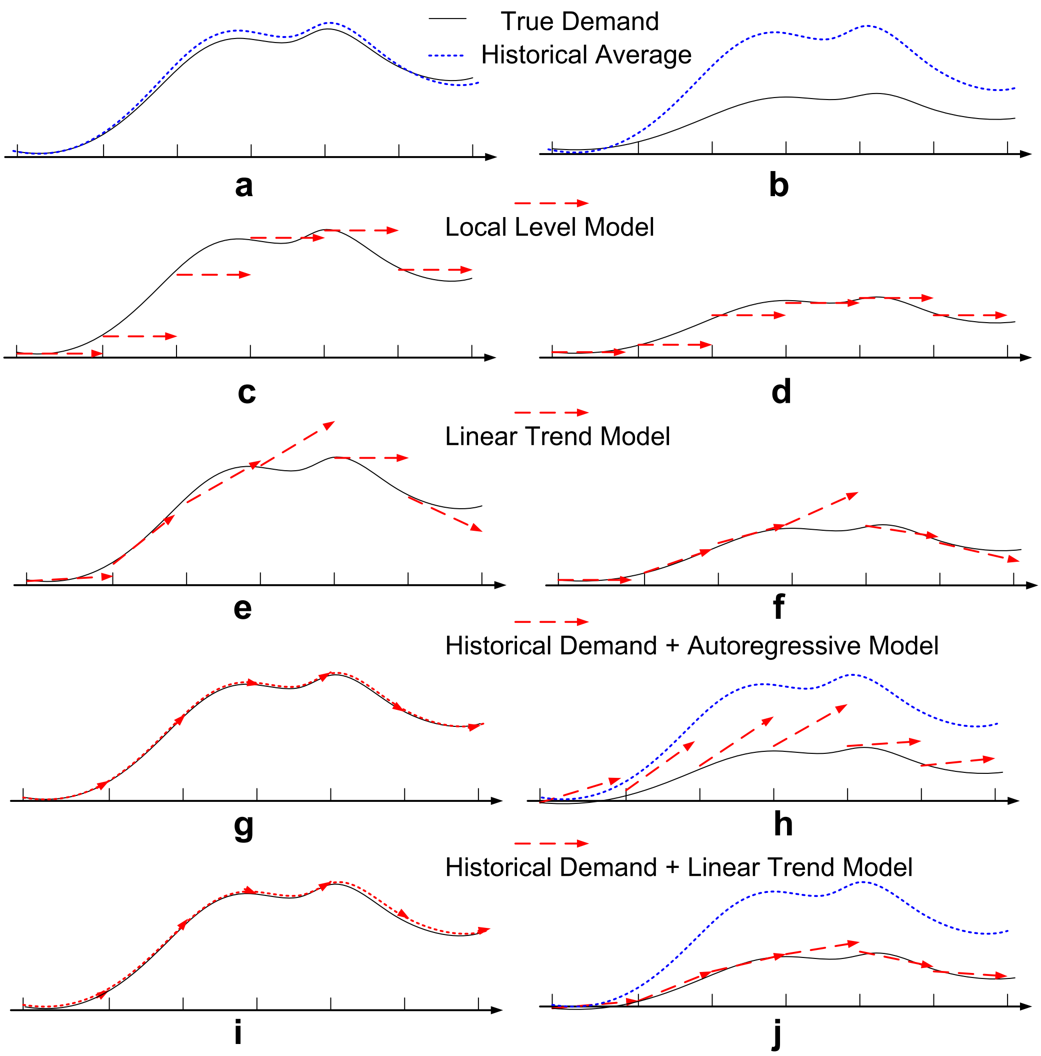

Fig. 3. Illustrative examples of different structural state models.

The notation used in the real-time OD estimation and prediction model is extended to the day-to-day context as follows:

d	index for day
D^r_d	state variable vector of regular OD demand pattern on day d, consisting of elements $D^r_{(j,\tau)}$
ξ_d	day-to-day evolution variance on day d
$\hat{D}_d$	vector of the real-time demand estimate on day d, consisting of estimates $\hat{D}_{(j,\tau)}$
η_d	measurement variance matrix on day d
$\hat{D}^r_{d,d-1}$	predicted state variable vector D^r_d using observations up to day $d-1$, consisting of elements $\widetilde{D}^r_{(j,\tau)}$
$\hat{D}^r_{d,d}$	estimated state variable vector D^r_d using observations up to day d
$\Sigma_{d,d-1}$	predicted state covariance matrix for the regular demand pattern on day d
$\Sigma_{d,d}$	estimated state covariance matrix for the regular demand pattern on day d
K_d	Kalman gain matrix for using real-time demand estimates on day d
$\hat{M}_d$	vector of estimated demand deviations on day d, with elements $\hat{\mu}_{(j,\tau)}$
K'_d	Kalman gain matrix for using real-time observations on day d
C_d	vector of traffic observations on day d, consisting of elements $c_{(i,t)}$
LP_d	link proportion matrix on day d, consisting of elements $LP_{(i,t),(j,\tau)}$

The transition and measurement equations for the day-to-day demand evolution can be written as

Transition equation:

$$D^r_{d+1} = D^r_d + \xi_d. \tag{35}$$

Measurement equation:

$$\hat{D}_d = D^r_d + \eta_d. \tag{36}$$

Assumption 4. ξ_d and η_d are white noise terms uncorrelated with the initial state $D_0{}^r$ and with each other, where $\xi_d \sim N(0, Q_d)$ and $\eta_d \sim N(0, R_d)$.

According to transition Eq. (35), the regular demand pattern can evolve smoothly from day-to-day, where stochastic day-to-day evolution is captured by the evolution random term ξ_d with zero mean. In the measurement equation, since the true demand state cannot be directly observed, the new real-time demand estimate $\hat{D}_d$ is considered as "measurement" incoming everyday. Following the standard Kalman filtering algorithm, the updating procedure can be summarized as follows.

Algorithm 3. Day-to-day updating for regular demand pattern estimate

Step 1: (Initialization) Set up $\hat{D}^r_{0,0}$ and $\Sigma_{0,0}$ as the initial estimated mean and covariance of the regular demand. Let $d = 1$.

Step 2: (Computation of a priori estimate) The *a posteriori* estimate $\hat{D}^r_{d-1,d-1}$ on previous day $d-1$ is used as the a priori estimate for current day d. The corresponding covariance matrix is updated by taking evolution noise into account

$$\hat{D}^r_{d,d-1} = \hat{D}^r_{d-1,d-1}, \tag{37}$$

$$\Sigma_{d,d-1} = \Sigma_{d-1,d-1} + Q_d. \tag{38}$$

Step 3: (Real-time OD estimation and prediction) Run the real-time OD estimation and prediction module in conjunction with real-time DTA simulators, to obtain new estimates $\hat{M}_d$ and $\hat{D}_d$ for day d.

Step 4: (Update of gain matrix) Compute the gain matrix using predicted state covariance matrix and measurement variance matrix

$$K_d = \Sigma_{d,d-1}(\Sigma_{d,d-1} + R_d)^{-1}. \tag{39}$$

Step 5: (Update of mean and covariance) Update the estimated mean and covariance matrix for the regular demand state vector

$$\hat{D}^r_{d,d} = \hat{D}^r_{d,d-1} + K_d(\hat{D}_d - \hat{D}^r_{d,d-1}) = \hat{D}^r_{d,d-1} + K_d\hat{M}_d, \tag{40}$$

$$\Sigma_{d,d} = (I - K_d)\Sigma_{d,d-1}. \tag{41}$$

Step 6: Move to the next day, $d = d + 1$, and then go back to Step 1.

An important point is that $\hat{D}_d - \hat{D}^r_{d,d-1}$ in the above day-to-day updating algorithm is equivalent to $\hat{M}_d$, which consists of the demand deviation estimate $\hat{\mu}_{(j,\tau)}$ generated from the real-time estimation and prediction algorithm on day d. It can be shown that $\Sigma_{d,d} \leqslant \Sigma_{d,d-1}$, that is, the conditional demand estimate contains less uncertainty than the corresponding a priori estimate for the regular demand pattern, after incorporating additional information from the new real-time estimation result. More importantly, the above recursive updating algorithm naturally integrates with the proposed structural model in the previous section, and it is able to accumulate the information from the real-time estimator on a daily basis. Compared to the simple moving algorithm method, which applies the same weight factor for all the OD pairs, the proposed optimal updating process can adaptively determine different factors for different OD pairs. Assuming the measurement variance and evolution variance of OD pairs are independent in Eq. (39), we can calculate the updating factor for each individual OD pair based on the corresponding magnitude of the previous estimation variance and the current measurement variance. For example, for an OD pair with high a priori variance, higher weight should be given to the new demand estimate based on traffic measurements to speed up the learning process, and vice versa.

In order to make this recursive algorithm operational, the next question is how to specify the values of evolution variance Q_d and measurement varianceR_d. By using the multi-day OD estimation method proposed by Zhou et al. (2003), the variance of the measurement noise in Eq. (39) can be obtained by evaluating the variance of estimated OD demands across several days. Adaptive Kalman filtering (see Chui and Chen, 1991) is another approach to identify measurement varianceR_d, which utilizes statistical properties of the measurement residual sequence to dynamically estimate the measurement variance. Determining the day-to-day variance is generally more difficult, since we cannot directly observe the day-to-day demand evolution process. Recognizing that the proposed day-to-day evolution process can be described as a random walk plus noise model, existing time series techniques are applied to choose appropriate values of process variance. A common approach is to first assume a constant signal to noise ratio $\lambda = \frac{Q_d}{R_d}$, indicating the ratio of inherent system variance with respect to observational variance. Based on calibrated R_d, $Q_d = \lambda R_d$, so we can select an appropriate signal to noise ratio so as to minimize average prediction errors in the training data sets. The reader is referred to West and Harrison (1997) for a comprehensive treatment. In early stages of applying this updating mechanism, considerable uncertainty in the predicted state covariance $\Sigma_{d,d-1}$ results in a high gain factor, implying that the new real-time estimates receive relatively large weighting. After a certain number of iterations, the gain factor becomes stable as it is gradually reaching a steady state. If constant Q and R are assumed, the following limiting behavior for the Kalman gain matrix can be derived (West and Harrison, 1997):

$$\lim_{d\to\infty} K_d = \frac{\lambda}{2}\left(\sqrt{1+4/\lambda} - 1\right). \tag{42}$$

A typical value of λ can be 0.05, leading to $\lim_{d\to\infty} K_d = 0.2$, so the most recent real-time estimate receives relatively small weighting eventually. If $\lambda = 0.5$, corresponding to limiting gain factor of 0.5, then the a priori estimate and the new real-time estimate share equal importance in determining priori demand information for the next day.

Another approach to update the historical demand estimate is to directly utilize link observations instead of real-time demand estimates. The modified transition and measurement equations are shown below.

Transition equation:

$$D^r_{d+1} = D^r_d + \xi_d. \tag{43}$$

Measurement equation:

$$C_d = LP_d D^r_d + \eta'_d. \tag{44}$$

Assumption 5. ξ_d and η'_d are white noise terms uncorrelated with the initial state D^r_0 and with each other, where $\xi_d \sim N(0, Q_d)$ and $\eta'_d \sim N(0, R'_d)$.

Note that the above transition equation is identical to the one using real-time estimates. As the measurement equation uses link proportions to link real-world measurements and the regular demand pattern, the measurement variance R'_d should be recalibrated accordingly. The gain matrix, mean and covariance updating formulation in Steps 4 and 5 of Algorithm 3 should also be changed to the following:

$$K'_d = \Sigma_{d,d-1} LP^{\mathrm{T}}_d (LP_d \Sigma_{d,d-1} LP^{\mathrm{T}}_d + R'_d)^{-1}, \tag{45}$$

$$\hat{D}^r_{d,d} = \hat{D}^r_{d,d-1} + K'_d (C_d - LP_d \hat{D}^r_{d,d-1}), \tag{46}$$

$$\Sigma_{d,d} = (I - K'_d)\Sigma_{d,d-1}. \tag{47}$$

This Kalman filtering formulation provides a least-squares unbiased estimator for the regular demand pattern, with the optimal weights on the a priori estimate and new information. It is important to note that this recursive prediction–correction algorithm only requires a priori mean and covariance statistics at each iteration instead of the entire historical data series, resulting in efficient storage implementation for on-line applications. Practically, this updating method can be viewed as a moving average method with adaptive weights, depending on the respective reliability of the a priori and real-time information sources.

5. Application to Irvine test bed data set

Numerical experiments are presented to illustrate the application of the proposed model and algorithms to the Irvine, CA test bed traffic network, which consists of three freeway corridors (I-5, I-405, Highway 133) and other main arterials. As shown in Fig. 4, this network includes 61 OD zones, 326 nodes and 626 links, where traffic counts are measured at 30-second intervals on 19 freeway links and at 5-min interval on 28 arterial links during three days selected from May and June, 2001. In addition, the a priori estimate of the regular demand pattern is constructed by the off-line OD estimation method (Tavana and Mahmassani, 2001; Zhou et al., 2003) using the first day data. Real-world observations on the second day are used to calibrate the system and measurement variances in the real-time OD estimation and prediction model. The third day data are used to validate the proposed real-time OD estimation and prediction algorithm. The time of interest in the following experiments is the morning peak period (4:00 AM–10:00 AM), while the demand departure time interval and roll period are 5 and 15 min, respectively.

First, a first-order polynomial trend model (i.e. local linear trend model) is applied to estimate the demand deviations from the a priori estimate of the regular demand pattern. In Fig. 6, the a priori regular pattern estimate and the corresponding demand deviations are displayed for the OD pair from zone 53 to zone 40, which carries the highest trips for all among all the OD pairs in the study network. On average, the a priori demand data underestimates the real-time demand on the third day for this selected OD pair, but the prior information still shows similar time-varying dynamic patterns. As expected, the demand deviations exhibit a much slower changing pattern than the corresponding real-time demands over the same time. Essentially, the estimated structural demand deviations are caused by the day-to-day dynamics, but the deviations shown in this case can be also due to the estimation noise in the a priori demand data, which only utilizes one-day observations.

Considering the smooth trend for demand deviations in Fig. 5, it is desirable to further reduce the first-order polynomial model to the zeroth-order polynomial model. To assess and compare the estimation performance of alternative models, we define root mean square (RMS) error in density as

$$\mathrm{RMS}_t = \sqrt{\frac{\sum_i (c_{i,t} - \hat{c}_{i,t})}{N_{\mathrm{obs}}}} \tag{48}$$

where $c_{(i,t)}$ is the density measured on link i, during observation interval t, $\hat{c}_{(i,t)}$ is the simulated density from the real-time DTA estimator on link i, during observation interval t.

The RSE errors at every 5 min are plotted in Fig. 6, for the zeroth and first-order polynomial models, respectively. The average RMS error of the zeroth-order model during the study horizon (10.2608) is marginally greater than the average RMS error of the first-order model (10.8588) by 1.8%. Basically, these two

Fig. 4. Irvine test bed network.

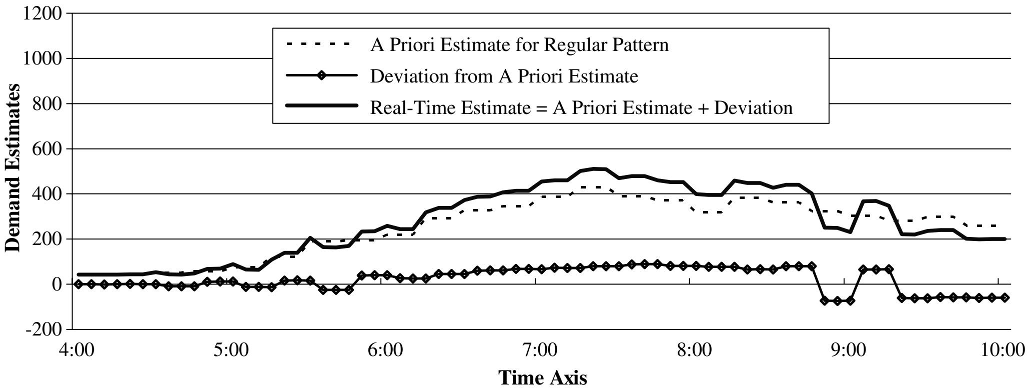

Fig. 5. Dynamic demand estimates for OD pair from zone 53 to zone 40.

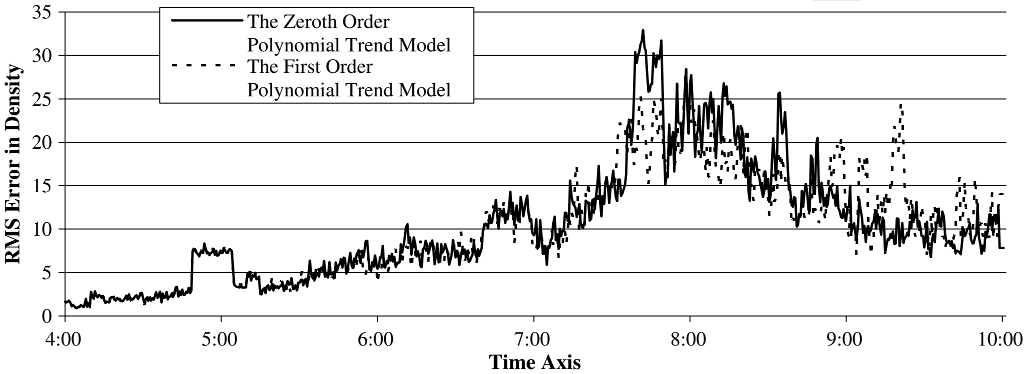

Fig. 6. RMS error in density for the zeroth and first-order polynomial models.

models perform better in the early morning (from 4:00 AM to 6:00 AM), compared to the peak hour period (from 7:00 AM to 9:00 AM). Such time-dependent RMS measures can be explained by the increasing variability in the peak hour demands and the high dynamics in the related traffic flow propagation processes. Based on experiment results from the third day data, the zeroth order polynomial model seems to be more attractive than the first-order model, since it offers acceptable accuracy with considerable enhanced computation efficiency. On the other hand, if real-time response constraints can be satisfied, keeping a high-order polynomial model is always preferable, because it can capture non-linear patterns in the demand structure changes.

Fig. 7 plots simulated density, predicted density and the observed density on link 212, using 20-min prediction horizon. Specifically, the simulated density and the predicted density are generated from the DTA

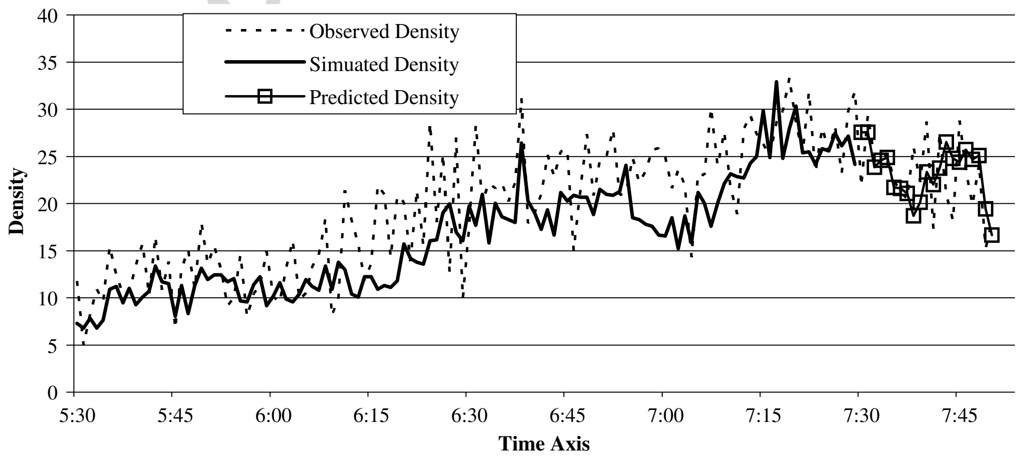

Fig. 7. Observed density vs. simulated density and predicted density on link 212.

network state estimation module and the DTA state prediction module, respectively. Link 212 is a freeway link going northbound, and its location is marked in Fig. 4. The DTA network state estimator is able to capture the time-varying trends of real-world traffic, while the DTA network state predictor can forecast dynamic flow propagation with acceptable quality. The above results further validate the effectiveness of the proposed real-time OD estimation and prediction framework, and illustrate the role of the real-time traffic measurement database in enabling and improving this process.

Table 1 compares network-wide simulation results and real-world observations in terms of the root mean squared errors of density, volume and speed for observed links. In particular, link density and speed measures are processed at 1-min time intervals, while link volume is processed based on 15-min time intervals to obtain reliable samples. The consistent error reduction in these three major traffic measures again demonstrates that the on-line network state estimator is able to utilize real-time measurements to improve the final quality of traffic state estimation.

In the following experiments, the dynamic demand matrix calibrated from the observations in 2001 is used as an initial historical demand pattern. We apply the proposed day-to-day updating framework to utilize traffic measurements obtained from the PeMS database in 2005. PeMS is a general Performance Measurement System, and initially applied to California freeways (Chen et al., 2001; Varaiya, 2002). Fig. 8 clearly shows a "day of week" pattern in the Orange County in terms of vehicle miles traveled. As a result, we focus on traffic measurements from 12 continuous Monday (namely March 07, 2005 to May 23, 2005), instead of continuous weekdays, to demonstrate the proposed day-to-day updating procedure. It should be remarked that, the year 2001 data cover both arterials and freeways, but the year 2005 data from PeMS only provide freeway traffic measurements.

Fig. 9 indicates that, the optimal updating mechanism produces significantly lower estimation errors than the simple moving average method in terms of RMSE of link density in the first 6 weeks, while the simple moving average method uses a constant gain factor of 0.2 for all OD pairs. The error reduction can be attributed to the fact that the optimal updating mechanism can quickly provide an up-to-date estimate of regular demand pattern with the least variance. After 7 weeks, both simple updating and optimal updating models

Table 1
Average RMSE in on-line estimation vs. a priori estimation

	A priori estimation	On-line estimation	Percentage improvement
Density	11.6	10.5	9.5
Volume	288.8	208.5	27.8
Speed	16.7	14.1	25.6

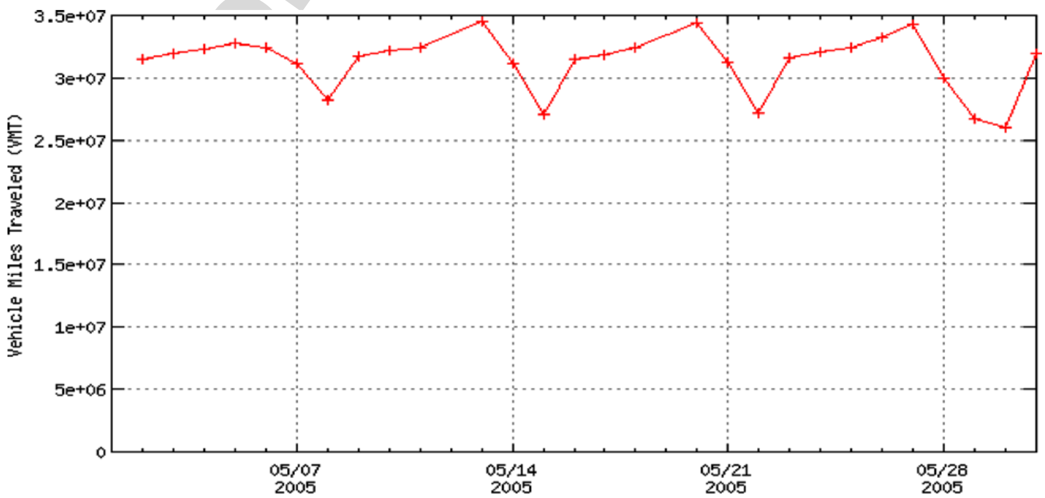

Fig. 8. Vehicle miles traveled in Orange County in May, 2005 (from PeMS).

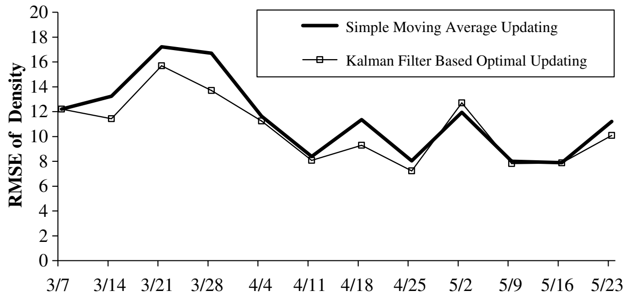

Fig. 9. Estimation quality under different updating frameworks.

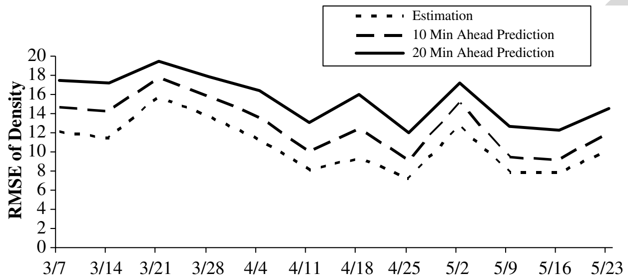

Fig. 10. Prediction performance under Kalman filter based optimal updating model.

demonstrate similar estimation performance, suggesting that the simple updating model is also able to converge to the up-to-date regular demand pattern but with more updating iterations. Fig. 10 shows the prediction precision for different prediction horizons under the Kalman Filter based optimal updating model. As expected, a longer prediction horizon leads to larger prediction errors. Additionally, the magnitude of prediction errors at different days highly depends on the corresponding estimation quality, which highlights the need for high-quality real-time OD demand estimators.

6. Concluding comments

Real-time OD estimation and prediction is an important component in real-time dynamic traffic assignment for ATMS/ATIS network applications. This paper exploits the potential of using a structural state space model to systematically incorporate regular demand pattern information, structural deviations and random fluctuations. The contributions of this study include the following aspects. First, a polynomial trend filter is developed to estimate and predict demand deviations from the a priori estimate of the regular demand pattern, so as to utilize valuable historical information and adaptively response to possible structural deviations in demands. Second, based on a Kalman filtering framework, an optimal recursive procedure is proposed for updating the regular demand pattern estimate with new real-time estimates and observations obtained every day. These models can be naturally integrated into the real-time DTA system and provide an effective and efficient approach to utilize the real-time traffic data continuously in the operational settings. Third, the application to the Irvine network provides an illustration of the usefulness of the real-time traffic sensor data to calibrate and test advanced network modeling tools intended for ITS planning and operation.

One particularly attractive opportunity would be to integrate our dynamic network modeling capability explicitly with real-time traffic sensor systems into the decision support system. The day-to-day updating framework would then be used to regularly update the modeling tools, keeping an up-to-the-minute version ready for application to analyze scenarios and contemplated actions for unfolding conditions. An important capability in this process would be the identification, through the extensive traffic sensor data warehouse, of most comparable days or patterns in order to provide a starting point, to be updated using the methods presented in this work.

Acknowledgements

This paper is based in part on research funded through the US Department of Transportation. The work has benefited from the contributions of several current and former graduate students, including Xiao Qin, Robert Mahfoud, Hossein Tavana, Ying Kang, Khaled Abdelghany and Hayssam Sbayti. The final version of the paper has benefited from the comments of two anonymous referees. The authors are of course responsible for all results and opinions expressed in this paper.

References

Ashok, K., 1996. Estimation and prediction of time-dependent origin–destination Flows. Ph.D. Dissertation, MIT, USA.

Ashok, K., Ben-Akiva, M., 1993. Dynamic origin–destination matrix estimation and prediction for real-time traffic management systems. In: Daganzo, C. (Ed.), Proceedings of the 12th International Symposium on Transportation. Elsevier.

Ashok, K., Ben-Akiva, M., 2000. Alternative approaches for real-time estimation and prediction of time-dependent origin–destination flows. Transportation Science 34 (1), 21–36.

Brookner, E., 1998. Tracking and Kalman Filtering Made Easy. John Wiley and Sons, Inc., New York.

Cascetta, E., Inaudi, D., Marquis, G., 1993. Dynamic estimators of origin–destination matrices using traffic counts. Transportation Science 27 (4), 363–373.

Chang, G.L., Tao, X.D., 1999. An integrated model for estimating time-varying network origin–destination distributions. Transportation Research Part A 33 (5), 381–399.

Chang, G.L., Wu, J., 1994. Recursive estimation of time-varying origin–destination flows from traffic counts in freeway corridors. Transportation Research Part B 28 (2), 141–160.

Chen, C., Petty, K., Skabardonis, A., Varaiya, P., Jia, Z., 2001. Freeway performance measurement system: mining loop detector data. Transportation Research Record 1748, 96–102.

Chui, C., Chen, G., 1991. Kalman Filtering with Real-Time Applications. Springer-Verlag.

Cremer, M., Keller, H., 1981. Dynamic identification of flows from traffic counts at complex intersections. In: Hauer, E., Hurdle, V. (Eds.), Proceedings of the 8th International Symposium on Transportation and Traffic Theory. University of Toronto Press, Toronto, pp. 121–142.

Cremer, M., Keller, H., 1987. A new class of dynamic methods for the identification of origin–destination flows. Transportation Research Part B 21 (2), 117–132.

Harvey, A.C., 1989. Forecasting Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Kang, Y., 1999. Estimation and prediction of dynamic origin–destination demand and system consistency control for real-time dynamic traffic assignment operation, Ph.D. Dissertation, University of Texas at Austin, USA.

Mahmassani, H.S., 2001. Dynamic network traffic assignment and simulation methodology for advanced system management applications. Networks and Spatial Economics 1 (2), 267–292.

Mahmassani, H.S., Hu, T.-Y., Peeta, S., Ziliaskopoulos, A., 1994. Development and testing of dynamic assignment and simulation procedure for ATIS/ATMS applications. Report DTFH61-C-00074-FG, Center for Transportation Research, University of Texas, Austin, TX.

Mahmassani, H.S., Hawas, Y., Abdelghani, K., Abdelfatah, A., Chiu, Y.-C., Kang, Y., 1998. DYNASMART-X; Volume II: analytical and algorithmic aspects. Technical Report ST067-85, Center for Transportation Research, The University of Texas at Austin, USA.

Morrison, N., 1969. Introduction to Sequential Smoothing and Prediction. McGraw-Hill, New York.

Okutani, I., Stephanedes, Y., 1984. Dynamic prediction of traffic volume through Kalman filtering theory. Transportation Research Part B 18 (1), 1–11.

Peeta, S., Ziliaskopoulos, A., 2001. Foundations of dynamic traffic assignment: the past, the present and the future. Networks and Spatial Economics 1 (2), 233–266.

Tavana, H., 2001. Internally-consistent estimation of dynamic network origin–destination flows from intelligent transportation systems data using bi-level optimization. Ph.D. Dissertation, University of Texas at Austin, USA.

Tavana, H., Mahmassani, H.S., 2001. Estimation of dynamic origin destination flows from sensor data using bi-level optimization method. Transportation Research Board CD-ROM Paper Preprints (TRB Paper No. 01-3241), National Research Council, Washington, DC.

Varaiya, P., 2002. California's performance measurement system: improving freeway efficiency through transportation intelligence. TR News 218, 18–24.

West, M., Harrison, B.J., 1997. Bayesian Forecasting and Dynamic Models, second ed. Springer-Verlag.

Zhou, X., Mahmassani, H.S., 2005. Recursive approaches for online consistency checking and O–D demand updating for real-time dynamic traffic assignment operation. Journal of the Transportation Research Board 1923, 218–226.

Zhou, X., Qin, X., Mahmassani, H.S., 2003. Dynamic origin–destination demand estimation using multi-day link traffic counts for planning applications. Transportation Research Record 1831, 30–38.