Improving Scalability of Generic Online Calibration for Real-Time Dynamic Traffic Assignment Systems

Prakash, A. Arun; Seshadri, Ravi; Antoniou, Constantinos; Pereira, Francisco C.; Ben-Akiva, Moshe

Published in:
Transportation Research Record

Link to article, DOI:
10.1177/0361198118791360

Publication date:
2018

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Improving Scalability of Generic Online Calibration For Real-time Dynamic Traffic Assignment Systems

Article in Transportation Research Record Journal of the Transportation Research Board - January 2018
DOI: 10.1177/0361198118791360

5 authors, including:

Arun Prakash Akkinepally
Massachusetts Institute of Technology
40 PUBLICATIONS 215 CITATIONS
See Profile

Constantinos Antoniou
Technische Universität München
297 PUBLICATIONS 3,339 CITATIONS
See Profile

Ravi Seshadri
Singapore-MIT Alliance for Research and Technology
33 PUBLICATIONS 245 CITATIONS
See Profile

Francisco C. Pereira
Technical University of Denmark
233 PUBLICATIONS 2,526 CITATIONS
See Profile

Some of the authors of this publication are also working on these related projects:

Choice Set Formation View project

Socio-economic evaluation of Grand Paris Express and other large transport infrastructure projects View project
IMPROVING SCALABILITY OF GENERIC ONLINE CALIBRATION FOR REAL-TIME DYNAMIC TRAFFIC ASSIGNMENT SYSTEMS

A. Arun Prakash  
Corresponding Author  
Massachusetts Institute of Technology  
arunakkin@gmail.com

Ravi Seshadri  
Singapore-MIT Alliance for Research and Technology (SMART)  
ravi@smart.mit.edu

Constantinos Antoniou  
Technical University of Munich  
c.antoniou@tum.de

Francisco C. Pereira  
Technical University of Denmark  
camara@mit.edu

Moshe Ben-Akiva  
Massachusetts Institute of Technology  
mba@mit.edu

Word Count: 6387 words + 1500 (5 figures + 1 table) = 7887

Submission Date: 15 November 2017
ABSTRACT
Flexible calibration of Dynamic Traffic Assignment (DTA) systems in real-time has important applications in effective traffic management. However, the existing approaches are either limited to small networks or to a specific class of parameters. In this light, this study presents a framework to systematically reduce the dimension of the generic online calibration problem making it more scalable. Specifically, we propose a state-space formulation of the problem in the reduced dimension space. Following which the problem is solved using the Constrained Extended Kalman Filter, which is made tractable because of the low dimensionality of the formulated problem. The effectiveness of the proposed approach is demonstrated using a real-world network leading to better state estimation by 13% and better state predictions by 11% —with a 50 fold dimensionality reduction. Insights into choosing the right degree of dimensionality reduction are also discussed. This work has the potential for a more wide-spread application of real-time DTA systems in practice.
INTRODUCTION
The pervasive problem of traffic congestion in urban transportation networks worldwide has led to an increasing emphasis on developing tools for real-time traffic management and operations. In this context, simulation-based DTA model systems (e.g. DynaSMART [1], DynaMIT [2]) have proven effective solutions for traffic management, which includes providing effective traveler guidance and optimizing network control strategies. However, the efficacy of these DTA systems depends on their ability to accurately estimate and predict traffic conditions. A crucial component in traffic state estimation and prediction is online calibration which involves adjusting the demand and supply parameters of a DTA system in real time—using observed traffic measurements—so as to accurately estimate and predict traffic states. One widely used approach for solving the calibration problem is the extended Kalman Filter (EKF) [3], which is appealing as it can incorporate both demand and supply side parameters and any type of measurement data. However, a key challenge in adopting the EKF for online applications is its scalability to large networks—an issue that this work attempts to address.

The motivation for this study is twofold. First, the computationally intensive nature of the EKF which poses challenges for online deployment. This is due in large part to the need for computing numerical derivatives in real–time that involves expensive simulation runs—an issue that is further exacerbated as the number of calibration parameters increases. The second motivating factor is applications—such as traveler guidance generation and control strategy optimization—which require real–time performance of the DTA system.

In view of these motivating considerations, this study pursues the following objectives: 1) To develop a framework to reduce the dimensionality of the generic online calibration problem, 2) To formulate the state–space model and the EKF solution approach for online calibration in the reduced parameter space, and 3) To examine the performance of the principal component based EKF on the Singapore expressway network using real data and perform a sensitivity analysis on the degree of dimensionality reduction.

This paper contributes to the existing literature in the following respects. First, the state space model and EKF solution approach for the online calibration problem is formulated in a reduced parameter space. Specifically, the approach is demonstrated using the principal component analysis for dimensionality reduction. Second, a case study conducted on the Singapore expressway network based on real data demonstrates that using principal components which explain 95% of the variance achieves a 52 fold dimensionality reduction whilst improving state estimation and prediction accuracy. The dimensionality reduction makes the EKF tractable and feasible for real time application. Thirdly, the numerical experiments and sensitivity analysis yield several insights regarding the effect of using principal components within the EKF based formulation.

We make the following salient assumptions in applications discussed in the paper. First, we assume that the principal component directions are time-invariant. Second, we assume a linear transition equation. Third, we adapt a sequential online calibration approach where the estimated parameters of the previous interval are assumed to be fixed. Each of these assumptions are discussed in greater detail in the subsequent sections.

LITERATURE REVIEW
The calibration of simulation-based dynamic traffic assignment (DTA) models is critical in obtaining accurate traffic state estimates and predictions and has traditionally been studied in two contexts, offline and online. The offline calibration problem involves estimating historical values
for simulation parameters so as to best replicate average or historical traffic conditions for a given network. Solution methodologies for the dynamic origin-destination (OD) estimation problem (which focuses on demand-side parameters) typically involve optimization formulations including Generalized Least Squares (GLS) approaches [4, 5, 6] and state space formulations [7, 8]. The GLS approach, however, does not lend itself well to the combined calibration of demand and supply parameters and to address this, Balakrishna et al. [9] proposed a more generic solution method for the offline problem based on the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm that can incorporate any type of measurement data. We refer the reader to [10] for a more detailed discussion of offline calibration and focus our attention on online calibration.

The online calibration problem involves updating the historical simulation parameters (such as origin-destination flows on the demand side and segment capacities/traffic dynamics parameters on the supply side) in real-time to match simulated and prevailing traffic conditions as closely as possible [11]. The existing literature can broadly be categorized into studies that consider demand-side parameters, supply-side parameters and the combination of demand and supply side parameters.

Several studies on the online calibration problem have focused solely on the calibration of origin-destination demands using sensor count data. Ashok and Ben–Akiva [12] proposed a Kalman filtering approach formulated in terms of deviations of OD flows from historical values (as opposed to the OD flows themselves). This enables incorporation of apriori information regarding underlying structural relationships between the OD flow variables. Ashok and Ben–Akiva [13] extended this approach to explicitly account for stochasticity in the assignment matrix (which maps OD flows to traffic counts). Zhou and Mahmassani [14] also use the Kalman filter for the real time OD estimation problem where the transition equation is a polynomial trend filter that captures historical trends and structural deviations. Zhang et al. [15] explicitly modeled constraints (such as non-negativity of OD flows) in the Kalman Filter based approach and demonstrated significant improvements in accuracy for large networks where the constraints are often violated.

With regard to the calibration of supply–side parameters, Zhou and Mahmassani [16] proposed a dynamic programming approach that uses an approximation of the simulator based on macroscopic models to adjust flow propagation. Antoniou et al. [17] proposed the use of the extended, limiting, and unscented Kalman filters for the calibration of speed–density parameters.

A key challenge that arises in using the extended Kalman filter (EKF) for online applications is the linearization of the measurement equation which necessitates the computation of numerical derivatives that is computationally intensive, particularly when the number of calibration parameters is large. To address this issue, Antoniou et al. [3] proposed limiting EKF as a practical alternative; the limiting EKF which utilizes offline computed derivatives and drastically reduces the computational complexity of the EKF, without sacrificing significantly on accuracy. The authors simultaneously calibrate supply and demand parameters. Hashemi and Abdelghany [18] also address the simultaneous calibration problem and propose an approach where the OD demands are calibrated using a least squares approach and supply side parameters are calibrated using a feedback controller.

The scalability of online calibration approaches has received some attention in the recent literature. Frederix et al. [19] proposed a hierarchical approach that decomposes the OD estimation problem for the entire network and enables the application of different OD estimation methods to different subnetworks. Djukic et al. [20, 21] apply the concept of principal components analysis to the dynamic OD estimation problem using a Kalman filtering approach. Another study
that addresses dimensionality reduction is that of Prakash et al. [22] who propose a principal components based GLS approach that significantly reduces dimensionality of the online calibration problem. However, our work differs from that of earlier studies in three aspects. First, the proposed framework is generic and can incorporate (i) any technique for dimensionality reduction (like auto-encoders and kernel principal component analysis), (ii) any set of parameters (pertaining to both supply models and demand models), and (iii) a wide variety of measurements (like sensor counts, AVI, and GPS). Second, we examine the performance of dimensionality reduction in the context of online calibration (with a focus on both traffic state estimation and prediction) as opposed to offline calibration. Third, we utilize real–world data to test the performance of the principal component based formulation and perform a detailed sensitivity analysis to gain insights on the effect of the number of principal components on estimation and prediction accuracy.

In summary, a large body of work on the calibration of DTA systems utilizes extended kalman filtering based approaches which are plagued by issues of scalability which have not been systematically addressed in an online context. This paper attempts to address this gap.

**ONLINE CALIBRATION: STATE SPACE PROBLEM FORMULATION**

The online calibration problem is briefly formulated in the following paragraphs. The reader is referred to [3, 13] for a more detailed discussion. We present the calibration problem as a standard state-space formulation which is a classical approach to model dynamical systems.

Consider an analysis period \( T \) which is divided into equal intervals \( h = 1, 2, 3 \ldots n \) of size \( t \). The transportation network is represented by \( G(N, L, S) \), where \( N \) represents the set of nodes, \( L \) represents the set of links, and \( S \) represents the set of segments. The network has \( n_N \) nodes, \( n_L \) links, and \( n_S \) segments. The segments are sections of road with homogeneous geometry; a link comprises one or more segments. The set of OD pairs are represented by \( K \) and are \( n_K \) in number. Further, \( n_s \) of the \( n_S \) segments are assumed to be equipped with surveillance sensors.

The state-space formulation consists of three main components: (i) a state vector that succinctly characterizes the system, (ii) a transition equation that captures the evolution of the system—through the state-vector—in time, and (iii) a measurement equation that captures the relationship between the state-vector and the measurements or observations of the system. Following the proposal in [13], we express the state-space formulation in terms of the deviations of the relevant variables from their historical values. As mentioned in [3], by modelling in terms of the deviations in parameters, we can capture the structural information of the trip and network patterns present in the historical values.

In the current context, let \( \pi_h \) represent vector containing the parameters of the DTA system in the time interval \( h \); it can contain the OD flow variables along with behavioral and supply parameters. Let \( \pi_h^H \) represent the historical values of the parameters in interval \( h \). The historical values \( \pi_h^H \) are generally obtained through offline calibration [10]. The state-vector is then denoted by \( \Delta \pi_h = \pi_h - \pi_h^H \), whose evolution can be represented through the following generic transition equation:

\[
\Delta \pi_h = f (\Delta \pi_{h-1}, \ldots, \Delta \pi_{h-q}) + \eta_h
\]

where \( q \) denotes the number of previous interval’s states that influence the current interval’s state; and, \( \eta_h \) is the vector random errors in the transition equation.

Further, let \( M_h \) denote the vector of measurements in interval \( h \); these can be both point-based measurements (like flows and speeds) or spatial measurements (like GPS or AVI). As before,
let $M^H_h$ represent the historical values of the measurements in interval $h$. The historical values of measurements are generally obtained by running the DTA system with the historical parameters or they can also be actual measurements used for offline calibration. The measurement-vector in deviations is then denoted by $\Delta M_h = M_h - M^H_h$, which is related to the state-vector through the following generic measurement equation:

\[
\Delta M_h = g(\Delta \pi_h, \ldots, \Delta \pi_{h-p}) + \zeta_h
\]  

where $p$ denotes the number of previous interval’s states that influence the current interval’s measurements; and $\zeta_h$ is the vector random errors in the measurement equation.

Equations (1) and (2) together form the state-space formulation of the generic online calibration problem. In practice, two assumptions are made on the generic formulation: (i) the transition equation is approximated through a linear autoregressive process, and (ii) the measurement equation is expressed in terms of the simulator. After applying the above assumptions, the resulting transition and measurement equations are as follows:

\[
\Delta \pi_h = \sum_{i=h-q}^{h-1} F^h_i \Delta \pi_i + \eta_h
\]

\[
\Delta M_h = S(\Delta \pi_h + \pi^H_h, \ldots, \Delta \pi_{h-p} + \pi^H_{h-p}) - S(\pi^H_h, \ldots, \pi^H_{h-p}) + \zeta_h
\]

In equation (3a), $F^h_i$ represents a matrix relating the parameter estimates of interval $i$ to the estimates of interval $h$, and $q$ denotes the degree of the autoregressive process in the deviations. In equation (3b), $S(.)$ represents the simulation model whose inputs are the parameters and outputs are the simulated measurements in the current interval, and $p$ represents the the maximum number of previous interval’s parameters that influence the measurements in the current interval.

**Online OD-flow estimation problem**

The scalable version of the extended Kalman Filter proposed in this study can be applied to any set of parameters and measurements. To demonstrate the proposed approach we choose the OD flow estimation problem which is a special case of the online calibration problem in (3). The parameters of interest are the time-dependent OD flows and the measurements are the sensor-flow counts.

Let $x_h$ represent vector of OD-flows in interval $h$, whose length is $n_K$. Let $x^H_h$ represent the historical OD-flow values in interval $h$. Then the state-vector in the OD flow estimation problem is given by $\Delta x_h = x_h - x^H_h$. Similarly, let $y_h$ represent vector of sensor-flow counts in interval $h$ whose length is $n_s$; and let $y^H_h$ denote the historical sensor-flow counts in interval $h$. The measurement-vector in deviations is then denoted by $\Delta y_h = y_h - y^H_h$.

The transition and measurement equations for the state-space formulation of the OD-flow estimation problem can be given by

\[
\Delta x_h = \sum_{i=h-q}^{h-1} F^h_i \Delta x_i + u_h
\]

\[
\Delta y_h = S(\Delta x_h + x^H_h, \ldots, \Delta x_{h-p} + x^H_{h-p}) - S(x^H_h, \ldots, x^H_{h-p}) + v_h
\]
where $F^h_i$ is an $n_K \times n_K$ matrix relating the OD estimates of interval $i$ to the OD estimates of interval $h$; $u_h$ is the vector of random errors in the transition equation of OD-flows with covariance matrix $\Omega_{u_h}$; $v_h$ is the vector of random errors in the measurement equation of sensor-flow counts with covariance matrix $\Omega_{v_h}$.

In standard OD-flow estimation problem the measurement equation (4b) is linearized and written as

$$\Delta y_h = \sum_{j=h-p}^{h} A^h_j \Delta x_j + v_h$$

using an assignment matrix $A^h_j$ which assigns the proportion of OD flows from interval $j$ to measurements in interval $h$. We refrain from making this assumption to demonstrate that the proposed procedure is applicable even when seemingly linear relationships between measurements and parameters do not exist.

In the context of online systems, the problem in (4) is solved to obtain an estimate of only the OD flow deviations in interval $h$, $\Delta x_h$. The OD-flow estimates of the earlier intervals $h-1, h-2, \ldots$ are not re–estimated. In effect, the OD-flow estimates of the previous intervals are treated as constants in the current interval; this is inherently assumed in the subsequent discussions.

The sensor flow counts in the current time interval might contain information about OD–flow parameters from the previous time intervals. Therefore, ideally, the OD-flow estimates from the previous time intervals need to be corrected based on the sensor-flow counts in the current interval. In the context of state space formulation, this correction of the parameters estimated in previous time intervals is formulated using state augmentation approach [7]. We would like to emphasize that the proposed approach can be extended to the state augmentation formulation where the state comprises the principal components corresponding to multiple time-intervals. Although state augmentation results in better estimates of the parameters, it can be computationally intensive as it requires ‘rolling back’ the simulator in real–time. In other words, the simulator needs to re–run with the new parameter estimates (from the previous time intervals) to update the simulated measurements. Ashok and Ben-Akiva [13] found that the state augmentation can be reasonably approximated using the sequential estimation procedure we described in the paper.

To predict the OD flow vectors in the subsequent intervals $h+1, h+2, \ldots$, the estimates calculated using equation in (4a) are used. For example, the prediction for the time interval $h+1$, if the current interval is $h$, is calculated as

$$x_{h+1} = x_{h+1}^{hf} + \sum_{i=h-q+1}^{h} F^h_i \Delta \hat{x}_i$$

(5)

where $\Delta \hat{x}_i$ represent the a posteriori estimates of OD-flow deviations in interval $i$.

**Extended Kalman Filter**

The Kalman Filter based approaches are a natural and efficient way to recursively solve the system of equations in (4). The Kalman Filter efficiently determines the estimate of the current interval using the estimate from the previous interval and the measurements in the current interval. The classical Kalman Filter — which is a minimum mean square estimator — is applicable to linear state-space models. For the non-linear state-space models — as in the problem in (4) — the Extended Kalman Filter (EKF) was introduced. The EKF linearizes the non-linear transition or measurement
equations around the apriori estimates and adopts the procedure for linear state-space models to estimate the state-vector. The main conceptual drawback of the EKF is that it is not an optimal estimator, however, its practical application has been demonstrated in various studies [11]. Note that for the application of the Kalman Filter based methods, we impose an assumption that the error terms $u_h$ and $v_h$ — from equations (4a) and (4b) — are zero mean Gaussian variables and that they are independent over time.

There are three main steps in the traditional EKF algorithm when applied to the OD-flow estimation problem: (i) time-update, (ii) linearization, and (iii) measurement update. In the time-update step, the prediction for OD-flow deviation vector $\Delta x_{h|1-h-1}$ and its covariance matrix $P_{h|1-h-1}$ are made using the transition equation and the optimal estimates from the previous interval $h$, $\Delta x_{h-1|1-h-1}$ and $P_{h-1|1-h-1}$. These estimates — $\Delta x_{h|1-h-1}$ and $P_{h|1-h-1}$ — are termed a priori estimates as they use data only until the previous time-interval $h - 1$. In the second step, the measurement equation is linearized around the a priori estimates of the current interval — $\Delta x_{h|1-h-1}$ — resulting in

$$
\Delta y_h = S(\Delta \hat{x}_{h|1-h-1} + x_h^H \Delta x_{h-1} + x_h^H \Delta x_{h-1} \ldots \Delta x_{h-p} + x_h^H) - S(x_h^H, \ldots, x_{h-p}^H) + H_h \Delta x_h - H_h \Delta \hat{x}_{h|1-h-1} + v_h
$$

where

$$
H_h = \nabla S(\Delta x_h^H \ldots \Delta x_{h-p}^H) \bigg|_{\Delta x_{h} = \Delta x_{h|1-h-1}}
$$

represents the $n_s \times n_K$ Jacobian of the simulator with respect to $\Delta x_h$ at $\Delta x_{h|1-h-1}$. As discussed, $\Delta x_{h-1} \ldots \Delta x_{h-p}$ are held fixed in the current estimation interval $h$. In the third step, the a priori estimates are updated using the linearized measurement equation and the matrix $H_h$ to obtain the a posteriori estimates for OD-flow deviation vector $\Delta x_{h|1-h}$ and its covariance matrix $P_{h|1-h}$. Please refer to [11] for a more detailed discussion of the procedure.

To main bottleneck to apply the EKF algorithm to the online calibration problem is the calculation of the Jacobian of the measurement equation i.e., the matrix $H_h$. As measurement equation represents the simulator, it has no closed-form expression. Therefore, $H_h$ is calculated through the numerical derivatives of $S(.)$. Determining the centered numerical derivate involves perturbing each of the parameters and running the simulator $2n_K$ times, which can be computationally intractable. The procedure outlined in the next section alleviates this problem.

**STATE SPACE FORMULATION IN THE REDUCED PARAMETER SPACE**

This section discusses the approach to reduce the dimensionality of online calibration problem and solve it using the state-space formulation in the reduced parameter space. The dimensionality reducing technique discussed the principal component analysis, however, the approach is generic and can incorporate other techniques. Subsequently, the online calibration problem in formulated as a state-space formulation with the principal components as the state-vector. Finally, the estimates of parameter vector are constructed back from the estimates of the principal components.

**Principal components**

The principal components are the linear combinations of the parameters vector that explain most of the variability of the parameters. The co-efficients of the linear combinations are the principal component directions. The principal components also have the following desirable properties: (i) they are orthogonal to each other — in the sample data and (ii) the first principal component explains the most variance followed by the second principal component and so on. Thus, the
principal components not only approximate the parameter vector, they also decouple them. We will
discuss how such decoupling of parameters is beneficial in the next section where we formulate
the problem. Please refer to [22] for a detailed discussion on the construction of the principal
components and their interpretation in the context of online calibration.

In the subsequent discussion, let $\mathbf{V}$ represent a $n_K \times n_d$ the matrix whose columns are the
first $n_d$ chosen principal components of the OD-flow vector:

$$
\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \cdots & \mathbf{v}_{n_d} \end{bmatrix}
$$

(8)

Then the $n_d \times 1$ principal component vector $\mathbf{z} = [z_1, z_2, \ldots, z_{n_d}]^T$ of the OD flow vector $\mathbf{x}$ can be
written as

$$
\mathbf{z} = \mathbf{V}^T \mathbf{x}
$$

(9)

and the OD flow vector $\mathbf{x}$ can be approximately reconstructed as

$$
\mathbf{x} \approx \mathbf{Vz}
$$

(10)

In the following discussion, we assume that the principal components directions $\mathbf{V}$ is constant
across time-intervals. This implies that the statistical correlations between the OD demands is
assumed to be constant across time intervals $h$. But note that the structural differences and
evolution in OD demands (and their principal components) across time intervals are captured using
the transition equation (equations (4a) and (12)). Indeed, the statistical correlations between the
OD demands can change with the time-of-day. In which case, we would need to use principal
components directions specific to the time-intervals $\mathbf{V}_h$. However, calculating the time-interval or
time-period specific principal components, though desirable, can have high data requirements.

**Problem formulation**

Similar to the state-space formulation in terms of the OD-flow deviations, we formulate the on-
line calibration problem in terms of the deviations of the principal components. The principal
component deviations can be written as

$$
\Delta \mathbf{z}_h = \mathbf{z}_h - \mathbf{z}_h^H
$$

(11)

where $\mathbf{z}_h^H = \mathbf{V}_h^T \mathbf{x}_h^H$. The transition equation expressing the evolution of the principal component
state-vector is given by

$$
\Delta \mathbf{z}_h = \sum_{i=h-q}^{h-1} \mathbf{G}_i^h \Delta \mathbf{z}_i + \mathbf{w}_h
$$

(12)

where $\mathbf{G}_i^h$ is an $n_d \times n_d$ matrix relating the principal component estimates of interval $i$ to the
principal component estimates of interval $h$; and $\mathbf{w}_h$ represents the error vector of size $n_d \times 1$ in the
transition equation of principal components with covariance matrix $\mathbf{\Omega}_w^h$. As before, we assume
that $\mathbf{w}_h$ are zero mean normal random variables that are independent over time intervals.

Note that the transition equation in (12) assumes that the only the considered $n_d$ principal
components are related through the autoregressive process. The other $n_K-n_d$ principal components
are assumed not to represent the structural variation in the data and are assumed to be inherently
random.
In practice, three assumptions are made in the transition equations —both in (4a) and (12): (i) the matrices in the transition equations are assumed to be diagonal, (ii) the error components in the error vector are assumed to be independent, and (iii) the errors are assumed to be normally distributed. These three assumptions are more justifiable for the principal component based formulation as discussed below.

The matrices in the transition equations are assumed to be diagonal as it is generally difficult to estimate the relationships between different parameters across time intervals when the state-vector is large. Fortunately, assuming that $G^h_i$ in equation (12) is diagonal is more reasonable compared to assuming that $F^h_i$ in equation (4a) is diagonal. This follows from the property that the principal components are orthogonal to each other: a principal component does not explain—at least linearly— another principal component. Therefore, the non-diagonal terms in matrix $G^h_i$ can be ignored which simplifies its estimation. Note that the dependence between the OD-flows is implicitly captured in (12) through the principal components. By following the same reasoning—that the principal components are orthogonal to each other—the random vector $w^h$ can reasonably be assumed to consist of $n_d$ independent random variables. Finally, as the principal components are a linear combination of many OD-flows they can be treated as normal random variables by invoking the central limit theorem.

Substituting (10) in equation (4b), the measurement equation in terms of the principal components can be written as

$$
\Delta y^h = S(V\Delta z^h + Vz^H_{h-1} + Vz^H_{h-p}) - S(Vz^H_{h} + \ldots + Vz^H_{h-p}) + v^h \tag{13}
$$

The set of equations (12) and (13) can be solved recursively using the EKF algorithm. The EKF procedure is discussed in the following section.

**Constrained EKF algorithm for the principal component formulation**

In this section we present the steps of the EKF algorithm to determine the estimate of the principal components. We also incorporate the enhancement to the EKF algorithm—which was proposed in [15]—to impose constraints on the principal components. These constraints ensure that the OD-flows constructed from the estimated principal components are non-negative. After determining the estimates of the principal components in the current time-interval, we reconstruct the OD-flows which are assigned to the network.

The EKF algorithm presented above has five main steps. First, the time-update step performed through equations (16) and (17) obtains a priori estimates of the state vector $\Delta z_{h|h-1}$ and its covariance matrix $P_{h|h-1}$. The second step involves computing the Jacobian $J_h$ —which is a $n_s \times n_d$ matrix— of measurement equation with respect to the state-vector of the current interval at its a priori estimate, i.e. at $\Delta z_h = \Delta z_{h|h-1}$. In the third step, which is the measurement update, the posteriori estimates of state vector $\Delta z_{h|h}$ and its covariance matrix $P_{h|h}$ are determined as per the traditional Kalman Filter equations (refer equations (19), (20), and (21)). In the fourth step, we impose the constraint that the estimated state vector must result in non-negative OD-flows. This is done by solving an optimization problem to determine the most likely value in the a posteriori distribution obtained from the measurement update subject to the constraints. Finally, in the fifth step, we determine the OD-flows estimate by constructing them from the principal component estimates and the historical OD-flows as given in equation (23). We discuss the linearization and constraint imposition steps in the subsequent paragraphs.
Algorithm 1 Constrained Extended Kalman Filter in Principal Components

Initialization
\[
\Delta z_{0|0} = \Delta z_0 \quad (14)
\]
\[
P_{0|0} = P_0 \quad (15)
\]

for \( h = 1 \) to \( N \) do

Time Update
\[
\Delta z_{h|h-1} = \sum_{i=h-q}^{h-1} G_i^h \Delta z_{i|i} \quad (16)
\]
\[
P_{h|h-1} = \sum_{i=h-q}^{h-1} G_i^h P_{i|i} G_i^{hT} + \Omega_{wh} \quad (17)
\]

Linearization
\[
J_h = \nabla_{\Delta z_h} S(\mathbf{V}\Delta z_h + \mathbf{V}z_h^H, \ldots, \mathbf{V}\Delta z_{h-p} + \mathbf{V}z_{h-p}^H) \mid _{\Delta z_h = \Delta z_{h|h-1}} \quad (18)
\]

Measurement Update
\[
K_h = P_{h|h-1} J_h^T \left( J_h P_{h|h-1} J_h^T + \Omega_v \right)^{-1} \quad (19)
\]
\[
\Delta z_{h|h} = \Delta z_{h|h-1} + K_h \left[ \Delta y_h - \{ S(\mathbf{V}\Delta z_{h|h-1} + \mathbf{V}z_h^H, \ldots) - S(\mathbf{V}z_{h-1}^H, \ldots, \mathbf{V}z_{h-p}^H) \} \right] \quad (20)
\]
\[
P_{h|h} = P_{h|h-1} - K_h J_h P_{h|h-1} \quad (21)
\]

Imposing Constraints
\[
\hat{\Delta z}_{h|h} = \arg \max \{ (\Delta z - \Delta z_{h|h})^T P^{-1}_{h|h} (\Delta z - \Delta z_{h|h}) \mid \mathbf{V}\Delta z + x_h^H \geq 0 \} \quad (22)
\]

Constructing parameter estimates
\[
\hat{x}_{h|h} = \mathbf{V}\hat{\Delta z}_{h|h} + x_h^H \quad (23)
\]
end for

In the linearization step in equation (18), the Jacobian is computed for the function \( S(\cdot) \) which represents the simulator. Therefore, the Jacobian cannot be computed analytically. We compute the Jacobian numerically using a centered finite-difference method as follows. Let \( \delta z_i \) represent \( n_d \times 1 \) the perturbation vector which represents the perturbation of the \( i^{th} \) principal component, i.e. \( \delta z_i = [0, 0 \ldots c_i \ldots 0]^T \). In other words, \( \Delta z_{h|h-1} + \delta z_i \) represents the perturbation of only the \( i^{th} \) principal component by an amount \( c_i \) while holding all the other principal components the same. Let \( j_{ih} \) denote the \( i^{th} \) column in the Jacobian \( J_h \) —it represents the gradient of function \( S \) with respect to the \( i^{th} \) principal component. The vector \( j_{ih} \) is computed numerically as
\[
j_{ih} = \frac{S(\mathbf{V}(\Delta z_{h|h-1} - \delta z_i) + \mathbf{V}z_h^H, \ldots) - S(\mathbf{V}(\Delta z_{h|h-1} + \delta z_i) + \mathbf{V}z_h^H, \ldots)}{2c_i} \quad (24)
\]
We can construct the Jacobian matrix $J_h$ by concatenating the vectors $\mathbf{j}_i^h$, $\forall i = 1, 2, \ldots, n_d$ into matrix as $J_h = [\mathbf{j}^h_1, \mathbf{j}^h_2, \ldots, \mathbf{j}^h_{n_d}]$. Thus, to determine $J_h$ we need to run the simulator —for the duration of the current time interval $h$— $2n_d$ times. Note that, we have reduced the number of runs required to compute the Jacobian from $2n_K$ in OD-flow state-space formulation to $2n_d$ in the current principal component formulation. Further, as each of these runs of the simulator are independent of each other, we can run them in parallel. Further, as the number of parameters are reduced and only the most important ones are selected in the principal component based formulation, the error in the numerical approximation of the Jacobian might also be reduced.

In step four of the algorithm, constraints are imposed on the final estimate of the state vector. These constraints ensure that the OD-flow vector constructed from the estimated principal component vector is non-negative. In absence of these constraints, the procedure might result in negative OD-flows. In practice, the estimated negative OD-flows are truncated at zero. However, such truncation might result in erroneous and diverging results as discussed in [15]. Moreover, as the estimated principal components do not ensure that such truncation is minimal it is necessary to impose the non-negative OD-flow constraints. Note that constraints on any other kinds of parameters can be similarly imposed.

The constraints are imposed by correcting the posteriori estimates from the measurement update step. Specifically, the measurement update step determines the posteriori estimates of state $\Delta \mathbf{z}^h_{j|h}$ and its covariance matrix $\mathbf{P}^h_{j|h}$. As the state-vector is assumed to be normal random vector in the EKF, the measurement update step essentially determines the posteriori distribution of the state-vector $\Delta \mathbf{z}^h$ with mean $\Delta \mathbf{z}^h_{j|h}$ and covariance $\mathbf{P}^h_{j|h}$. We impose the constraints on the final state estimate by determining the value with the maximal probability in the posteriori distribution subject to the constraints. We determine such estimate by solving the optimality problem defined in equation (22). The constraints in equation (22) ensure that the OD-flows derived from the final state estimate $\Delta \hat{\mathbf{z}}^h_{j|h}$ are non-negative. Please refer to [15] for more details regarding the procedure.

**CASE STUDY ON SINGAPORE EXPRESSWAY NETWORK**

In this section, we apply the procedure presented in Section 5.3 to the Singapore Expressway Network. Firstly, the simulation setup and data is explained followed by the estimation of the inputs to the online calibration. Finally, the results are discussed.

**Overview**

The case study is conducted on the Singapore Expressway Network and the real–time DTA system used is DynaMIT-R [2]. The road network is depicted in Figure 1, which has 939 nodes, 1157 links, and 3906 segments. The network specification also contains information about segment lengths, segment curvatures, speed limits, lanes specifications, lane–connections, and dynamic tolling gantries which are replicated from the real–world.

The network has 4121 OD pairs, whose locations and historical values were determined by an earlier work through offline calibration [23]. The work also determined the supply–side parameters, which include the modified Greenshield’s speed–density equation parameters and the segment capacities. The network has 357 sensors, each of which is associated with a segment; these video camera based sensors count the vehicular flow for a period of 5 minutes. The Land Transport Authority (LTA) of Singapore provided the measured sensor counts after preprocessing the raw-data.

For the current case study, the simulation time–period was taken from 06:00 hrs to 12:00 hrs.
which includes the morning peak period and also the peak to off-peak transition. The estimation interval was 5 minutes and the prediction interval —to estimate future traffic states and provide guidance— was 15 minutes. Thus, we have 72 estimation intervals with a total of $72 \times 4121 = 102,312$ variables.

For this study, sensor count data of 30 weekdays in August-September 2015 was used. The 30 day data was divided into a training set of the first 25 days and the calibration procedures were tested on the last 5 days.

**Estimating the inputs to the calibration**

To solve the online OD-flow estimation problem in (4), estimates of the following input variables are required: covariance matrix of the errors in the transition equation $\Omega_u$, covariance matrix of the errors in the measurement equation $\Omega_v$, and the autoregressive matrices $F^h_i$ for $i = h, h-1, \ldots, h-q$. Additionally, to solve the principal component based estimation problem in equations (12) and (13) following variables are needed: the principal component directions of the OD flow vector $V$, covariance matrix of the errors in the transition equation $\Omega_w$, and the autoregressive matrices $G^h_i$ for $i = h, h-1, \ldots, h-q$.

The covariance matrix $\Omega_v$, which measures the variance in sensor measurements was determined based on sensor credibility. The covariance matrix $\Omega_u$ and the autoregressive matrices $F^h_i$ for $i = h, h-1, \ldots, h-q$ were determined as follows. First, the online OD-flow estimation problem was solved for the days 11 to 25 using the Generalized Least Squares (GLS) approach. The initial 10 days —from day 1 to day 10— were used to determine the inputs to the formulated GLS problem. Then, using the determined OD-flow estimates, the autoregressive equation in (4a) was estimated for each OD pair assuming that the OD pairs are independent of each other; thus estimating the values of $F^h_i$ for $i = h, h-1, \ldots, h-q$ and $\Omega_u$. Note that a random walk was assumed when no autoregressive relation could be established. For more details regarding the procedure please refer to [11, 22].

The OD-flow estimates from days 11 to 25 were used to determine the principal component directions of the OD-flow vector. As discussed in [22], a data matrix was constructed —with $n_p = 72 \times 15 = 1080$ data points— on which the principal component analysis was carried out.
The results of the principal component analysis are presented in Figure 2.

![Figure 2: Principal Component Analysis of OD-flow vector using the estimated values from day 11 to day 25](image)

From Figure 2, we see that very few principal components explain majority of the variance in the OD-flow vector. Specifically, 80 principal components explain 95% of the variance and only 150 principal components explain around 99% of the variance. Further, after 80 principal components, we see that the percentage variance explained in minimal.

For the principal component based calibration, covariance matrix of the errors in the transition equation $\Omega_w$, and the autoregressive matrices $G^h_i$ for $i = h, h-1, \ldots, h-q$ were estimated as follows. First, the OD-flows estimated for days 11 to 25 were transformed into principal components. Then, an autoregressive equation was fit to each principal component independently. As before, if no autoregressive relation could be established for a principal component a random walk was assumed.

**Calibration Results**

The results from the online calibration on the 5 tests days for the Singapore Expressway network are discussed in this section. The performance measures adopted were the Normalized Root Mean Squared (RMSN) Errors and Mean Absolute Percentage Errors (MAPE) which are defined as

$$RMSN = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} y_i}}$$ (25a)

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right|$$ (25b)

where $y_i$ represents actual measurement and $\hat{y}_i$ represents simulated measurement. Note that as MAPE is not normalized to consider the size of measurement, it can be high when the sensor-flows are small even though the deviation from the observations is acceptable.
<table>
<thead>
<tr>
<th>Method</th>
<th>RMSN Est</th>
<th>RMSN Pred</th>
<th>MAPE Est</th>
<th>MAPE Pred</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>step1</td>
<td>step2</td>
<td>step3</td>
<td>step1</td>
</tr>
<tr>
<td>Hist</td>
<td>0.423</td>
<td>0.420</td>
<td>0.419</td>
<td>28.21</td>
</tr>
<tr>
<td>OD-EKF</td>
<td>0.287</td>
<td>0.295</td>
<td>0.305</td>
<td>19.77</td>
</tr>
<tr>
<td>PC-EKF 20</td>
<td>0.240</td>
<td>0.287</td>
<td>0.318</td>
<td>20.44</td>
</tr>
<tr>
<td>PC-EKF 40</td>
<td>0.241</td>
<td>0.284</td>
<td>0.310</td>
<td>20.34</td>
</tr>
<tr>
<td>PC-EKF 60</td>
<td>0.250</td>
<td>0.276</td>
<td>0.295</td>
<td>19.93</td>
</tr>
<tr>
<td>PC-EKF 80</td>
<td>0.250</td>
<td>0.265</td>
<td>0.275</td>
<td>18.46</td>
</tr>
</tbody>
</table>

The number of principal components used is depicted in the method column.

**TABLE 1**: Aggregate values of RMSNs and MAPEs of sensor-flow counts (5 min) for historical, OD-EKF, and PC-EKF calibration methodologies over the 5 test days. Note that method ‘Hist’ refers to the offline calibrated parameter values.

We compared performance of the OD-based calibration with EKF algorithm (OD-EKF), principal component based calibration with EKF algorithm (PC-EKF), and no online calibration. The no online calibration alternative represents the errors in the historical values which were obtained from the offline calibration. Further, the PC-EKF procedure was run by varying the number of chosen principal components. Specifically, we had four cases where we chose the first 20, 40, 60, and 80 principal components; these components cumulatively explained 55%, 75%, 90%, and 95% of the variance in the data respectively. The aggregate RMSNs and MAPEs of the procedures for estimation and prediction across the five test days are presented in Table 1.

In the context of estimation, the OD-EKF on average exhibits an RMSN of 0.287 and MAPE of 28.21%. It improves over historical by 32% in RMSN and 30% in MAPE. The PC-EKF with 80 principal components exhibited an RMSN of 0.25 and MAPE of 18.46%; its performance is better compared to OD-EKF by 13% in RMSN and 7% in MAPE. It appears that PC-EKF with sufficient principal components can perform well in estimation of the current state. In addition, the time-dependent RMSNs of the estimated sensor-flow counts for historical, OD-EKF and PC-EKF are presented in Figure 3a. We see that the performance of PC-EKF is better than OD-EKF for all the time-intervals considered.

Comparing the estimation performance of PC-EKF by varying the number of chosen principal components, we see curiously that the decreasing the number of principal components decreases the error in estimation. Specifically, we see the RMSN when 20 principal components were used is 0.24 and RMSN when 80 principal components were used is 0.25. This seemingly counter intuitive finding can be explained by noting that as the number of the principal components decrease the weight given to the a priori estimates in the objective function —of the Kalman Filter algorithm— also decreases. Therefore, decreasing the number of principal component essential increases the weight given to the current measurements resulting in the observed trend. It should be noted that this can happen only if the fewer number of chosen parameters can influence most of the observations. If this conjecture were true, we should observe poor predictions when the chosen principal components are too low in number even though the estimations are good —which we confirm in the following paragraphs.

The results from the predictions are represented in three steps, which represent the three 5
minute intervals in the complete 15 minute prediction interval. In the context of prediction, from Table 1, the OD-EKF on average exhibits an RMSN of 0.290, 0.295, and 0.3 for the three steps and MAPE of 24.60%, 26.44%, and 27.71% for the three steps. By comparing the RMSNs and MAPEs after calibration with those of the historical values, we see that RMSNs after calibration are significantly better than those of the historical, but the MAPEs after calibration are only marginally better than those of the historical. It implies that calibration improves on the predictions of sensor-flow counts of the sensors with higher flows than those with the lower flows. The PC–EKF with 80 principal components on average exhibited RMSNs of 0.252, 0.265, and 0.275 for the three steps and MAPEs of 22.13%, 24.31%, and 25.66% for the three steps. We see that PC-EKF outperforms OD-EKF in prediction in all the three steps of prediction. Specifically, PC-EKF with 80 principal components does better than OD-EKF by 11% in terms of RMSNs and by 8.5% in terms of MAPEs. The better predictive performance of PC-EKF is because of the following three

FIGURE 3: Plots of sensor count RMSNs for OD-EKF and PC–EKF with respect to time–of–day. The figure includes separate plots for estimation interval and each of the three steps in the prediction interval.
reasons. First, the PC-EKF formulation has lower ‘variance’ compared to OD-EKF formulation, thus reducing the effect of overfitting. Secondly, the transition equation in PC-EKF is estimated taking into consideration the structural dependencies between OD-flows, which are ignored in OD-EKF. In addition, such an estimation also results in a better estimation of the error covariance structure $\Omega_w$ as the principal components are orthogonal. Thirdly, the numerical approximation of the Jacobian in PC-EKF is better than that in OD-EKF. The time-dependent sensor count RMSNs in prediction are presented in Figures 3b, 3c, and 3d. These trends show that PC-EKF with 80 principal components almost consistently performs better than the EKF in predicting the sensor counts.

Comparing the predictive performance of PC-EKF by varying the number of chosen principal components, we see that decreasing the number of principal components deteriorates the predictive performance. Specifically, the average predictive performance in RMSNs of PC-EKF with 20, 40, 60, and 80 principal components is 0.286, 0.283, 0.274, and 0.263 respectively. This is in line with the findings discussed in the earlier paragraphs. Decreasing the principal components introduces bias and gives more importance to the measurements in the current interval, leading to poorer predictions. This trend is also observed in time-dependent error plots in Figures 3c and 3d where the prediction errors of PC-EKF get worse as the chosen number of principal components decrease.

The scatter plots representing the estimated and predicted sensor-flow counts from PC-EKF versus actual sensor-flow counts are presented in Figures 4 and 5. Each figure contains plots corresponding to the four levels of principal components chosen in PC-EKF. The complete five day’s sensor-flow count values are represented in a single plot as a heat map. A line was fit between the estimated/predicted and the actual values in each of the plots and its equation is presented. From Figure 4, we see that the estimation accuracy does not depend much on the number of principal components chosen. This corroborate the earlier finding that the even a low number of principal components can influence many of the measurements. However, from Figure 5, we see that PC-EKF with the lower number of principal components has poor predictive performance. Specifically, we see that by using PC-EKF we introduce a bias in the formulation eventhough we reduce the variance —as in the typical bias-variance trade-off of the models. This bias can be quantified by the slope of the line fit between the predicted values and the actual values. We see that as the chosen number of principal decreases the bias increases. Specifically, the slope of the line with 20 principal components is 0.950 compared to the slope of the line with 80 principal components which is 0.977. Thus chosen the right number of principal components while applying PC-EKF is crucial so that the net benefit —from reducing variance but introducing bias— compared to OD-EKF is positive.

**CONCLUSIONS**

This study presented a generic approach to make the complete online calibration of DTA systems scalable. Specifically, we presented a scalable version of the problem formulation which incorporated parameters from both supply models and demand models. To demonstrate the approach, we adopted the calibration of OD-flows as the problem of interest and principal component analysis as the method to reduce the dimensionality of the parameter vector. The proposed state-space formulation was solved using the Constrained Extended Kalman Filter which can incorporate the non-linear relationships between parameters and the corresponding measurements. Finally, we demonstrated the approach on a real world transportation network of Singapore. We assessed the
FIGURE 4: Comparison of estimation PC–EKF procedure with different principal components through the scatter plots of 5 minute estimated vs. actual sensor counts on all the 5 test days. The equations of the line-fit are depicted on the plots. The darker the cell, the higher the number of points in it.

Performance of the proposed approach rigorously which lead to following insights

- Few principal components explain very high percentage of the variance. Specifically, 80 principal components explain 95% of the variance (dimensionality reduction by 52 times) and only 150 principal components explain around 99% of the variance (dimensionality reduction by 27 times).

- The proposed approach (PC-EKF) makes the application of Extended Kalman Filter feasible for application in real-time. The traditional approach (OD-EKF) is intractable for real-time applications even with present computational resources.
In state estimation, PC-EKF—with 80 principal components—outperformed OD-EKF by 13% in RMSNs and 7% in MAPEs. In state predictions, PC-EKF—with 80 principal components—outperformed OD-EKF across the three steps on average by 11% in RMSNs and 8.5% in MAPEs. The better overall performance of PC-EKF can be explained as follows: (i) it lends itself to better formulation through more accurate transition equations and error covariances; (ii) it reduces the errors in the numerical approximation of Jacobian in EKF leading to better performance of the algorithm; and, (iii) it tends to reduce the "variance" of the model leading to better overall performance.

Sensitivity analysis of principal components lead to following insights. First, dimension-
ality reduction introduces bias into the model which can be beneficial if the variance of the base model is too high. However, choosing very few principal components can bias the model too much leading to poor predictive performance.

ACKNOWLEDGMENT
This research is supported by the National Research Foundation, Prime Minister’s Office, Singapore, under its CREATE program, Singapore-MIT Alliance for Research and Technology (SMART) Future Urban Mobility (FM) IRG.
REFERENCES


