Methodology for efficient real time OD demand estimation on large scale networks

Tamara Djukic
Delft University of Technology
Department of Transport and Planning
Stevinweg 1, 2628 CN Delft, The Netherlands
+31 15 278 1723
t.djukic@tudelft.nl
(Corresponding author)

Hans van Lint
Delft University of Technology
Department of Transport and Planning
Stevinweg 1, 2628 CN Delft, The Netherlands
j.w.c.vanLint@tudelft.nl

Serge P. Hoogendoorn
Delft University of Technology
Department of Transport and Planning
Stevinweg 1, 2628 CN Delft, The Netherlands
s.p.hoogendoorn@tudelft.nl

November 15, 2013

Word count:
Number of words in abstract 174
Number of words in text (including abstract) 5221
Number of figures and tables 8 * 250 = 2000
Total 7395

Submitted to the 93rd Annual Meeting of the Transportation Research Board, 12-16 January 2014, Washington D.C.
ABSTRACT
In previous work, we have explored the idea of dimensionality reduction and approximation of OD demand based on principal component analysis (PCA). In particular, we have shown how we can apply PCA to linearly transform the high dimensional OD matrices into the lower dimensional space without significant loss of accuracy. Next, we have defined a new transformed set of variables (demand principal components) that is used to represent the OD demand in lower dimensional space. These new variables are defined as state variable in a novel reduced state space model for real time estimation of OD demand. In this paper, we review previous work and continue this line of research. Based on the previous results, we demonstrate the quality improvement of OD estimates using this new formulation and a so-called, ‘colored’ Kalman filter approach for OD estimation, in which correlated observation noise is accounted. Moreover, we provide a thorough analysis of the model performance and computational efficiency using real data from a large network, and method for obtaining a reduced set of state variables.
INTRODUCTION

This paper focuses on the efficient estimation of OD matrices for large scale networks, since they will be used for on-line applications such as dynamic traffic management. Much of the work in OD matrix estimation and prediction so far has focused on improving estimation and prediction of OD matrices with more sophisticated and less time consuming algorithms (1), (2), (3) and by including additional data, ranging from traffic counts to automatic identification data (4), (5), (6), or data from Bluetooth devices (7), to name a few. Lately, decomposition of network into smaller subareas has been proposed by (8), (9) to deal with high dimensional OD estimation problem. A convenient way of understanding the complexity of the OD estimation problem and proposed solution is to illustrate both methodologies in the generic way.

In general terms, all dynamic OD demand estimation and prediction methods aim to find the most probable OD matrix $X_k$, given previous estimates $X_{k-n}$, $n = 1, 2, ..., $ historical OD matrices $X_{prior}$, the available (sensor) data $Y$ and all the other assumptions $H$ related to for example the assignment method and/or the assumed temporal evolution of the OD patterns. The common methodology with inputs and outputs into an OD matrix estimation and prediction is illustrated in Figure 1. This generic methodology can be used for off-line and on-line applications. The most widely used sensor data are link traffic counts $y$ that would be available for the entire analysis period (all the departure intervals) or at the end of each interval $k$. For off-line application, the entire set of link traffic counts for the analysis period would be used to simultaneously estimate OD matrices for all time intervals. For on-line application, at the end of each interval $k$, only the counts corresponding up to $k^{th}$ time interval would be used to sequentially estimate OD matrix for current time interval. Finally, for the on line application, predictions of OD matrices are generated for intervals $k + 1, k + 2, ....k + T$.

![FIGURE 1 Overview of common OD estimation (and prediction) methodology]

The OD estimation problem is computationally intensive of the complexity of the estimation and prediction methods and the fact that time dependent OD matrices for real-life transport networks typically constitute high dimensional data structures. One of the problems with high-dimensional datasets is that, in many cases, not all the measured variables are "important" for understanding the underlying phenomena of interest. In other words, these high-dimensional data may consist of multiple, indirect measurements of an underlying source. Therefore, one possible solution approach to solve this "curse of dimensionality" is to map the high dimensional OD matrices into a space of lower dimensionality, such that most of the structural information about the demand is preserved.

Figure 2 illustrates the efficient methodology for on-line OD demand estimation and prediction. This method, first applies the Principal Component Analysis (10) or PCA, to any data set of historical OD flows or generated from detailed demand microsimulation system. This historical OD dataset can be represented, without loss of generality, as a linear combination of a set of only a few orthonormal vectors (eigenvectors) and principal demand components. We extract off-line these eigenvectors that capture the...
trip-making patterns and their spatial and temporal variations, whereas the principal demand components capture the contribution of each eigenvector to the realization of a particular OD flow. These principal demand components are used as state variables instead of the OD flows themselves. For the on-line application, at the end of each interval $k$, the traffic counts corresponding up to $k^{th}$ time interval would be used to sequentially update principal demand components for current time interval. Finally, the estimated principal demand components are used to obtain the estimates of OD matrix.

Reducing the problem dimensionality through PCA replaces the usual approach of using prior OD matrices by structural information obtained either from data or from a detailed demand microsimulation system. The importance and originality of this approach lie in the possibility to capture the most important structural information without loss of accuracy and considerably decreasing the model dimensionality and computational complexity.

The paper is organized as follows. In the first part of the paper, we summarize the idea of dimensionality reduction and approximation of OD demand based on PCA. In the second part of the paper, we present the state space formulation of the OD estimation model with principal demand components as the state variables. Next, we analytically explore the properties of the colored noise Kalman filter to solve the proposed OD estimation method with time-correlated measurements. In the third part of paper, we demonstrate the performance of the proposed OD estimation model on a large-scale network (Vitoria, Spain). The paper closes with a discussion on further application perspectives of the OD demand estimation model and further research directions.

**FIGURE 2** Overview of proposed OD estimation (and prediction) methodology

**REDEFINING THE STATE VARIABLES FOR OD ESTIMATION**

**The idea of dimensionality reduction**

Since OD matrices are high dimensional multivariate data structures, the specification and estimation of OD matrices is both methodologically and computationally cumbersome for real time applications. There are three factors that increase the computational effort: the size of the state vector, the complexity of model components (e.g. assignment matrix, covariance matrices, etc.), and the number of measurements to be processed. For example, the Kalman filter algorithm is commonly used method to estimate and predict the OD matrices $(11), (12), (13)$. Since the computational complexity of the Kalman filter is typically in the order of $O(n_{ij}^3)$, where in the simplest case $n_{ij}$ is the total number of the OD pairs in the network, this can
represent a potential computational bottleneck. In addition, each traveler takes a certain time to complete his/her trip in large scale networks, and the resulting travel time can be very long depending on trip length between OD pairs and prevailing traffic conditions (i.e. congestion level in the network). The effect of time lag indicates that the traffic flow at the current observation time interval can include OD flows departing from previous time interval, leading to an enormous computational strain. For example, if we assume that the number of lagged time intervals is s, the size of the state vector increase for at least \((s \times n_{ij})\), and manipulation of vectors and corresponding covariances becomes cumbersome. Lately, approximation of OD flows to handle the lagged OD flows has been proposed by (14). The approximation is based on the conjecture that much of the information about an OD flow is likely to be provided the first time it is counted. If this were true, OD flows corresponding to prior departure intervals could be held constant at their prior estimated values and only the flows for the current departure interval need to be estimated. Alternatively, a polynomial trend model proposed by (3) can offer a compact representation of lagged demands. However, the polynomial trend model is still very computationally intensive for large scale networks.

Clearly, reducing the dimensionality of the state vector, is a way to improve computational efficiency. For example, let us assume that OD flows have been estimated for several previous days or months. These flows subsume in them various kinds of information, about trip making patterns and their spatial and temporal variations. Therefore, the key idea in our approach is to reduce the dimensionality of the OD matrix, in such way that the structural and temporal patterns are preserved. With this approach the computational cost can be speeded up dramatically, without significant lost of accuracy. One commonly used method of dimensionality reduction is a linear transformation technique known as Principal Component Analysis (PCA). PCA has found application in traffic and transportation science before, for example for the dimensionality reduction in calibration of travel demand from traffic counts (15).

In the remainder of this section, we will explain how we derived series of OD matrices and how we organized these data in appropriate form for PCA method. Finally, we define the new set of state variables for OD estimation.

The dimensionality reduction based on PCA

In the previous section, we have discussed the problems that can arise in attempts to estimate and predict OD matrices in high dimensional spaces, and the potential improvements which can be achieved by first transforming data into a space of lower dimensionality. In this section we demonstrate the remarkable fact that any OD matrix has a concise representation when expressed in terms of an orthonormal basis of \((n_{ij} \times 1)\) vectors \(e_i, i = 1, 2, ..., n_{ij}\) that can be derived using PCA.

Our goal is to map vectors of the OD demand \(X \in \mathbb{R}^{n_{ij}}\) onto the new vector in an \(n_m\)-dimensional space, where \(n_m < n_{ij}\). To illustrate dimensionality reduction based on PCA we follow the same rational as in (10).

Let \(X\) be the data matrix set defined such that one dimension in data represents the dimension in which we are seeking to find structure (e.g. columns of the matrix \(X_j\)), and the other dimension represents the dimension in which realizations of this structure are stored (e.g. rows of the matrix \(X_i\)). For example, suppose that we have used a microsimulation-based demand model to generate a large sample of OD demand observations \(r\) (e.g. observations can represent a daily OD demand, or OD demand per departure time interval) in a network, each being a realization of the \(n_{ij}\)-dimensional OD demand vector \(x_r = (x_1, x_2, ..., x_{n_{ij}})\). Thus, we have a \((r \times n_{ij})\) OD demand matrix \(X\), where each row \(i, i = 1, 2, ..., r\) contains the vector of OD flows per time (e.g. for whole day, per departure time interval) and column \(j, j = 1, 2, ..., n_{ij}\) denotes the realizations of the \(n_{ij}\)-th OD pair over time, in the following form:
Once we generate the matrix $X$, we apply off-line the PCA algorithm to extract the eigenvectors $e_i, i = 1, 2, ..., n_{ij}$ and eigenvalues $\lambda_i, i = 1, 2, ..., n_{ij}$. In practice, the PCA algorithm proceeds by first computing the mean of the vectors $x_r$ and then subtracting this mean. The new centered data set is computed as

$$X = X - \bar{X}$$

$$\bar{X} = \frac{1}{r} \sum_{i=1}^{r} X_i$$

(1)

Then the covariance matrix $S$ of the set of vectors $x_r$ given by $\sum_r (x_r - \bar{x})(x_r - \bar{x})^T$ is calculated and its eigenvectors and eigenvalues are found, i.e.

$$S e_i = \lambda_i e_i$$

(2)

where $e_i$ for $i = 1, 2, ..., n_{ij}$ are the eigenvectors and $\lambda_i$ for $i = 1, 2, ..., n_{ij}$ are the eigenvalues of the covariance matrix $S$.

Since the covariance matrix of $X$ is real and symmetric, its eigenvectors $e_1, e_2, ..., e_{n_{ij}}$ can be chosen as an orthonormal basis. Therefore, any vector $x$, or actually $(x - \bar{x})$, can be represented, without loss of generality, as a linear combination of a set of $n_{ij}$ orthonormal vectors $e_i$

$$x - \bar{x} = c_1 e_1 + c_2 e_2 + ... + c_{n_{ij}} e_{n_{ij}} = \sum_{i=1}^{n_{ij}} c_i e_i$$

(3)

where the vectors $e_i$ satisfy the orthonormality relation

$$e_i^T e_j = \delta_{ij}$$

(4)

in which $\delta_{ij}$ is the Kronecker delta symbol. Explicit expressions for the coefficients $c_i$ in (3) can be found by using (4) to give

$$c_i = e_i^T (x - \bar{x})$$

(5)

which can be regarded as a simple rotation of the coordinate system from original $x$’s to a new set of coordinates given by $e$’s. Through sorting the eigenvectors in decreased order by the size of the eigenvalue, we can retain the first $n_m$, where $n_m \leq n_{ij}$, eigenvectors which captures the maximum data variance. However, since the covariance matrix of observed OD demand in general can be very large, it is inconvenient to evaluate and store it explicitly. To avoid this we can use efficient algorithms, which find the $n_m$ largest eigenvectors of the covariance matrix such as the orthogonal iteration and power method (16). An intuitive explanation of (5) is that the eigenvectors are used as weights on each of the original variables to compute the principal demand components.

Once the $n_m$ largest eigenvectors $e_1, e_2, ..., e_{n_m}$ are found, a new low dimensional representation of the OD demand can be expressed as follows

$$(\hat{x} - \bar{x}) = \sum_{i=1}^{n_m} c_i e_i$$

(6)
where $\hat{X}$ is the approximated OD demand constructed using the first $n_m$ eigenvectors. The representation of $(X - \hat{X})$ on the orthonormal basis $e_1, e_2, ..., e_{n_m}$ is thus given by principal demand components $c_1, c_2, ..., c_{n_m}$. Thus, we define a new set of variables, principal demand components $c_i$ that capture the contribution of each eigenvector $e_i$ to the particular observations of OD demand. In turn, the eigenvectors $e_i$ capture the common behavior of travelers over the all OD pairs. The eigenvectors then define the fixed structure of our OD matrices, which we then update on-line from traffic counts. Thus, the explicit expression for the approximated OD demand $\hat{X}$ in lower dimensional space can be found by using equation (6) to give

$$\hat{x} = \sum_{i=1}^{n_m} c_i e_i + \bar{x}$$

(7)

In the next section we formulate a state space OD estimation model, where the eigenvectors $e_i$ define the fixed structure of our OD matrices and principal demand components are updated on-line from traffic counts.

A REDUCED STATE SPACE OD ESTIMATION MODEL FORMULATION

In this section we demonstrate how the approximated OD demand presented in previous section can be viewed and defined as a state-space based formulation.

Following the idea presented in previous section, we define our state to be a $(n_m \times 1)$ vector of principal demand components, $c_k$, where $n_m$ represent the reduced number of variables in state vector. The principal demand components represent the approximated OD demand, where each principal demand component $c_i$, for $i = 1, 2, ..., n_m$ captures the contribution of each eigenvector $e_i$ to the particular observations of OD demand. Therefore, the OD demand state in the network at time $k$ is uniquely described by the vector of the principal demand components $c_k$ in $n_m$-dimensional space, where $n_m < n_{ij}$.

The state space model formulation consists of the process and observation equations. Clearly, we have to specify the process equation that captures the temporal evolution of the state, and observation equation that uses whatever new information (i.e. observation) is available to estimate the state.

The process equation is based on the autoregressive process on the principal demand components, which provides preliminary estimate of the OD flow. We define the process equation as follows:

$$c_k = \sum_{p=k-q}^{k-1} \phi_p^p c_p + \omega_k$$

(8)

where $\phi_p^p$, a $(n_m \times n_m)$ is the process matrix that represents the effects of previous states $c_p$ on current state $c_k$, $q'$ is a degree of the autoregressive process and $\omega_k$ is a vector of a random variables capturing the unobserved deviations in process. The process noise vector $\omega_k$ depicts some known Gaussian noise term defined with following assumptions:

- mean $E[\omega_k] = 0$;
- variance $E[\omega_k^2] = \theta_k \delta_k$, where $\theta_k$ is a $n_m \times n_m$ variance covariance matrix, with eigenvalues on the diagonal stored in decreasing order, and the $\delta_k$ is the Kronecker symbol.

The state-space model formulation furthermore uses an observation equations. We define the observation equation as a linear relationship between the state variables (principal demand components) and the observations (traffic counts):

$$y_k = \sum_{p=k-q'}^{k} A_p^p x_p + v_k$$

(9)

where $y_k \in \mathbb{R}_{n_l}$ denotes a vector of link traffic counts for time interval $k$, and $A_p^p$ is a $(n_l \times n_{ij})$ matrix, known as assignment matrix, mapping OD flows departing during intervals $p$ to link traffic counts observed.
during interval $k$. Further, $p'$ is the maximum number of time intervals needed to travel between any OD pair, and $\upsilon_k$ is a vector of random variables capturing the observations error on detectors during interval $k$.

Following the lower dimensional representation of OD demand by principal demand components and substituting (6) in (9), we can reformulate the observation equation (9) as:

$$y_k = \sum_{p=k-p'}^{k} A^p_k (e_p e_p + \bar{x}_p) + \upsilon_k$$

$$= \sum_{p=k-p'}^{k} H^p_k c_p + \bar{y}_p + \upsilon_k$$

(10)

where $H^p_k = \sum_{p=k-p'}^{k} A^p_k e_p$ is a $(n_l \times n_m)$ matrix called observation matrix, mapping the principal demand components during intervals $p$ to traffic counts observed during interval $k$. Note that the observation matrix $H^p_k$ in equation (10) is not the same as the assignment matrix $A^p_k$ given in (9). Finally, the matrix $H^p_k$ is used for the linearization of the model; it equals the transform of the assignment matrix $A^p_k$ to the orthonormal basis matrix of eigenvectors $e_p$. The observation noise $\upsilon_k$ depicts some known Gaussian noise term defined with following assumptions:

- mean $E[\upsilon_k] = 0$;
- variance $E[\upsilon^2_k] = R_k \delta_{km}$, where $R_k$ is a $(n_l \times n_l)$ variance covariance matrix, and the $\delta_{km}$ is the Kronecker symbol.

In conclusion, we might mention that this model uses following input variables: process transition matrix $\phi^p_k$, process error covariance matrix $\theta_k$, observation error covariance matrix $R_k$, and assignment matrix $A^p_k$. These input data are usually derived from, for example, existing historical data on OD demand and observations. In transport modeling for the real time applications, it is considered that data would be available over multiple days, and hence, we would be able to calibrate model inputs. We revisit this issue later in case study section.

**SOLUTION APPROACH: ESTIMATION AND PREDICTION**

It is convenient to start the presentation of the solution approach with reference to the idea of variables reduction in state vector and state-space model given in previous section. Then, we provide a solution approach when correlated observation noise is accounted due to reduction of variables in state vector.

Temporal correlation between observations introduced by dimensionality reduction

Equations (8) and (10) constitute a discrete time liner Kalman Filter. The solution approach of such a system of equations may seem fairly standard at first glance. However, since there are practical points which are not entirely obvious, we illustrate them here before presenting a solution algorithm. Reducing the state variables introduces additional uncertainty in the process, and this noise increases as the reduced number of state variables increases. In order to explain the potential reasons of the temporal correlation between observations introduced by the dimensionality reduction of the state vector, we analytically derive the observation noise correlation. Here, we omit the effect of lagged time intervals $p$ in observation equation (10) for the sake of simplicity.

The given observation equation (10) for reduced number of state variables $n_m$ over time interval $k$ can be expressed as
\[
y_k = A_k \sum_{i=1}^{n_m} c_{i,k} e_{i,k} + A_k \sum_{i=n_m+1}^{n_{ij}} c_{i,k} e_{i,k} + v_k
\]
\[
y_{k+1} = A_k \sum_{i=1}^{n_m} c_{i,k+1} e_{i,k+1} + v_{k+1}
\]
\[
\xi_k = A_k \sum_{i=n_m+1}^{n_{ij}} c_{i,k} e_{i,k} + v_k
\]
\[
\xi_{k+1} = A_{k+1} \sum_{i=n_m+1}^{n_{ij}} (c_{i,k} + \omega_k) e_{i,k+1} + v_{k+1}
\]
\[
\xi_{k+1} = A_{k+1} \sum_{i=n_m+1}^{n_{ij}} \omega_k e_{i,k+1} + v_{k+1}
\]

where, \(\xi_k\) represent the observation noise that consists of additional noise introduced by dropped state variables from \(n_m + 1\) till \(n_{ij}\) at time interval \(k\).

Further, observation equation (10) for reduced number of state variables \(n_m\) for the next time interval \(k + 1\) can be expressed as

\[
y_{k+1} = A_{k+1} \sum_{i=1}^{n_m} c_{i,k+1} e_{i,k+1} + A_{k+1} \sum_{i=n_m+1}^{n_{ij}} c_{i,k+1} e_{i,k+1} + v_{k+1}
\]
\[
\xi_{k+1} = A_{k+1} \sum_{i=n_m+1}^{n_{ij}} (c_{i,k} + \omega_k) e_{i,k+1} + v_{k+1}
\]

where, \(\xi_{k+1}\) represent the observation noise at time interval \(k + 1\) that consists of additional noise introduced by omitted state variables from \(n_m + 1\) till \(n_{ij}\) in previous time interval \(k\). Therefore, \(\xi_k\) and \(\xi_{k+1}\) represent the temporal correlated observation noise. It is well known that this condition destroys the assumption of independency between process and observation noise that underlies the standard Kalman filter. The objective of this section is to find an effective method to deal with this kind of correlation.

**Colored noise Kalman filter solution algorithm**

When the observation errors are temporally correlated, as we show in previous subsection, the time differencing approach, which was first introduced in 1968 by Bryson and Henrikson (17) is commonly applied as a way to model correlated observation noise in state-space model representation. The core idea behind this filter is the elimination of the time-correlated observation noise terms \(\xi_k\) using a pseudo-observation equation \(z_k\) whose error is white and is given by

\[
z_k = y_{k+1} - \Psi y_k
\]
\[
= (H_{k+1}\Phi_k - \Psi_k H_{k+1}) c_k + H_k \omega_k + v_k
\]
\[
= H_k^* c_k + v_k^*
\]

where correlation matrix \(\Psi\) is equivalent to the process transition matrix \(\Phi_k\) for time correlated errors, and \(v_k\) is a observation noise vector assumed to be uncorrelated with the process noise vector \(\omega_k\), and \(H_k^* = H_k\phi_k - \psi_k H_k\). The new observation noise is given as

\[
v_k^* = H_{k+1}\omega_k + v_k
\]
with mean $E[\nu_k^*] = 0$ and covariance matrix $R_k^*$.

Further, the decorrelation technique from [11] is applied on process equation (11) to eliminate the correlation that now exists between the new observation noise $\nu_k^*$ (14) and the process noise $\omega_k$. A new process equation can be written as

$$
c_k = \phi_{k-1}c_{k-1} + \omega_{k-1} + J_{k-1}(z_{k-1} - H_k^*c_{k-1} - \nu_{k-1}^*)
= \phi_{k-1}^*c_{k-1} + J_{k-1}z_{k-1} + \omega_{k-1}^*
$$

(15)

where the new state process matrix is expressed as $\phi_{k-1}^* = \phi_{k-1} - J_{k-1}H_k^*$. Now, the new process noise error is defined as

$$
\omega_{k-1}^* = \omega_{k-1} - J_{k-1}\nu_{k-1}^*
$$

(16)

with mean $E[\omega_k^*] = 0$ and covariance matrix $\Theta_k^*$.

At this time, for the given problem we have a state space model depicted by equations (13) and (15) which satisfy the assumptions of standard Kalman filter. Clearly, the new process noise $\omega_k^*$ and observation noise $\nu_k^*$ are independent, zero-mean, Gaussian noise processes of covariance matrices $\Theta_k^*$ and $R_k^*$ respectively. Algorithm 1 summarizes the colored Kalman Filter equations as a solution of such a system:

**Algorithm 1: The colored Kalman Filter**

**Initialization:**

$$
\hat{e}_{0|0} = E[e_{0|0}] \text{ and } P_{0|0} = E[e_{0|0} - E[e_{0|0}]^T]
$$

In case no additional information is available, $P_{0|0}$ is usually initialized as a matrix with a large diagonal entries, reflecting the fact that we are highly uncertain about our initial estimate of $\hat{e}_{0|0}$.

**For** $k = 1, 2, \ldots$ **do:**

**Compute the Kalman Gain:**

$$
K_k = P_{k|k-1}H_k^T(H_k^*P_{k|k-1}H_k^* + R_k^*)^{-1}
$$

(17)

**Correct mean and covariance:**

$$
c_{k-1|k} = c_{k-1|k-1} + K_k(z_k^* - H_k^*c_{k-1|k-1})
$$

(18)

$$
P_{k-1|k} = (I - K_kH_k^*)P_{k-1|k-1}(I - K_kH_k^*)^T + K_kR_k^*K_k^T
$$

(19)

**Update mean and variance of state variables:**

$$
c_{k|k} = \phi_k^*c_{k-1|k} + J_kz_k
$$

(20)

$$
P_{k|k} = \phi_k^*P_{k-1|k}\phi_k^T + \Theta_k^*
$$

(21)

**End**

Note that the time differencing solution algorithm uses a 1 time interval latency in the observation updating because the observation in time interval $k$ has to be used to update the state vector in previous time interval, $k - 1$. Therefore, following the Kalman filter terminology, $c_{k-1|k}$ denotes correction of the state variable for time interval $k - 1$, using the information from link traffic counts for interval $k$, and $P_{k-1|k}$ depict updated state error covariance matrix. The Kalman filter gain in equation (17) evaluates the importance of the new information obtained from link traffic counts at time interval $k$ and it can be interpreted as the weight given to the latest information. The equations (18) and (19) reflect the our corrected knowledge on the system state at time interval $k - 1$ with obtaining the link traffic counts for interval $k$. In the update step, the our knowledge on evolution of state and observations is used to update prior correction. Therefore,
the equations (20) and (21) reflect the our estimate (best knowledge) on the system state $c_{k|k}$ and $P_{k|k}$ error covariance matrix at time interval $k$ including the information on link traffic counts for time interval $k$.

Finally, the result of the colored Kalman filter, the estimated a posterior state vector $c_{k|k}$, is used to estimate the OD demand by applying equation (7). For a more detailed derivation of colored noise Kalman Filter, and derivation of covariance matrices $\Theta^*$, $R^*$ and $M^*$ we refer to (18) and (19).

CASE STUDY

In this section we will first describe the input data used by method, e.g. historical OD demand generation and state variables reduction procedure. We consider two assessment scenarios in terms of number of variables in state vector (i.e. with reduction of state variables and without reduction). These scenarios will be discussed in more detail below. Numerical experiments are performed on large-scale network, (Vitoria, Basque Country, Spain) with real data to evaluate the performance of the proposed model and solution algorithm.

Network topology

Prior to method evaluation, we define a Vitoria network that consists of 57 centroids, 3249 OD pairs with a 600km road network, 2800 intersections and 389 detectors presented with black dots in Figure 3. This network is available in the mesoscopic version of the Aimsun (20) traffic simulation model for reproducing the traffic propagation over the network. The true OD demand is available for this network, which allows analyst to assess the performance of proposed model. The true assignment matrix and traffic counts on detectors are derived from assignment of true OD matrix in Aimsun for evening period from 19:00 to 20:00 reflecting the congested state at the network. The simulation period is divided in 15 minutes time intervals with additional warm-up time interval, $T = 5$. The link flows resulting from the assignment of the true OD demand are used to obtain the traffic count data per observation time interval. The trips between some of the OD pairs are not completed within one time interval due to congestion on network or the distance between OD pairs. In this way a vehicle entering the network in a particular departure time interval needs more than one time interval to reach a traffic detector where the departure time interval and detection time are different. In our study network, the maximum travel time between OD pairs observed on network takes four time intervals, which leads to very sparse assignment matrices, and the number of lagged time intervals $p' = 4$.

![Figure 3 The Vitoria network, Basque Country, Spain](image)
Simulating historical daily OD demand

A major problem with all method assessments is obtaining meaningful evaluations of the algorithms results and performance, because the true sources of data are not available for comparison when working with real data. One solution is to use simulated OD demand data, where underlying sources and phenomena are known. To generate a simulated OD demand per departure time interval dataset for our case study requires us to define an arbitrary model for OD demand generation, which represents a common spatial and temporal behavior of travelers.

Here, we perform the Logit model in sequence in order to introduce the correlation in OD demand data. First, we defined the set of traveler’s decisions before making a trip, including decisions to make a trip or not, destination choice and departure time choice. Then, for each of these decisions we have defined the set of alternatives available to travelers. The activity and traveling intentions of traveler $t_r$ are presented in the Figure 4. The main principle of this model is that a large number of simulations are performed for varying model inputs, reflecting the variability’s in the travelers behavior and consequently in OD demand based on Monte Carlo simulations.

The total number of trips per origin from available true OD matrix is assumed as an initial number of travelers per origin in simulations. Subsequently, we generate 10000 observations, each being a realization of the 3249-dimensional OD demand vector, where the total number of travelers per origin is equal to true OD matrix while their distribution over destinations is varied. Each generated OD demand vector per departure time interval is stored in OD demand matrix where each row represents one observation of OD demand, as we defined in the definition of state variables section.

The state vector reduction

To examine the effect of reducing the number of principal demand components in state vector, we applied the PCA on the OD demand data matrix $X_k$ over $k = 5$ departure time intervals. Once we perform the PCA, we obtain the set of eigenvectors $e_{i,k}$ for $i = 1, 2, \ldots, 3249$ and eigenvalues $\lambda_{i,k}$ for $i = 1, 2, \ldots, 3249$ per time interval $k$.

We have seen in previous sections that we can use eigenvalues to explore the data reduction potential, for instance by considering the total (cumulative) percentage of total variation explained (e.g. 95%), Figure 5. We can observe that the 90% of the variance of the data is captured by first 50 eigenvectors out of 3249. This result indicates that we can reduce the state vector by more then 90% and still capture the temporal and spatial variance in data.
We have performed the PCA on OD demand data set per time interval, such that in every time interval we can identify potential number of variables in state vector that describe the 95% of variance in data set. In Table 1, we show the number of state variables \( n_m \) that describe the 95% of variance in data set per departure time interval.

<table>
<thead>
<tr>
<th>Departure time interval</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of state variables</td>
<td>40</td>
<td>61</td>
<td>37</td>
<td>39</td>
<td>41</td>
</tr>
</tbody>
</table>

It clearly appears from Table 1 that we have obtained different values of variables in state vector over time intervals. Therefore, we define the number of state variables as \( m = \max(n_{m_k}) \), for \( k = 1, 2, \ldots, T \), since omitting the principal demand components with highest captured variance in OD demand will lead to non effective dimensionality reduction of the state vector. In the next subsection we will compare the performance of the colored Kalman filter for reduced variables in state vector (for \( m = 61 \)) and no reduced variables in state vector.

**Method performance**

We have performed the experiments for a Vitoria network, Spain, given in Figure 3, for following two scenarios:

- Case 1: in this experiment run, we omit the state variables (principal demand components) from the state vector. Since the principal demand components in the state vector are arranged in decreasing order of an eigenvalues, we remove the principal demand components that capture the lowest variance and keep first \( m = 61 \) state variables;
- Case 2: in this experiment run we keep all state variables (principal demand components) in the state vector, such that \( m = n_{ij} = 3249 \).
In Table 2, we represent: (1) the root mean square error (RMSE) per departure time interval and (2) mean absolute error (MAE) per departure time interval, that is

\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\bar{x}_i - x_i)} \]  

\[ MAE = \frac{1}{n} \sum_{i=1}^{n} |\bar{x}_i - x_i| \]

for \( \bar{x}_k \) estimated OD demand per time interval depending on the number of the variables in state vector, \( n \).

It appears clearly from Table 2 that reducing the variables in state vector yield overestimation of OD demand. However, we can observe that reducing the dimensionality of the state vector by more than 90%, the colored Kalman filter produces a reasonable reduction in accuracy. In real time applications it is always a question of trade-off between the computational efficiency and result’s accuracy. Therefore, it is of interest to examine the optimal number of variables in a state vector, such that the lower bound is define as a minimum number of variables that capture the 95% of the variance in data set, while the upper bound is given by the computation time preferences. In addition, the larger errors relate to the observability problem introduced by state variables reduction. Under conditions of non-observability, the effect of the initial estimates do not disappear with time and therefore it is critical to obtain accurate results or initial values. Therefore, the state identifiablity must be taken into account in the optimal number of state variables computation to achieve the Kalman filter convergence.

Table 3 reports that significant CPU computation time reduction can be achieved by the reduction of state variables. These times have been obtained by running MATLAB on Dell with Intel Xeon, Quad Core processor, 8GB (1600mHz) memory.

<table>
<thead>
<tr>
<th>Time interval</th>
<th>Case 2: No reduction of stat.var</th>
<th>Case 1: Reduction of stat.var</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE ( t_1 )</td>
<td>6.803</td>
<td>14.674</td>
</tr>
<tr>
<td>RMSE ( t_2 )</td>
<td>0.537</td>
<td>9.774</td>
</tr>
<tr>
<td>RMSE ( t_3 )</td>
<td>0.493</td>
<td>10.026</td>
</tr>
<tr>
<td>RMSE ( t_4 )</td>
<td>0.493</td>
<td>9.296</td>
</tr>
<tr>
<td>RMSE ( t_5 )</td>
<td>0.527</td>
<td>9.553</td>
</tr>
<tr>
<td>MAE ( t_1 )</td>
<td>3.039</td>
<td>8.186</td>
</tr>
<tr>
<td>MAE ( t_2 )</td>
<td>0.288</td>
<td>4.330</td>
</tr>
<tr>
<td>MAE ( t_3 )</td>
<td>0.243</td>
<td>4.501</td>
</tr>
<tr>
<td>MAE ( t_4 )</td>
<td>0.233</td>
<td>4.268</td>
</tr>
<tr>
<td>MAE ( t_5 )</td>
<td>0.278</td>
<td>4.409</td>
</tr>
</tbody>
</table>

Note that initial idea in our work is to solve the computational complexity of the OD estimation problem for real time applications. Therefore, in Table 3 we show the run time of colored Kalman filter for each scenario (e.g. no reduction of state variables in state vector and reduced number of state variables) on Vitoria network.
CONCLUSIONS
From the results presented in this contribution we can conclude that PCA can be used to linearly transform high dimensional OD matrices into the lower dimensional space without significant loss of estimation accuracy. We have proposed a new OD estimation method that uses the eigenvectors and principal demand components as state variables instead of OD flows. These variables can be used to construct a state space model that can be solved with recursive solution approaches such as the Kalman filter.

The proposed state space model, however, appears to be sensitive to the reduction of the dimensionality due to the induced temporal measurement correlation. We have explored and derived an analytical solution for the so-called colored noise Kalman filter algorithm that accounts for temporal correlated measurement noise to avoid this limitation.

In this paper we show that reduction of state variables in proposed OD estimation model for large-scale networks will lead to computational efficiency with an acceptable degradation in result’s accuracy. An improvement of the algorithm presented in this paper can be seen in two directions: (1) definition of the optimal number of principal demand components in state vector such that the computational efficiency, results accuracy and state observability are satisfied, (2) adaptation of the model when additional data (i.e. speeds, density, travel times from different technological sources) can be considered to improve the quality of the estimated OD demand.

ACKNOWLEDGMENT
This research is partly funded by the ITS Edulab, a collaboration between TUDelft and Rijkswaterstaat. Also, this research is supported and partly funded by the EU COST Action TU0903 MULTITUDE Methods and tools for supporting the Use caLibration and validaTIon of Traffic simUlation moDEls project in collaboration with Delft University of Technology and KTH University.

REFERENCES


