

# Travel time forecasting from clustered time series via optimal fusion strategy

Andres Ladino, Alain Kibangou, Hassen Fourati, Carlos Canudas de Wit

► **To cite this version:**

Andres Ladino, Alain Kibangou, Hassen Fourati, Carlos Canudas de Wit. Travel time forecasting from clustered time series via optimal fusion strategy. 15th European Control Conference (ECC 2016), Jun 2016, Aalborg, Denmark. <10.1109/ECC.2016.7810623>. <hal-01296525>

**HAL Id: hal-01296525**

**<https://hal.archives-ouvertes.fr/hal-01296525>**

Submitted on 1 Apr 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Travel time forecasting from clustered time series via optimal fusion strategy

Andres Ladino<sup>1</sup>, Alain Kibangou, Hassen Fourati<sup>2</sup>, and Carlos Canudas de Wit<sup>3</sup>

**Abstract**—This paper addresses the problem of travel time forecasting within a highway. Several measurements are captured describing travel times for multiple origin-destination (OD) pairs. A network model is then proposed to infer travel time between origin and destination based on a reduced number of states. The forecast strategy is based on current day and historical data. Historical data is organized into several clusters. For each cluster, a predictor is designed based on the Kalman filtering strategy. Then these predictions are fused, in a best linear unbiased estimation sense, in order to get the best prediction. The performance of the proposed method is evaluated using traffic data from the South Ring of the Grenoble city in France.

## I. INTRODUCTION

Traffic forecasting is one of the most desired tools for traffic management requested by operators and commuters. In the era of data deluge in which we are, measurements collected by sensors are important sources of information that require analysis, classification, and processing in order to detect patterns and behaviours that can be exploited for traffic prediction [7], [14]. The collected information can be classified by clusterization algorithms, such as K-means, where each cluster collects traffic patterns which in some cases characterize typical regimes such as congestion. Several indicators like travel time, queue length, density, delay are used as performance indexes to determine the status of a traffic network [16].

In this paper we focus on models that rely on historical travel time information. As shown in [3], typical real patterns can be found in large data sets of historical data. For example [10] proposes a forecasting of travel time based on an Adaptive Kalman Filter (AKF) strategy in which observations are built from historical data sets of speed and flow. [11] applies a similar strategy in flow prediction and presents the problem of multi-step ahead forecasting based on clustered time series by applying several predictors such as Gaussian maximum likelihood (GML) and AKF. [17] presents an approach for short term flow forecasting using multiple ARMAX based predictors obtained from clustered data. The ARMAX model is adapted independently to different groups of flow time series and a single prediction is selected based on one criteria that considers minimum error estimation for the predicted signals. [15] uses Link Node Cell Transmission

Model calibrated via Monte Carlo methods in order to generate a prediction using the expectation maximization algorithm. Although all these methods present strategies for the selection of the predicted sequence, regimes described by clustered data are not totally separable and studies like [19] have shown the improvement of performance with combined forecasts.

Different models can be inferred from different clusters of data. A series of new methods have been emerging to combine information from these models. For instance [18] proposes a forecast based on interactive multiple models by combining different individual forecasting methods. [4] proposes an adaptive fusion method combining historical information and current day data. Other hybrid forecasts includes support vector machine methods like in [6] and model based approaches as in [2].

Generally, forecast algorithms in the literature are designed to satisfy a set of constraints given by the forecast problem. Most of the algorithms take into account availability of a full set of measurements for all possible locations and time instants within the traffic network and they overcome the problems of missing data, low penetration ratio or unbalanced spatial coverage by introducing additional steps such as imputation algorithms [5]. Moreover, the great majority of literature provides fixed forecasting scenarios in which predefined OD schemes are considered [3], [9]. Some recent approaches have emerged considering flexible OD, moreover the possibility to exploit internal relations of the network to reconstruct information [4]. In this paper we formulate the problem by considering prediction over all OD pairs of the highway. To reduce the dimensionality of the problem, we actually predict the travel time for the internal state of the network. In addition, since travel time measurements for all the OD pairs cannot be available all the time, we face a missing data problem. To overcome this issue, we resort to a data imputation based on a dictionary learning approach [8]. From the imputed data, a clusterization is achieved, defining then different clusters characterized by a centroid containing the mean of the data and a given dispersion around it. The evolution of the centroid can be used as future observation, herein called pseudo-observation, that can feed a Kalman filter. Therefore the prediction problem is solved as a filtering one. However, the main question is, how to associate the current day data to a specific cluster since we don't know its future? To solve this issue, we run Kalman filters for each cluster and then we make the fusion of the obtained forecasts. The benefits of this solution are: First, the interaction of main features in each cluster are captured when

<sup>1</sup>Andres Ladino is with NeCS team and GIPSA-Lab [andres.ladino-lopez@gipsa-lab.fr](mailto:andres.ladino-lopez@gipsa-lab.fr)

<sup>2</sup>Alain Y. Kibangou, Hassen Fourati are with NeCS team and Univ. Grenoble Alpes GIPSA-Lab [{alain.kibangou,hassen.fourati}@gipsa-lab.fr](mailto:{alain.kibangou,hassen.fourati}@gipsa-lab.fr)

<sup>3</sup>Carlos Canudas de Wit is NeCS team at CNRS, GIPSA-Lab Grenoble, France [carlos.canudas-de-wit@gipsa-lab.fr](mailto:carlos.canudas-de-wit@gipsa-lab.fr)

the fusion is performed, and the classification of the forecast is not always assigned to a single group a priori. Second, applying a kalman filter incorporates measured statistics in historical data into the forecast. Finally, under the appropriate considerations the model can be extended into other type of variables such as flow since the process model is mainly stochastic. The main contribution of this paper is to develop a fusion method controlled by the error covariance of the local Kalman filters. The performance of the proposed method is evaluated using real traffic data from the Grenoble Traffic Lab (GTL) [1].

The paper is organised as follows. Section II presents the travel time network model while Section III is devoted to the missing information and clustering problems. The travel time forecasting method is described in Section IV, evaluated in Section V with real data before concluding the paper in Section VI.

## II. NETWORK TRAVEL TIME MODEL

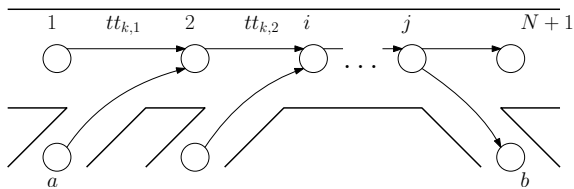


Fig. 1. Network architecture of a highway

Consider a highway constituted with a set of nodes  $\mathcal{N} = \{1, 2, \dots, N+1\}$  (See. Fig. 1). An origin is defined as a node having an on-ramp whereas a destination is a node with an off-ramp. We define by  $\mathcal{O} \subset \mathcal{N}$  and  $\mathcal{D} \subset \mathcal{N}$  the set of origins and destinations respectively. The highway can be represented as a directed path graph  $\mathcal{S} = (\mathcal{N}, \mathcal{E})$ . Each directed edge  $e \in \mathcal{E}$  connects two consecutive nodes. Therefore  $|\mathcal{E}| = N$ .

For a path starting from node  $i$  and ending at node  $j$ , we define the vector  $\mathbf{f}(i, j) \in \mathbb{R}^N$  whose  $q$ th entry equals 1 if the  $q$ th directed edge is crossed, else it is zero. As a consequence,  $\|\mathbf{f}(i, j)\|_0$  is equal to the number of edges between  $i$  and  $j$ . Let  $tt_{k,q}$  be the travel time experienced at time  $k$  for crossing to  $q$ th edge. Then for any OD pair  $(i, j)$  we can associate the instantaneous travel time  $TT_k(i, j)$  as:

$$TT_k(i, j) = \mathbf{f}^T(i, j) \mathbf{tt}_k, \quad (1)$$

where  $\mathbf{tt}_k = [tt_{k,1} \quad tt_{k,2} \quad \dots \quad tt_{k,N}]^T$ . By concatenating the travel times of all the OD pairs for a particular day  $d$  in the vector

$$\mathbf{TT}_{k,d} = \begin{bmatrix} TT_k(1,2) \\ \vdots \\ TT_k(1,N+1) \\ \vdots \\ TT_k(i,j) \\ \vdots \\ TT_k(N,N+1) \end{bmatrix}$$

$(i, j) \in \mathcal{O} \times \mathcal{D}$ , we get:

$$\mathbf{TT}_{k,d} = \mathbf{\Phi} \mathbf{tt}_k, \quad (2)$$

where  $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$  results on the concatenation of the vectors  $\mathbf{f}^T(i, j)$  associated with  $M$  possible OD pairs.

The travel time related to an OD pair can be measured using fixed sensors (magnetic sensors, bluetooth,...) and/or crowdsourcing. In all cases, due to network data issues, and availability of some measurements it is not possible to collect the travel times of all OD pairs each time. The purpose of this paper is to predict the travel time of any OD pair given historical and current measurements of some OD pairs.

## III. HISTORICAL DATA PROCESSING & FORECASTING PROBLEM FORMULATION

### A. A dictionary learning approach for data imputation

Given available information at time  $k$  we want to predict the travel time for the next  $\Delta_f$  time samples. For this purpose, we analyse available historical information in the interval of time  $[k - \Delta_p + 1, k + \Delta_f]$ . The size of the time window in the past to be considered is  $\Delta_p$ . It is standard to split historical information in order to detect typical patterns for different regimes [16]. Let define the matrix  $\mathbf{\Gamma} = [\mathbf{\Gamma}_1 \quad \mathbf{\Gamma}_2 \quad \dots \quad \mathbf{\Gamma}_D]$  of historical information:

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{TT}_{k-\Delta_p+1,1}^T & \mathbf{TT}_{k-\Delta_p+1,2}^T & \dots & \mathbf{TT}_{k-\Delta_p+1,D}^T \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{TT}_{k+\Delta_f,1}^T & \mathbf{TT}_{k+\Delta_f,2}^T & \dots & \mathbf{TT}_{k+\Delta_f,D}^T \end{pmatrix},$$

where we have assumed that, for the  $M$  OD pairs we have measurements for  $D$  days, meaning that  $\mathbf{\Gamma}$  has dimensions  $(\Delta_p + \Delta_f) \times MD$ . Assume that the travel time measurements associated with each OD pair can be clustered in  $Q$  clusters and define  $\mathbf{C} \in \mathbb{R}^{(\Delta_p + \Delta_f) \times MQ}$  the matrix of the centroids of these clusters. We can then notice that

$$\mathbf{\Gamma} = \mathbf{CS} + \mathbf{V},$$

where  $\mathbf{S}$  is a cluster selection matrix in which all columns entries are zero but one while  $\mathbf{V}$  stands for the dispersion around the centroid. However, since some of the measurements of all the OD pairs are not available at each time step,  $\mathbf{\Gamma}$  has several missing values. Let define by  $P_\Omega(\mathbf{\Gamma})$  the available information; specifically  $\Omega$  defines the set of  $(m, n)$  indices of  $\mathbf{\Gamma}$  for which the measurements are available. Therefore if  $(m, n) \notin \Omega$  the  $(m, n)$  entry of  $P_\Omega(\mathbf{\Gamma})$  is set to zero. We can solve the problem as a dictionary learning problem [12]. The problem is to estimate  $\mathbf{C}$  and  $\mathbf{S}$  from the available data:

$$\min_{\mathbf{C}, \mathbf{S}} \|P_\Omega(\mathbf{\Gamma} - \mathbf{CS})\|_F^2 + \lambda \sum_{q=1}^{MQ} \|\mathbf{S}(:, q)\|_1. \quad (3)$$

We minimize the Frobenius norm of the error between available data and their reconstruction through  $\mathbf{C}$  and  $\mathbf{S}$ . The norm one regularization aims to take into account the sparse nature of  $\mathbf{S}$ . This problem can be solved by alternating a LASSO step  $(\mathbf{S}_t = \arg \min \mathcal{J}(\mathbf{C}_{t-1}, \cdot))$  with

$\mathcal{J}(\mathbf{C}_{t-1}, \cdot) = \|P_{\Omega}(\mathbf{\Gamma} - \mathbf{C}_{t-1}\mathbf{S})\|_F^2 + \lambda \sum_{q=1}^{MQ} \|\mathbf{S}(:, q)\|_1$  and a constrained least-squares problem ( $\mathbf{C}_t = \operatorname{argmin} \mathcal{J}(\cdot, \mathbf{S}_t)$ ) with  $\mathcal{J}(\cdot, \mathbf{S}_t) = \|P_{\Omega}(\mathbf{\Gamma} - \mathbf{C}_t\mathbf{S}_t)\|_F^2$ . Both  $\mathcal{J}(\cdot, \mathbf{S}_t)$  and  $\mathcal{J}(\mathbf{C}_{t-1}, \cdot)$  are convex functions. Under mild conditions, the alternating minimization process converges to a stationary point.

After convergence, the imputed matrix can be computed as:

$$\hat{\mathbf{\Gamma}} = P_{\Omega}(\mathbf{\Gamma}) + P_{\bar{\Omega}}(\mathbf{C}\mathbf{S}), \quad (4)$$

$\bar{\Omega}$  being the complement of  $\Omega$ . We can note that it exists a permutation matrix  $\mathbf{\Pi}$  of dimensions  $(\Delta_p + \Delta_f)M \times (\Delta_p + \Delta_f)M$  such that:

$$\begin{pmatrix} \mathbf{T}\mathbf{T}_{k-\Delta_p+1,d} \\ \mathbf{T}\mathbf{T}_{k-\Delta_p+2,d} \\ \vdots \\ \mathbf{T}\mathbf{T}_{k,d} \end{pmatrix} = \mathbf{\Pi} \operatorname{vec}(\mathbf{\Gamma}_d).$$

where  $\operatorname{vec}(\mathbf{\Gamma}_d)$  stands for the operator that vectorizes the matrix  $\mathbf{\Gamma}_d$  by stacking its columns.

### B. Multi-OD travel time clustering

In the previous section, the imputation was based on the estimation of the matrix  $\mathbf{C}$  assumed to be constituted with centroids of each OD pair. Building composite centroids from  $\mathbf{C}$  can yield a huge number of clusters. Instead, from the imputed matrix, we aim to clusterize the set of vectors  $\mathcal{S} = \{\operatorname{vec}(\mathbf{\Gamma}_1), \operatorname{vec}(\mathbf{\Gamma}_2), \dots, \operatorname{vec}(\mathbf{\Gamma}_D)\}$ . For this, let define by  $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_Q\}$  the set of centroids. Each one of the clusters are found by minimizing the distance between each day  $\operatorname{vec}(\mathbf{\Gamma}_i)$  with respect to the centroid  $\boldsymbol{\mu}_j$ . The main objective is to gather particular regimes and most representative conditions and patterns of the travel time profiles. The generation of  $\boldsymbol{\mu}_i$  is performed by the K-means algorithm that minimises the following criteria:

$$\operatorname{argmin}_j \sum_{i=1}^Q \sum_{i=1}^D \|\operatorname{vec}(\mathbf{\Gamma}_i) - \boldsymbol{\mu}_j\|^2. \quad (5)$$

The classification algorithm is applied over the set  $\mathcal{S}$ . The set of centroids can be structured as  $\boldsymbol{\mu}_q = [\boldsymbol{\mu}_{k-\Delta_p+1,q}^T \quad \boldsymbol{\mu}_{k-\Delta_p+2,q}^T \quad \dots \quad \boldsymbol{\mu}_{k+\Delta_f,q}^T]^T$ .

### C. Forecasting problem

The forecasting problem can then be stated as follows: Given the measurements of travel time  $\mathbf{\Gamma}_d$  within the day  $d$  for some OD pairs in the past interval defined from  $k - \Delta_p + 1$  up to  $k$ , and given a set of  $D$  historical days with associated travel time data for some origins and destinations  $\hat{\mathbf{\Gamma}}$ . The aim is to obtain a single prediction  $\hat{\mathbf{T}}_{k+l,d}$  where  $1 \leq l \leq \Delta_f$ . In order to perform the task we produce a separation of the matrix  $\hat{\mathbf{\Gamma}}$  in clusters and then construct individual forecasts for each one of the clusters. Afterwards we look for a fusion method based on best linear unbiased estimator principles to create the desired forecast.

## IV. PREDICTION ALGORITHM

### A. Local Kalman Filter

In general the elements belonging to the  $q$ th cluster can be written as  $\mathbf{T}\mathbf{T}_{k,d} = \boldsymbol{\mu}_{k,q} + \mathbf{w}_{k,q}$ . Assuming that the current day belongs to a cluster  $q$ , we adopt a random walk model for the time evolution of the elements of the cluster:

$$\mathbf{T}\mathbf{T}_{k+1,d} = \mathbf{T}\mathbf{T}_{k,d} + \mathbf{v}_{k,q} \quad (6a)$$

$$\mathbf{v}_{k,q} \sim \mathcal{N}(\Delta\boldsymbol{\mu}_{k,q}, \mathbf{Q}_{k,q}), \quad (6b)$$

where  $\Delta\boldsymbol{\mu}_{k,q} = \boldsymbol{\mu}_{k+1,q} - \boldsymbol{\mu}_{k,q}$  corresponds to mean of the process noise for the dynamic model for the cluster  $q$ .  $\mathbf{Q}_{k,q}$  is the process noise covariance computed as:

$$\mathbf{Q}_{k,q} = \frac{1}{|q| - 1} \sum_{i \in q} (\mathbf{T}\mathbf{T}_{k,i} - \boldsymbol{\mu}_{k,q})(\mathbf{T}\mathbf{T}_{k,i} - \boldsymbol{\mu}_{k,q})^T,$$

with  $|q|$  denoting the number of elements belonging to the  $q$ th cluster. Taking into account (2), we can note that:

$$\mathbf{t}\mathbf{t}_{k+1,d} = \boldsymbol{\Psi}\mathbf{t}\mathbf{t}_{k,d} + \boldsymbol{\Phi}^\dagger \mathbf{v}_{k,q}, \quad (7)$$

with  $\boldsymbol{\Psi} = \boldsymbol{\Phi}^\dagger \boldsymbol{\Phi}$ , where we have assumed  $\boldsymbol{\Phi}$  to be full rank, in this case  $\boldsymbol{\Phi}^\dagger$  denotes the pseudo-inverse of  $\boldsymbol{\Phi}$ . This equation defines the dynamics of the travel time from the  $q$ th cluster point of view.

In particular for the current day  $d$  and up to time  $k$ , we retrieve measurements of OD travel times given by  $\mathbf{z}_{k,d} = \mathbf{H}\mathbf{T}\mathbf{T}_{k,d}$  where  $\mathbf{H}$  is a row selection matrix of dimension  $P \times M$  with  $P < M$ . We assume that if a travel time belongs to a given cluster  $q$  then  $\mathbf{T}\mathbf{T}_{k+1,d} - \mathbf{T}\mathbf{T}_{k,d} = \Delta\boldsymbol{\mu}_{k,q} + \mathbf{m}_{k,q}$ , with  $\mathbf{m}_{k,q} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{k,q})$ . As a consequence:

$$\mathbf{T}\mathbf{T}_{k+j,d} = \mathbf{T}\mathbf{T}_{k,d} + \sum_{i=1}^j \Delta\boldsymbol{\mu}_{k+i,q} + \mathbf{m}_{k+i,q} \quad j \geq 1,$$

which yields  $\mathbf{z}_{k+j,d} = \mathbf{z}_{k,d} + \mathbf{H} \sum_{i=1}^j (\Delta\boldsymbol{\mu}_{k+i,q} + \mathbf{m}_{k+i,q})$ . Let represent by:

$$\mathbf{y}_{k+j,d} = \mathbf{z}_{k,d} + \mathbf{H} \sum_{i=1}^j (\Delta\boldsymbol{\mu}_{k+i,q}), \quad j = 1, \dots, \Delta_f,$$

the pseudo-observations of the travel time from the point of view of the  $q$ th cluster. Since by definition,  $\mathbf{z}_{k+j,d} = \mathbf{H}\mathbf{T}\mathbf{T}_{k+j,d} = \mathbf{H}\boldsymbol{\Phi}\mathbf{t}\mathbf{t}_{k+j,d}$ , we can conclude that:

$$\mathbf{y}_{k+j,d} = \mathbf{H}\boldsymbol{\Phi}\mathbf{t}\mathbf{t}_{k+j,d} + \sum_{i=1}^j \mathbf{m}_{k+i,q}. \quad (8)$$

Strictly speaking, from cluster  $q$ , the system evolves for day  $d$  as:

$$\mathbf{t}\mathbf{t}_{k+1,d} = \boldsymbol{\Psi}\mathbf{t}\mathbf{t}_{k,d} + \boldsymbol{\Phi}^\dagger \mathbf{v}_{k,q} \quad (9)$$

$$\mathbf{y}_{k+j,d}^q = \mathbf{H}\boldsymbol{\Phi}\mathbf{t}\mathbf{t}_{k+j,d} + \mathbf{w}_{k+j,q}, \quad (10)$$

with  $\mathbf{w}_{k+j,q}$  a zero mean noise with covariance  $\check{\mathbf{R}}_{k+j,q} = \sum_{i=k}^{k+j} \mathbf{R}_{i,q}$ , where,

$$\mathbf{R}_{k,q} = \frac{1}{|q| - 1} \sum_{i \in q} (\Delta\mathbf{T}\mathbf{T}_{k,i} - \Delta\boldsymbol{\mu}_{k,q})(\Delta\mathbf{T}\mathbf{T}_{k,i} - \Delta\boldsymbol{\mu}_{k,q})^T.$$

From the model defined by (9) and (10) we intend to run separate predictions for each one of the clusters. Thanks to the pseudo-observations built from clustered historical data, the prediction problem can then be viewed as a filtering one (See. [11]). For each one of the cluster, separate models will evolve in the future according to the following Local Kalman Filter (LKF) strategy as:

$$\mathbf{t}\hat{\mathbf{t}}_{k+1|k,q} = \Psi\hat{\mathbf{t}}_{k|k} + \Phi^\dagger \Delta\boldsymbol{\mu}_{k,q} \quad (11a)$$

$$\mathbf{P}_{k+1|k,q} = \Psi\hat{\mathbf{P}}_{k|k,q}\Psi^T + \Phi^\dagger \mathbf{Q}_{k,q}\Phi^{\dagger T} \quad (11b)$$

$$\mathbf{K}_{k+1|k,q} = \mathbf{P}_{k+1|k,q}\Phi^T \mathbf{H}^T \left( \mathbf{H}\Phi\mathbf{P}_{k+1|k,q}\Phi^T \mathbf{H}^T + \check{\mathbf{R}}_{k+1,q} \right)^{-1} \quad (11c)$$

$$\hat{\mathbf{t}}_{k+1|k+1,q} = \mathbf{t}\hat{\mathbf{t}}_{k+1|k,q} + \mathbf{K}_{k+1|k,q}(\mathbf{y}_{k+1,D}^q - \mathbf{H}\Phi\mathbf{t}\hat{\mathbf{t}}_{k+1|k,q}) \quad (11d)$$

$$\hat{\mathbf{P}}_{k+1|k+1,q} = (\mathbf{I} - \mathbf{K}_{k+1,q}\mathbf{H}\Phi)\mathbf{P}_{k+1|k,q}, \quad (11e)$$

where  $\hat{\mathbf{P}}_{k|k,q} \in \mathfrak{R}^{n \times n}$  represents the error covariance matrix. The evolution of the error covariance matrix provides information regarding the convergence of the estimate given by (11) to the true value. We are interested in considering this information in order to fusion the forecasts obtained for each one of the clusters. We remark the fact that LKF strategies are connected as a feedback system (See subsection IV-B).

### B. Fusion Algorithm

Sometimes the regular performance of the clustering algorithms not always separates individual patterns. The main objective by introducing a fusion algorithm is justified in the fact that the combined forecast might improve its performance as it is discussed in [19], particularly during peak times. In order to make the fusion of the forecasts provided by the clusters, we take into account the fact that  $\tilde{\mathbf{t}}_{k,q} = \hat{\mathbf{t}}_{k|k,q} - \mathbf{t}\hat{\mathbf{t}}_k$ , the estimation error for each one of the LKF strategies, has variance  $\mathbf{P}_{k|k,q}$ . We assume that each forecast is a noisy version of the actual travel time. The algorithm is detailed in Fig. 2.

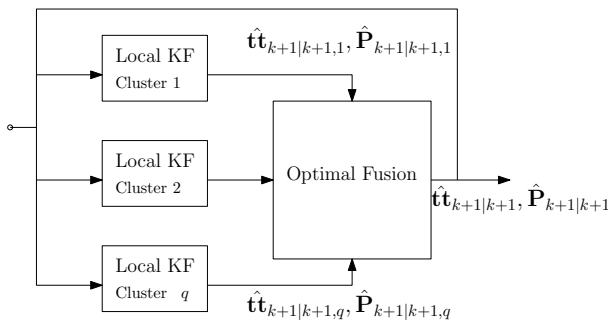


Fig. 2. Feedback fusion mixing strategy

In what follows, we aim to fuse the forecasts provided by each cluster with a weighting term according to their error covariance matrix. The weighting information will provide weight to those clusters with more confidence for the future horizon. The measurement for the confidence is given by the information of the covariance error. For this purpose, let

concatenate all the forecasts in a vector  $\boldsymbol{\theta}_k \in \mathfrak{R}^{NQ}$ ,  $Q$  being the number of clusters. We aim to find

$$\hat{\mathbf{t}}_k = \underset{\mathbf{t}_k}{\operatorname{argmin}} \|\boldsymbol{\theta}_k - \mathbf{G}\mathbf{t}_k\|_{\mathbf{U}_k}^2, \quad (12)$$

with  $\mathbf{G} = [\mathbf{I} \ \mathbf{I} \ \dots \ \mathbf{I}]^T$  a  $NQ \times N$  matrix, and  $\mathbf{U}_k$  the forecasts error covariance matrix. The main objective of the problem (12) is to perform a weighted data fitting problem based on measurements provided by the individual cluster forecasts and weighted by  $\mathbf{U}_k$  which is conformed by  $NQ \times NQ$  block matrices denominated  $\mathbf{u}_{[ij]_k}$  denoting the error covariance between the estimates from two different clusters  $i$  and  $j$ . In this case  $\mathbf{u}_{[ij]_k}$  is given by the following proposition:

*Proposition 1:* Consider the individual estimates  $\boldsymbol{\theta}_k$  from the Local Kalman filters and the optimal fusion given by (12). Then the error covariance matrix of the fusion is given by:

$$\mathbf{u}_{[ij]_{k+1}} = \Upsilon_i \Psi \mathbf{u}_{[ij]_k} \Psi^T \Upsilon_j^T + \delta(i-j) \left( \Upsilon_i \mathbf{Q}_{k,j} \Upsilon_j^T + \mathbf{K}_{k+1|k,i} \check{\mathbf{R}}_{k+1,j} \mathbf{K}_{k+1|k,j}^T \right), \quad (13)$$

where  $\delta(i-j)$  is the Kronecker delta function and  $\Upsilon_i = \mathbf{I} - \mathbf{K}_{k+1|k,i}\mathbf{H}\Phi$ .

*Proof:* Consider the prediction equation (11d) that computes the update in the filter. Consider also the replacement of the terms  $\hat{\mathbf{t}}_{k+1|k,q}$  and  $\mathbf{y}_{k+1,D}^q$  given by (11a) and (10)

$$\hat{\mathbf{t}}_{k+1|k,j} = \mathbf{A}_{k+1,j} \Psi \left( \hat{\mathbf{t}}_{k|k} + \Phi^\dagger \Delta\boldsymbol{\mu}_{k,j} \right) + \mathbf{K}_{k+1|k,j} \left( \mathbf{H}\Phi\mathbf{t}\hat{\mathbf{t}}_{k+1,j} + \mathbf{w}_{k+1,j} \right), \quad (14)$$

where  $\mathbf{A}_{k+1,j} = (\mathbf{I} - \mathbf{K}_{k+1|k,j}\mathbf{H}\Phi)$ . Subtracting the term  $\mathbf{t}\hat{\mathbf{t}}_{k,j}$  and using (6) on the rightside part of the equation we obtain.

$$\tilde{\mathbf{t}}_{k,j} = \mathbf{A}_{k+1,j} \left( \Psi \tilde{\mathbf{t}}_k^j + \Phi^\dagger \mathbf{v}_{k,j} + \Phi^\dagger \Delta\boldsymbol{\mu}_{k,j} \right) + \mathbf{K}_{k+1|k,j} \mathbf{w}_{k+1,j}. \quad (15)$$

We find the covariance between two clusters  $l, j$ . Furthermore, we group the vector  $\mathbf{X}_j = [\tilde{\mathbf{t}}_{k,j} \ \mathbf{v}_{k,j} \ \mathbf{w}_{k+1,j}]^T$  and consider the covariance between the affine term  $\mathbf{C}_j \mathbf{X}_j + \mathbf{D}_j$  where  $\mathbf{C}_j = [\mathbf{A}_{k+1,j}\Psi \ \mathbf{A}_{k+1,j}\Phi^\dagger \ \mathbf{K}_{k+1|k,j}]^T$ ,  $\mathbf{D}_j = \mathbf{A}_{k+1,j}\Phi^\dagger \Delta\boldsymbol{\mu}_{k,j}$  is given by:

$$E[\tilde{\mathbf{t}}_{k,i} \tilde{\mathbf{t}}_{k,j}] = \mathbf{C}_i E[\mathbf{X}_i \mathbf{X}_j^T] \mathbf{C}_j^T. \quad (16)$$

In equation (16) by previous hypothesis of the Kalman filter over independency between observation noise and process noise. (See. (9)). The expression leads to:

$$\mathbf{u}_{[lj]_{k+1}} = \mathbf{A}_{k+1,l} \Psi \mathbf{u}_{[lj]_k} \Psi^T \mathbf{A}_{k+1,j}^T + \delta(l-j) \mathbf{A}_{k+1,l} E[\mathbf{v}_{k,l} \mathbf{v}_{k,j}^T] \mathbf{A}_{k+1,j}^T + \delta(l-j) \mathbf{K}_{k+1|k,l} E[\mathbf{w}_{k,l} \mathbf{w}_{k,j}^T] \mathbf{K}_{k+1|k,j}^T, \quad (17)$$

which can be written as (13). ■

The optimal solution in this case is given by the best linear unbiased estimator as:

$$\hat{\mathbf{t}}_k^* = \boldsymbol{\Lambda}_k \boldsymbol{\theta}_k, \quad (18)$$

where  $\boldsymbol{\Lambda}_k = (\mathbf{G}^T \mathbf{U}_k \mathbf{G})^{-1} \mathbf{G}^T \mathbf{U}_k$ . The solution is then computed iteratively each sample time with updates in the

weighting matrix and the individual forecasts contained in  $\theta_k$ . The variance of the estimator given in (18) can be computed in this case as:

$$\hat{\Gamma}_k = \Lambda_k \mathcal{D}(\hat{\mathbf{P}}_{k|k}) \Lambda_k^T, \quad (19)$$

where  $\mathcal{D}(\hat{\mathbf{P}}_{k|k})$  is the block-diagonal matrix with  $\hat{\mathbf{P}}_{k|k,q}$  as blocks.

## V. EXPERIMENTAL SETUP FOR THE FORECASTING STRATEGY

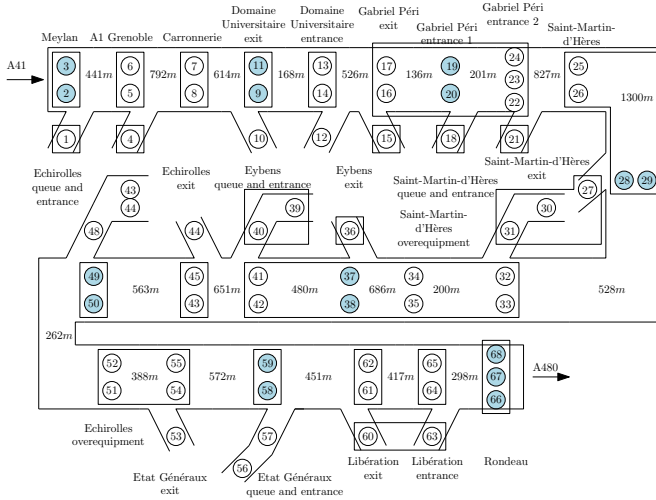


Fig. 3. Sensors distribution along Rociade-Sud

### A. Scenario Description

The experimental data used for the test is obtained from the Grenoble Traffic Lab (GTL). The network consists of 135 magnetometer sensors distributed along 10.5 Km connecting the highway A31 (north-west) to A480 (south) in Grenoble France. The network offers 68 possible locations for collecting speeds in fast and slow lanes. For further details see [1]. Fig. 3 provides details of the sensor distribution along the road. For this case we assume the scenario forecasting the travel time in a future horizon  $\Delta_f = 25$  minutes entering the highway at Meylan and leaving at Rondeau. Predictions are performed between 7:30am and 9:30am, time in which it is found peaks of congestion. Traveling times vary between 7 minutes in free flow conditions and 25 minutes in congested cases.

We constrain the set of historical data to 93 week days collected from April 2014 to May 2015. For all the tests, a fixed number of  $K = 3$  clusters was used, the value was selected based on the average number  $K$  obtained according to [13] and performing single clusterization of the full data set each 1 minute for the window of time 7:30-9:30am. In case of special conditions for empty clusters the algorithm reduces automatically the number  $K$  by one.

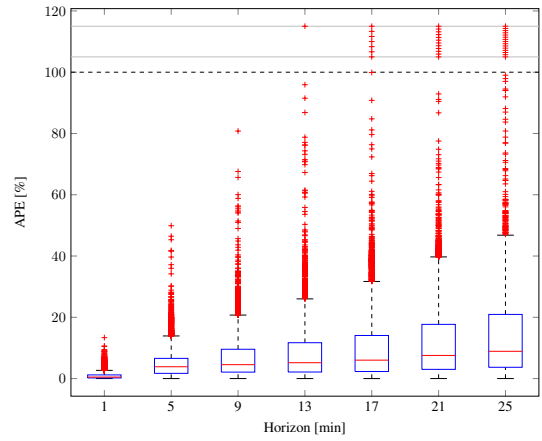


Fig. 4. Error distribution vs Horizon of prediction

### B. Performance Measurement

In order to assess performance of the method, we consider the Absolute Percentage Error (APE) as a reference. For some future horizon  $1 \leq \Delta_f$  the APE is given by:

$$APE_{k+h,d} = 100 \frac{|\widehat{TT}_{k+h,d} - TT_{k+h,d}|}{TT_{k+h,d}}. \quad (20)$$

A leave one out cross validation test is performed over the data set to perform validation. The element out of the set is the candidate day  $d$  for performing the prediction. The APE index in Eq. (20) is computed for each one of the forecasts and varying the departure time in steps of one minute until 9:30am. We select this type of measurement based on the fact that there might exist greater scenarios considering multiple OD that can be assessed under the same criteria.

### C. Results

The results condensate a total amount of 11160 forecasts that were launched with fixed prediction horizon of 25 minutes starting at 7:30am. Each minute 93 forecasts are performed from historical data by considering leave one cross validation schema. The departure time is then modified in steps of 1 minute until 9:30am, which is the normal time of the day when conditions return to free flow. In all cases the error is measured for the full prediction. Particularly we depict APE error from Eq. (20) for a fixed OD scenario that considers cars entering at Meylan and leaving at Rondeau. Fig. 4 illustrates the behaviour of the APE error according to the future horizon. Even though the prediction is done in steps of 1 minute, we summarize here the main statistical results of performance in steps of 5 minutes. The median of the data set is displayed as a red line within the blue boxes. Red crosses in each horizon time represent outliers of the prediction found when the samples are in distance longer than  $\pm 3\sigma$  for a normal distribution  $\mathcal{N}(\mu, \sigma^2)$ . These values may occur when K-means reduces the number of clusters since dispersion of the clusterized data sets is bigger. It is seen an increasing APE along the future horizon due to the high dependency on the previous sample in order to

perform the forecast. The maximum error obtained for any future horizon of 25 minutes is 40%. Indeed, this confirms the fact of high difficulties in accuracy with a long term forecasting since no prior information is known. However the maximum median at the same time horizon is 15.7% which establishes better generalized performance of the method with respect to the full population.

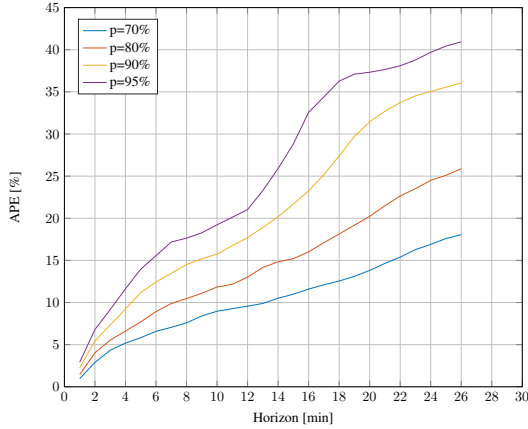


Fig. 5. Guaranteed Probability  $p(APE < p)$

On the other hand Fig. 5 displays the performance of the same method from the point of view of the experimental cumulative distribution function. In this case we analyze the cumulated APE error over the data set and we display several curves for guaranteed APE under some probabilities. The plot depicts the amount of error guaranteed with probability  $p = [70\ 80\ 90\ 95](\%)$ . From the plot we infer that in the future horizon of 25 minutes, the maximum error is 17.1% with 70% of probability, up to 38.3% when the probability is 95%.

## VI. CONCLUSIONS

This paper has addressed the problem of travel time forecasting over highways based on the fusion of individual forecasts performed for clusterized time series. The individual clustering classification is proposed to establish a framework for the prediction. Individual forecasts models are developed under a Kalman filter approach. A fusion algorithm is then proposed in order to merge the predictions. The assessment is applied over real data in which error is guaranteed to be less or equal than 35% in 90% of the cases. Future research on this work may include but is not limited to the addition of constraints into the optimal fusion algorithm, the study of statistical merge between model based approaches and the presented strategy and further studies on the effect of the horizon in relationship with the clusterized time series and the forecasting strategy.

## ACKNOWLEDGMENT

This work has been funded by the EU FP7 project SPEEDD (619435).

## REFERENCES

- [1] C. Canudas de Wit, F. Morbidi, L. L. Ojeda, A. Y. Kibangou, I. Bellicot, and P. Bellemain. Grenoble Traffic Lab: An Experimental Platform for Advanced Traffic Monitoring and Forecasting. *IEEE Control Systems*, 35(3):23–39, June 2015.
- [2] H. Chang, Y. Lee, B. Yoon, and S. Baek. Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences. *IET Intelligent Transport Systems*, 6(3):292, 2012.
- [3] R. Chrobok, O. Kaumann, J. Wahle, and M. Schreckenberg. Different methods of traffic forecast based on real data. *European Journal of Operational Research*, 155(3):558–568, June 2004.
- [4] L. Du, S. Peeta, and Y. H. Kim. An adaptive information fusion model to predict the short-term link travel time distribution in dynamic traffic networks. *Transportation Research Part B: Methodological*, 46(1):235–252, 2012.
- [5] F. G. Habtemichael and M. Cetin. Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transportation Research Part C: Emerging Technologies*, Sept. 2015.
- [6] W.-C. Hong, Y. Dong, F. Zheng, and C.-Y. Lai. Forecasting urban traffic flow by SVR with continuous ACO. *Applied Mathematical Modelling*, 35(3):1282–1291, Mar. 2011.
- [7] E. Lovisari, C. Canudas de Wit, and A. Y. Kibangou. Data fusion algorithms for Density Reconstruction in Road Transportation Networks. In *54th IEEE Conference on Decision and Control*, Osaka, 2015. IEEE.
- [8] J. Marial, F. Bach, and J. Ponce. Sparse Modeling for Image and Vision Processing. *Foundations and Trends in Computer Graphics and Vision*, 8(2):85–283, 2014.
- [9] T. Mu, J. Jiang, and Y. Wang. Heterogeneous Delay Embedding for Travel Time and Energy Cost Prediction Via Regression Analysis. *IEEE Transactions on Intelligent Transportation Systems*, 14(1):214–224, Mar. 2013.
- [10] L. L. Ojeda, A. Kibangou, and C. Canudas de Wit. Online Dynamic Travel Time Prediction using Speed and Flow Measurements. In *European Control Conference*, pages 4045–4050, Zurich, 2013. IEEE.
- [11] L. L. Ojeda, A. Y. Kibangou, and C. Canudas de Wit. Adaptive Kalman Filtering for Multi-Step ahead Traffic Flow Prediction. In *American Control Conference (ACC)*, pages 4731–4736, Washington, 2013.
- [12] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37(23):3311–3325, Dec. 1997.
- [13] D. T. Pham, S. S. Dimov, and C. D. Nguyen. Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1):103–119, Jan. 2005.
- [14] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias. Short-term traffic forecasting: Where we are and where we are going. *Transportation Research Part C: Emerging Technologies*, 43:3–19, June 2014.
- [15] N. Wan, G. Gomes, and R. Horowitz. Prediction on Travel-Time Distribution for Freeways Using Online Expectation Maximization Algorithm. In *Transportation Research Board 93rd Annual Meeting*, pages 14–3221, Washington, 2014.
- [16] W. Weijermars and E. van Berkum. Analyzing highway flow patterns using cluster analysis. In *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, pages 831–836. IEEE, 2005.
- [17] C.-j. Wu, T. Schreiter, and R. Horowitz. Multiple-clustering ARMAX-based predictor and its application to freeway traffic flow prediction. In *2014 American Control Conference*, pages 4397–4403, Portland, June 2014. IEEE.
- [18] Y. Zhang. Hourly Traffic Forecasts Using Interacting Multiple Model (IMM) Predictor. *IEEE Signal Processing Letters*, 18(10):607–610, Oct. 2011.
- [19] Y. Zhang and Y. Liu. Analysis of peak and non-peak traffic forecasts using combined models. *Journal of Advanced Transportation*, 45(1):21–37, Jan. 2011.