Short-Term Urban Rail Passenger Flow Forecasting: A Dynamic Bayesian Network Approach

Jérémy Roos
RATP / Univ Lyon, ERIC EA3083
F75012 Paris, France
Email: jeremy.roos@ratp.fr

Stéphane Bonnevay and Gérald Gavin
Univ Lyon, ERIC EA3083
F69100 Villeurbanne, France
Email: {stephane.bonnevay, gerald.gavin}@univ-lyon1.fr

***Abstract*—We propose a dynamic Bayesian network approach to forecast the short-term passenger flows of the urban rail network of Paris. This approach can deal with the incompleteness of the data caused by failures or lack of collection systems. The structure of the model is based on the causal relationships between the adjacent flows and is designed to take into account the transport service. To reduce the number of arcs and find the maximum likelihood estimate of the parameters, we perform the structural expectation-maximization (EM) algorithm. Then short-term forecasting is conducted by inference, using the bootstrap filter. An experiment is carried out on an entire metro line, using ticket validation, count and transport service data. Overall, the forecasting results outperform historical average and last observation carried forward (LOCF). They illustrate the potential of the approach, as well as the key role of the transport service.**

I. Introduction

RATP is the main public transport operator in the Paris region. It operates all 16 metro lines, sections of 2 RER (commuter rail) lines, 8 tramway lines and more than 350 bus lines. Currently, it uses several tools for passenger flow modeling, whose purpose is to assess the long-term effects of infrastructure or transport policy changes. However, these models are not designed for short-term forecasting and cannot take into account the impacts of unanticipated or non-recurrent events (e.g., service disruptions, unplanned closures of stations, crowd-attracting events). Furthermore, the diversity of available data is still largely untapped. Each service generally works with its own sources and thus only has a partial view of the mobility within the network.

Given these observations, we propose an approach to forecast the short-term passenger flows of the RATP urban rail network (metro and RER). Based on dynamic Bayesian networks, this approach is designed to harness the diversity of data and to forecast in case of incomplete data. By providing real-time predictions, it can cater for various applications in transport system management, such as operation planning, passenger flow regulation and passenger information.

First, we detail the mobility data collected by RATP. After a brief state of the art of short-term traffic forecasting, we introduce Bayesian networks and their temporal extension. Then we expose our modeling approach, which we apply to an entire Paris metro line. After presenting the results, we conclude the paper and discuss future work.

II. Mobility Data

RATP collects extensive data on the passenger mobility, which come from various sources: ticket validation, counts, surveys, transport service, expert knowledge, modeling results, etc. In this work, we focus on ticket validation, count and transport service data.

The RATP urban rail network can be divided into "public" and "controlled" areas. The public areas are accessible to all people, whereas the controlled areas require a valid ticket. Ticket validation takes place when passengers enter controlled areas from public or other controlled areas, and when they leave RER controlled areas. After adjusting for ticket evasion, it can be regarded as exhaustive counts.

Depending on the needs, count data can be collected anywhere in the network, manually or using automatic devices. Many counts are conducted to measure the number of passengers traveling by train between successive stop points. They are performed at train departures by specialized agents or, for a few lines, by on-board weighing systems.

The transport service data provide the train departure and arrival times at each stop point. These times are scheduled by the public transport operator.

Due to failures of collection systems and to the fact that these systems may be not used continuously, the data collected are incomplete. This issue affects the choice of the model, which must be able to forecast in any situation.

III. Short-Term Traffic Forecasting

According to Van Lint and Van Hinsbergen [1], forecasting the short-term traffic can be reduced to solving the following regression problem:

$$y^{(t)} = \mathcal{G}\left(x^{(k|k<t)}, e^{(k|k<t)}, \Theta^{(t)}\right) + e^{(t)} \qquad (1)$$

where $\mathcal{G}$ is the chosen model, $y^{(t)}$ is the vector of output variables at time slice t, $x^{(t)}$ is the vector of input variables at t (which can contain historical instances of y), $\Theta^{(t)}$ is the vector of adjustable parameters at t and $e^{(t)}$ is a noise process that represents unobservable factors at t.

There is a vast literature on short-term traffic forecasting, which can be classified into naïve, parametric and nonparametric methods [2]. Because of their easy implementation, naïve methods have been widely used, including historical average [3] and last observation carried forward (LOCF), also called random walk [4]. Among parametric methods, a great focus has been given to ARIMA models since the late 1970s [4], [5]. Kalman filter approaches have also been successfully applied [6], [7]. In recent years, particular attention has been afforded to nonparametric methods, such as nonparametric regression [8], [9] and neural networks [10], [11]. Their ability to better model nonlinear processes has contributed to their popularity.

Most research in short-term traffic forecasting has focused on vehicle flows in road networks. By contrast, little work has been devoted to passenger flows in public transport networks, the existing models being mainly designed for long-term planning practice [12]. Some authors have recently begun to tackle this issue, especially through neural network approaches [13]–[15].

Missing data is a common problem in many real-world situations. Although various methods have been proposed for traffic data imputation [16], [17], few of them are designed to operate in a real-time setting [18]. Hence, most of the models are not able to provide real-time predictions in case of incomplete data. Bayesian network approaches have been developed to address this problem, both on urban networks [19] and on highways [20].

IV. Bayesian Networks

A. Representation

Introduced by Pearl [21], Bayesian networks represent the conditional dependencies (and independencies) between random variables by a directed acyclic graph. These dependencies are described by a joint probability distribution, which decomposes into a product of local conditional distributions:

$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i | Pa(X_i)) \tag{2}$$

where $Pa(X_i)$ is the set of parents of X_i.

The flexibility of Bayesian networks allows to combine heterogeneous information sources. For example, an expert can estimate a part of the model from data and another part from his own knowledge [22]. Furthermore, the information propagation mechanism makes inference possible in case of incomplete data. This property is particularly useful in a real-time setting, where the implementation of an additional imputation process can be detrimental.

B. Linear Gaussian Bayesian Networks

When the variables are continuous, the local conditional distributions of a Bayesian network can be described as linear Gaussians [23]:

$$p(X_i | Pa(X_i)) = \mathcal{N}\left(\beta_{i,0} + \beta_i^\top Pa(X_i)\,, \sigma_i^2\right) \tag{3}$$

where $\beta_{i,0}$, β_i and σ_i^2 are the parameters to be estimated (for each i). The use of these distributions implies that the relationships between the variables are linear.

C. Dynamic Bayesian Networks

Dynamic Bayesian networks extend Bayesian networks to model the temporal relationships between variables [24]. Thus, they can represent systems that evolve over time. Assuming that this evolution occurs between discrete time slices, the joint distribution is expressed analogously to (2):

$$p\left(X^{(1)}, \ldots, X^{(T)}\right) = \prod_{t=1}^{T} \prod_{i=1}^{n} p\left(X_i^{(t)} \middle| Pa\left(X_i^{(t)}\right)\right) \tag{4}$$

where $X_i^{(t)}$ is the instantiation of X_i at time slice t and $X^{(t)} = \left\{X_1^{(t)}, \ldots, X_n^{(t)}\right\}$.

By extending the definition of Murphy [25] for first-order dynamic Bayesian networks, an rth-order ($r \geq 1$) dynamic Bayesian network is defined as an $(r+1)$-tuple of Bayesian networks $(\mathcal{B}_1, \ldots, \mathcal{B}_r, \mathcal{B}_\rightarrow)$, where $\mathcal{B}_1$ defines $p\left(X^{(1)}\right)$, $\mathcal{B}_t$ defines $p\left(X^{(t)} \middle| X^{(t-1)}, \ldots, X^{(1)}\right)$ for $2 \leq t \leq r$ and $\mathcal{B}_\rightarrow$ defines $p\left(X^{(t)} \middle| X^{(t-1)}, \ldots, X^{(t-r)}\right)$ for $t > r$.

V. Modeling Approach

A. Modeling Mechanism

Following Sun et al. [19], the most intuitive way to build the model is to consider that the passenger flows at given points have causal relationships with the flows located downstream in the transport network. Thus, we define a parameter d such that each flow at time slice t depends on its adjacent upstream flows at $t-1, \ldots, t-d$. As the historical values of the flows provide information on their trend, we also define a parameter m such that each flow at t depends on its values at $t-1, \ldots, t-m$ [19]. The resulting dynamic Bayesian network is of order $r = \max(d, m)$.

In a public transport network, the passenger flows are closely related to the transport service. Intuitively, the number of passengers on a train depends on the waiting time between this train and the previous one. Indeed, the longer this waiting time, the more passengers crowd the platform before boarding.

Let X be a flow of passengers traveling by train between two successive stop points and $\mathcal{D}_X^{(t)}$ the set of departure times from the stop point of origin of X during time slice t (several departures can take place during the same time slice). We define the "transport service variable" associated with X at t:

$$S_X^{(t)} = \begin{cases} \max \mathcal{D}_X^{(t)} - \max \bigcup\limits_{k|k<t} \mathcal{D}_X^{(k)}, & \text{if } \mathcal{D}_X^{(t)} \neq \emptyset \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $\max \mathcal{D}_X^{(t)}$ and $\max \bigcup\limits_{k|k<t} \mathcal{D}_X^{(k)}$ are the latest departure times during and before t respectively. $S_X^{(t)}$ can be regarded

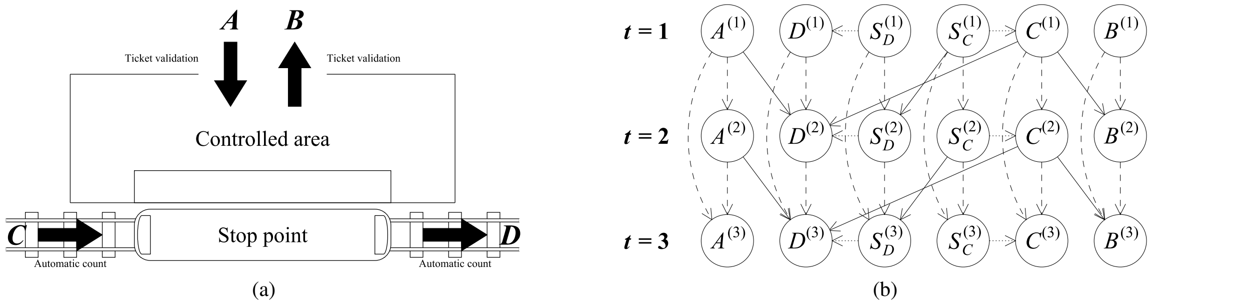

Figure 1. (a) Example of urban rail passenger flows. (b) Corresponding dynamic Bayesian network with $d = 1$ and $m = 2$, unrolled over 3 time slices.

as the total period during which passengers wait for the trains departing during t. In the dynamic Bayesian network, it directly influences $X^{(t)}$. As for passenger flows, it depends on its adjacent upstream transport service variables and on its values at the previous time slices.

All variables in the model are continuous. Assuming that their relationships are linear, we describe the local conditional distributions as linear Gaussians.

Fig. 1 provides an example of passenger flows and their corresponding dynamic Bayesian network with $d = 1$ (solid arcs in Fig. 1b) and $m = 2$ (dashed arcs). In this example, A is upstream of D and C is upstream of B and D. C and D are directly influenced by the transport service (dotted arcs).

B. Parameter Estimation

Given a training dataset, the parameters of a linear Gaussian Bayesian network can be estimated by maximum likelihood. However, the traditional estimation method is unable to provide analytical solutions in the presence of incomplete data. Moreover, if the missing data are widely scattered, listwise deletion may result to an excessive loss of information. The approach of Sun et al. [19] consists in replacing the variables whose values are missing by their parents in the dynamic Bayesian network. However, this method is also difficult to apply, because it implies that the parents are complete, which is not true in many cases.

Proposed by Dempster et al. [26], the expectation-maximization (EM) algorithm is an iterative method for finding the maximum likelihood estimate of the parameters when the dataset has missing (or hidden) values. Starting from an initial guess of the parameters, the EM algorithm performs, at each iteration k, the following two steps [27]:

- the expectation (E) step, where it calculates the expected value of the complete-data log-likelihood, with respect to the missing data $\mathcal{X}_m$ given the observed data $\mathcal{X}_o$ and the current estimate of the parameters $\Theta^{(k-1)}$:

$$Q\left(\Theta \middle| \Theta^{(k-1)}\right) = \mathbf{E}\left[\log p(\mathcal{X}_o, \mathcal{X}_m|\Theta) \middle| \mathcal{X}_o, \Theta^{(k-1)}\right]; \tag{6}$$

- the maximization (M) step, where it estimates the parameters that maximize this expectation:

$$\Theta^{(k)} = \arg\max_{\Theta} Q\left(\Theta \middle| \Theta^{(k-1)}\right). \tag{7}$$

As proved by Dempster et al. [26], each iteration is guaranteed to increase the log-likelihood until convergence to a local maximum.

C. Dimension Reduction

Depending on d and m, the number of arcs in the dynamic Bayesian network can be huge. In addition to increasing the computational complexity, this situation can lead to overfitting and decrease the forecasting performance. In order to avoid these problems, we need to select the best subset of arcs among those described in subsection V-A.

Proposed by Friedman [28], [29], the structural EM algorithm works similarly to the parametric version. Using the Bayesian information criterion (BIC) [30] as scoring function, it performs, at each iteration k, the following two steps:

- the E step, where it calculates the expected value of the BIC score, with respect to the missing data given the observed data and the current estimate of the model (i.e., the structure and the parameters) $\mathcal{B}^{(k-1)}$:

$$Q_{BIC}\left(\mathcal{B} \middle| \mathcal{B}^{(k-1)}\right) = \mathbf{E}\left[\log p(\mathcal{X}_o, \mathcal{X}_m|\mathcal{B}) \middle| \mathcal{X}_o, \mathcal{B}^{(k-1)}\right] - \frac{\log N}{2} \#\mathcal{B} \tag{8}$$

  where N is the number of observations in the dataset and $\#\mathcal{B}$ is the number of free parameters in $\mathcal{B}$;
- the M step, where it estimates the model that maximizes this expectation:

$$\mathcal{B}^{(k)} = \arg\max_{\mathcal{B}} Q_{BIC}\left(\mathcal{B} \middle| \mathcal{B}^{(k-1)}\right). \tag{9}$$

As for the EM algorithm, each iteration increases the BIC score until convergence to a local maximum [28].

In practice, an iteration of the structural EM algorithm consists in performing the EM algorithm to complete the data (E step) and then using the completed dataset to update the model (M step). When the data are complete, the decomposition of the log-likelihood into a sum of local terms allows to maximize the BIC score within each family independantly [28]. This property facilitates the computation and can be exploited by a greedy hill climbing procedure to gradually improve the structure [31].

In their paper, Friedman et al. [32] extend the structural EM algorithm to dynamic Bayesian networks. Because of the decomposition property of the BIC score, each of the $r+1$ structures that compose an rth-order dynamic Bayesian network can be improved independantly.

D. Short-Term Forecasting

The short-term forecasting process is an inference problem in the dynamic Bayesian network. However, the exact inference methods are generally too time-consuming for large models. In order to ensure real-time forecasting, it is necessary to fall back on approximate methods.

The bootstrap filter [33], also known as "survival of the fittest" [34], is a stochastic simulation algorithm that efficiently performs approximate inference in dynamic Bayesian networks. It generates weighted sample sequences by sampling the unobserved values and propagates them forward in time. At each time slice t, the algorithm randomly selects a set of sequences proportionally to their current weight. Each selected sequence is then used to generate a new sample for time slice $t+1$ (prediction step). Upon receiving the measures at $t+1$, the weight is updated to the likelihood of the evidence given the sequence, and so on.

Although the boostrap filter is often applied to first-order models [33], [34], it can be extended to high-order models ($r > 1$), as shown by Pan and Schonfeld [35].

VI. Experiment

A. Input Data

We apply the dynamic Bayesian network approach to the stations served by Paris metro line 2. More precisely, we focus on the flows of passengers entering, leaving and moving through these stations, on foot or by train, including those related to the connected metro and RER lines. In this experiment, we collect three types of data:

- ticket validation data, when passengers enter controlled areas from public areas (28 flows) or leave controlled areas to public or other controlled areas (7 flows);
- automatic count data, at train departures by on-board weighing systems (60 flows);
- transport service data (114 transport service variables).

Due to failures of collection systems and to the fact that some trains are not equipped with weighing systems, 80 of the 95 flows recorded are incomplete. Overall, their missing data rate is 4.8%, but it reaches more than 50% for some flows (by contrast, this rate is only 0.2% for the transport service variables). The missing data also have a high temporal dispersion, covering 99.7% of the time slices.

B. Experimental Method

We divide the dataset into training and test sets, which include the data of the first 24 days and the last 9 days respectively. The structure and the parameters of the dynamic Bayesian network are learned from the training set, using the structural EM algorithm described in subsection V-C. Then short-term forecasting is performed on the test set, using the bootstrap filter described in subsection V-D.

Since the transport service is scheduled by the public transport operator, we assume that the actual values of the transport service variables at time slice t are already known at $t-1$ and do not need to be predicted. This assumption may seem optimistic as unexpected changes can occur at the last moment. To be more realistic, we should preset the transport service variables at t to their values "expected" at $t-1$. However, this information is not stored and therefore cannot be exploited in this experiment.

To evaluate the forecast accuracy of the model, we adopt the weighted mean absolute percentage error (WMAPE):

$$WMAPE\left(x, \hat{x}\right) = \frac{\sum_{t=1}^{N} \left| x^{(t)} - \hat{x}^{(t)} \right|}{\sum_{t=1}^{N} x^{(t)}} \tag{10}$$

where $\hat{x}$ is the estimate of x. Like the mean absolute percentage error (MAPE), the WMAPE is easily interpretable. On the other hand, it weights the errors by the values of x and thereby favors models that are effective in predicting the high values of the flows.

In this experiment, we take the parameters $d = 2$ and $m = 3$, which empirically provide good forecasting results. In order to assess the individual contribution of each type of relationships, two partial versions of the model are tested in addition to the complete one: a version without transport service and a version without relationships between adjacent flows. The results are also compared to those of two naïve methods: historical average, which forecasts the flows by averaging their historical values at the corresponding time slices [3], and LOCF, which simply uses their last observed value [4].

C. Forecasting Results

The results of the experiment are listed in Table I. According to their location in the application area, the passenger flows are classified into three categories. For each of them, the results are expressed by the average WMAPE of the related flows.

Overall, the dynamic Bayesian network approach outperforms the two naïve methods, with average WMAPEs of 18.5% versus 32.1% for historical average and 49.6% for LOCF. The flows measured at train departures largely contribute to these good forecasting results. As mentioned in subsection V-A, they are closely related to the transport service. Therefore, they significantly benefit from the incorporation of this information, their average WMAPE

TABLE I
COMPARISON OF THE FORECASTING RESULTS FOR DIFFERENT METHODS AND CATEGORIES OF FLOWS (AVERAGE WMAPE IN %)

Passenger flows	**Dynamic Bayesian network**			**Historical average**	**LOCF**
	(complete)	(w/o transport service)	(w/o rel. between adj. flows)		
At train departures	17.8	37.3	21.1	40.3	63.7
From public to controlled areas	19.0	19.0	19.0	16.9	24.0
From controlled to public/controlled areas	22.6	23.7	24.8	22.2	31.6
All passenger flows	18.5	30.9	20.7	32.1	49.6

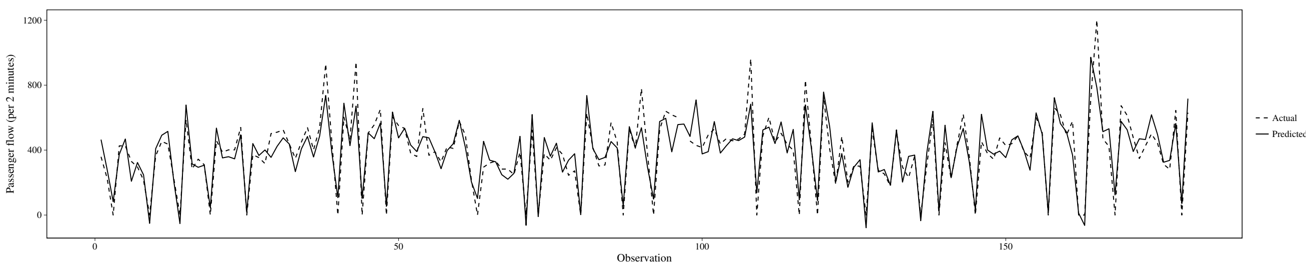

Figure 2. Actual and predicted values of the passenger flow from Blanche station to Place de Clichy station (April 7 to 9, 2015, from 7.30 to 9.30 am).

dropping from 37.3% to 17.8%. By contrast, historical average and LOCF perform poorly for this category of flows, with average WMAPEs of 40.3% and 63.7% respectively.

Fig. 2 shows the actual and predicted values of the flow from Blanche station to Place de Clichy station (WMAPE of 17.7%) for the test days from 7 to 9 April 2015. It well illustrates the ability of the model to fit the large fluctuations that are inherent in the transport service.

The relationships between the adjacent flows play an important role in the modeling. Their contribution is clearly visible for the flows measured at train departures (average WMAPEs of 17.8% versus 21.1% without these relationships) and, to a lesser extent, for those from controlled to public or other controlled areas (22.6 % versus 24.8 %). This is probably one of the reasons why our approach is slightly less effective than historical average for the passenger flows from public to controlled areas (average WMAPEs of 19.0% versus 16.9%). Indeed, these flows are located at the edges of the application area and do not have upstream flows. Hence, they only depend on their historical values and cannot exploit the full potential of the model.

The relatively good performance of historical average seems surprising, but it reflects some day-to-day regularity of the flows. This is also the case for the passenger flows leaving controlled areas to public or other controlled areas, for which the forecasting results of this method are approximately equivalent to those of our approach (average WMAPEs of 22.2% and 22.6% respectively).

VII. CONCLUSION AND FUTURE WORK

In this paper, we propose a dynamic Bayesian network approach for short-term urban rail passenger flow forecasting, which is able to provide real-time predictions in case of incomplete data. Based on the work of Sun et al. [19], our model exploits the spatiotemporal neighborhood of the flows to predict their future values. We extend this principle to public transport networks by proposing a structure that integrates the transport service. In the presence of incomplete data, we use the structural EM algorithm to learn both the structure and the parameters. Then short-term forecasting is performed efficiently by the bootstrap filter. The results of the experiment demonstrate the overall effectiveness of the approach and highlight the key role of the transport service.

Despite these encouraging results, our approach still has limitations. First, the choice of conditional linear Gaussian distributions implies that the relationships between the flows (and the transport service variables) are linear, which can be questionable. Therefore, we may wonder whether the use of distributions that allow to model nonlinear processes, such as Gaussian mixture models, would be more appropriate.

Another drawback is that the structure and the parameters of the model do not evolve over time. Yet in very disturbed conditions, the interactions between the flows may change to patterns that have never been observed (and thus learned). The difficulties of the model in fitting these situations may result in decreased forecasting performance.

As shown by the results of the experiment, our approach still needs to be improved, especially for the flows located at the edges of the application area. In future work, we could introduce new factors, such as temporal features (e.g., day of the week, month of the year, vacation) and external conditions (e.g., weather, sporting or cultural events). In order to investigate the performance of the model more deeply, we should also extend the application area to the flows related to other lines of the transport network and collected during other periods.

REFERENCES

[1] J. W. C. van Lint and C. P. I. J. van Hinsbergen, "Short-Term Traffic and Travel Time Prediction Models," *Artificial Intelligence Applications to Critical Transportation Issues*, vol. 22, pp. 22–41, 2012.

[2] C. P. I. J. van Hinsbergen, J. W. C. van Lint, and F. M. Sanders, "Short Term Traffic Prediction Models," in *Proceedings of the 14th World Congress on Intelligent Transport Systems*, Beijing, China, 2007, pp. 5013–5030.

[3] B. L. Smith and M. J. Demetsky, "Traffic Flow Forecasting: Comparison of Modeling Approaches," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 261–266, 1997.

[4] B. M. Williams, "Modeling and forecasting vehicular traffic flow as a seasonal stochastic time series process," Ph.D. dissertation, University of Virginia, 1999.

[5] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box-Jenkins techniques," *Transportation Research Record*, vol. 722, pp. 1–9, 1979.

[6] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research Part B: Methodological*, vol. 18, no. 1, pp. 1–11, 1984.

[7] J. Whittaker, S. Garside, and K. Lindveld, "Tracking and predicting a network traffic process," *International Journal of Forecasting*, vol. 13, no. 1, pp. 51–61, 1997.

[8] S. Clark, "Traffic Prediction Using Multivariate Nonparametric Regression," *Journal of Transportation Engineering*, vol. 129, no. 2, pp. 161–168, 2003.

[9] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transportation Research Part C: Emerging Technologies*, vol. 66, pp. 61–78, 2016.

[10] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211–234, 2005.

[11] K. Kumar, M. Parida, and V. K. Katiyar, "Short term traffic flow prediction for a non urban highway using Artificial Neural Network," *Procedia - Social and Behavioral Sciences*, vol. 104, pp. 755–764, 2013.

[12] Z. Ma, J. Xing, M. Mesbah, and L. Ferreira, "Predicting short-term bus passenger demand using a pattern hybrid approach," *Transportation Research Part C: Emerging Technologies*, vol. 39, pp. 148–163, 2014.

[13] H. B. Celikoglu and H. K. Cigizoglu, "Public transportation trip flow modeling with generalized regression neural networks," *Advances in Engineering Software*, vol. 38, no. 2, pp. 71–79, 2007.

[14] Y. Wei and M.-C. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 148–162, 2012.

[15] Q. Li, Y. Qin, Z. Wang, Z. Zhao, M. Zhan, Y. Liu, and Z. Li, "The Research of Urban Rail Transit Sectional Passenger Flow Prediction Method," *Journal of Intelligent Learning Systems and Applications*, vol. 5, no. 4, pp. 227–231, 2013.

[16] M. Zhong, P. Lingras, and S. Sharma, "Estimation of missing traffic counts using factor, genetic, neural, and regression techniques," *Transportation Research Part C: Emerging Technologies*, vol. 12, no. 2, pp. 139–166, 2004.

[17] L. Qu, L. Li, Y. Zhang, and J. Hu, "PPCA-Based Missing Data Imputation for Traffic Flow Volume: A Systematical Approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 512–522, 2009.

[18] J. Haworth and T. Cheng, "Non-parametric regression for space-time forecasting under missing data," *Computers, Environment and Urban Systems*, vol. 36, no. 6, pp. 538–550, 2012.

[19] S. Sun, C. Zhang, and G. Yu, "A Bayesian Network Approach to Traffic Flow Forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 124–132, 2006.

[20] M. E. Whitlock and C. M. Queen, "Modelling a Traffic Network with Missing Data," *Journal of Forecasting*, vol. 19, no. 7, pp. 561–574, 2000.

[21] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann Publishers, 1988.

[22] P. Leray, "Réseaux bayésiens : Apprentissage et diagnostic de systemes complexes," Accreditation to supervise research, Université de Rouen, 2006.

[23] R. D. Shachter and R. C. Kenley, "Gaussian influence diagrams," *Management Science*, vol. 35, no. 5, pp. 527–550, 1989.

[24] T. Dean and K. Kanazawa, "A Model for Reasoning About Persistence and Causation," *Computational Intelligence*, vol. 5, no. 3, pp. 142–150, 1989.

[25] K. Murphy, "Dynamic Bayesian Networks: Representation, Inference and Learning," Ph.D. dissertation, University of California, 2002.

[26] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorith," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[27] J. A. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," International Computer Science Institute, Tech. Rep., 1998.

[28] N. Friedman, "Learning Belief Networks in the Presence of Missing Values and Hidden Variables," in *Proceedings of the 14th International Conference on Machine Learning*, Nashville, TN, 1997, pp. 125–133.

[29] ——, "The Bayesian Structural EM Algorithm," in *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, Madison, WI, 1998, pp. 129–138.

[30] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[31] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning Bayesian Networks: The Combination of Knowledge and Statistical Data," *Machine Learning*, vol. 20, no. 3, pp. 197–243, 1995.

[32] N. Friedman, K. Murphy, and S. Russel, "Learning the Structure of Dynamic Probabilistic Networks," in *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, Madison, WI, 1998, pp. 139–147.

[33] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings-F - Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, 1993.

[34] K. Kanazawa, D. Koller, and S. Russel, "Stochastic simulation algorithms for dynamic probabilistic networks," in *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, Montreal, Canada, 1995, pp. 346–351.

[35] P. Pan and D. Schonfeld, "Visual Tracking Using High-Order Particle Filtering," *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 51–54, 2011.