Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 26 (2017) 53–61

Transportation Research Procedia

www.elsevier.com/locate/procedia

44th European Transport Conference 2016, ETC 2016, 5-7 October 2016, Barcelona, Spain

A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data

Jérémy Roos [a,b,*], Gérald Gavin [b], Stéphane Bonnevay [b]

[a] *RATP, 75012 Paris, France*
[b] *Université de Lyon, ERIC EA 3083, 69100 Villeurbanne, France*

Abstract

We propose an approach to forecast the short-term passenger flows of the urban rail network of Paris. Based on dynamic Bayesian networks, this approach is designed to perform even in case of incomplete data. The structure of the model is built so that the flows are predicted from their spatiotemporal neighbourhood, while the local conditional distributions are described as linear Gaussians. A first application carried out on a single station highlights the need to incorporate information on the transport service. In the presence of missing data, we perform the structural expectation-maximization (EM) algorithm to learn both the structure and the parameters of the model. Short-term forecasting is conducted by inference via the bootstrap filter. Finally, we apply the model to an entire Paris metro line, using on-board counting, ticket validation and transport service data. The overall forecasting results outperform the historical average and last observation carried forward (LOCF) methods. They also evidence the key role of the transport service in the modeling.

Peer-review under responsibility of the Association for European Transport

Keywords: short-term passenger flow forecasting; urban rail network; dynamic Bayesian networks; incomplete data.

1. Introduction

RATP is the main public transport operator in the Paris region. It operates 16 metro lines, sections of 2 RER (commuter rail), 8 tramway lines and more than 350 bus lines. Currently, it uses several tools for passenger flow modeling, including GLOBAL and IMPACT (Leblond and Garcia Castello, 2016). These models are mainly

* Corresponding author.
E-mail address: jeremy.roos@ratp.fr

Peer-review under responsibility of Association for European Transport
10.1016/j.trpro.2017.07.008

designed to assess the effects of infrastructure or transport policy changes over the long term. Hence, they do not provide short-term predictions and cannot take into account the effects of unanticipated or non-recurring events, such as service disruptions, closures of stations or crowd-attracting events. On the other hand, the diversity of available data is still largely untapped. Each service often works with a limited number of data sources and only has a partial view of the passenger mobility within the network. Moreover, these services have to face with incomplete data caused by failures or lack of collection systems. This issue leads to the implementation of imputation methods which can be time-consuming.

From these observations, we propose a model to forecast the short-term passenger flows of the RATP urban rail network (metro and RER). The chosen approach is based on dynamic Bayesian networks, which are able to combine heterogeneous information sources and make predictions even in case of missing data. The model is designed to cater for various applications in transport system management, such as operation planning, passenger flow regulation or passenger information.

In this paper, we first present a brief state of the art of short-term traffic forecasting. After introducing the Bayesian network representation, we perform a quick experiment on a single RER station with a complete set of ticket validation and counting data. Then we extend the model to the time factor, using dynamic Bayesian networks, and incorporate information on the transport service. The learning and inference processes are detailed for incomplete data and the application area is extended to an entire metro line. Finally, we conclude the paper and discuss the limitations of our approach.

2. Short-term traffic forecasting: brief state of the art

There is a vast scientific literature on short-term traffic forecasting. Various methods have been explored, which can be classified into naïve, parametric and nonparametric methods (van Hinsbergen et al., 2007). Due to their easy implementation, naïve methods have been widely used, including historical average (Smith and Demetsky, 1997) and last observation carried forward (LOCF), also called random walk (Williams, 1999). Among the parametric methods, ARIMA models have been explored since the late 1970’s (Ahmed and Cook, 1979). A great focus has also been given to Kalman filter algorithms (Okutani and Stephanedes, 1984). In recent years, nonparametric methods such as nonparametric regression (Clark, 2003) and neural networks (Vlahogianni et al., 2005) have been successfully applied. Their ability to better model nonlinear processes has contributed to their popularity.

In his thesis, Haworth (2014) points out the lack of consensus on the most suitable method. Indeed, each dataset has its own characteristics (e.g., spatiotemporal resolution, transport network, type of data), which makes comparison to other studies difficult. Moreover, for time-saving purposes, the authors generally compare their sophisticated model to basic implementations of other models.

Missing data is a common problem in real-world situations, including traffic flow forecasting. Surprisingly, few authors have dealt with this problem in a real-time setting (Haworth and Cheng, 2012). Hence, most of the short-term forecasting models are ill-equipped to provide real-time predictions in case of incomplete data. Among the few solutions that have been proposed, particular attention has been afforded to Bayesian networks, both on urban road networks (Sun et al., 2006) and on highways (Whitlock and Queen, 2000).

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 (Ma et al., 2014). Some authors have recently begun to tackle this issue, especially through neural network approaches (Celikoglu and Cigizoglu, 2007; Wei and Chen, 2012; Li et al., 2013). In these models, the flows are generally predicted on the basis of their historical values. The input data may also include other features, such as information from adjacent flows (Li et al., 2013), temporal factors (Wei and Chen, 2012; Li et al., 2013) or weather conditions (Li et al., 2013). Although the service provided by the public transport operator has a potential impact on the passenger flows, it seems that this information has not yet been considered.

3. Bayesian network representation

3.1. Bayesian networks

Introduced by Pearl (1988), Bayesian networks are probabilistic graphical models that describe the conditional dependencies (and independencies) between random variables. These dependencies are represented by a directed acyclic graph and "quantified" by a joint probability distribution, which decomposes into a product of local conditional distributions:

$$p(X_1, \dots, X_n) = \prod_{i=1}^{n} p\left(X_i \middle| \mathrm{Pa}_{X_i}\right) \tag{1}$$

where Pa_{X_i} is the set of parents of X_i. The flexibility of Bayesian networks allows to combine heterogeneous information sources. As an example, an expert can set a part of the model from his own knowledge, whereas the other part can be learnt from data (Leray, 2006). Furthermore, the information propagation mechanism makes inference (i.e., the forecasting process) possible even in case of incomplete data. This property is particularly useful in a real-time setting, where the implementation of an additional imputation process may be detrimental.

When the variables are continuous, the conditional probability distributions of a Bayesian network can be described as linear Gaussians (Shachter and Kenley, 1989):

$$p\left(X_i \middle| \mathrm{Pa}_{X_i}\right) = \mathcal{N}\left(\beta_{i0} + \beta_i^{\mathsf{T}} \mathrm{Pa}_{X_i}, \sigma_i^2\right) \tag{2}$$

where β_{i0}, β_i and σ_i are the parameters to be estimated for each i, usually by maximum likelihood. Note that choosing this type of distribution implies assuming that the interactions between the variables are linear.

3.2. Application to passenger flow forecasting

According to Sun et al. (2006), there is an intuitive causal relationship between the flows measured at given points and those located downstream in the transport network. Thus, by considering that each flow depends on its adjacent upstream flows, the structure of the Bayesian network derives directly from the topology of the transport network. We first test this principle on passenger flows of the urban rail station Nanterre-Préfecture (RER line A). The experiment is carried out on the basis of a complete set of ticket validation and counting data collected during weekdays and aggregated by 10 minutes. We focus on two families of flows:

- the flow of passengers moving from the entrance hall to the north-east wing of the station (manually counted), predicted from two upstream ticket validation flows (264 observations collected on December 5, 2013, March 11 and 27, 2014);
- the flow of passengers boarding trains from Cergy – Le Haut / Poissy to Paris (manually counted), predicted from three upstream flows: the passengers moving from the entrance hall to the north-east wing of the station and two other ticket validation flows (143 observations collected on March 11 and 27, 2014).

Given the limited number of available observations, the forecasting performance is evaluated by 10-fold cross-validation (Kohavi, 1995). This method consists in randomly dividing the dataset $\mathcal{X}$ into 10 subsets $\mathcal{X}_1, \dots, \mathcal{X}_{10}$ of (approximately) equal size. For each $k \in \{1, \dots, 10\}$, the model is trained on $\mathcal{X} \setminus \mathcal{X}_k$ and tested on $\mathcal{X}_k$. The final performance is estimated by averaging the 10 obtained accuracy measures. In this paper, we adopt the weighted mean absolute percentage error (WMAPE):

$$WMAPE(x, \hat{x}) = \frac{\sum_{t=1}^{N} \left| x^{(t)} - \hat{x}^{(t)} \right|}{\sum_{t=1}^{N} x^{(t)}} \tag{3}$$

where $\hat{x}$ is the estimate of x and N is the number of observations in the dataset. Like the mean absolute percentage error (MAPE), the WMAPE is easy to interpret. On the other hand, it favors models that are effective in predicting the high values of the flows.

When we focus on the passengers moving from the entrance hall to the north-east wing of the station, the model provides encouraging results, with an estimated WMAPE of 16.3% (see Fig. 1(a)). However, the performance significantly decreases when considering the passengers boarding trains from Cergy – Le Haut / Poissy to Paris, with an estimated WMAPE of 48.2%. As can be seen in Fig. 1(b), the large fluctuations of this flow are poorly fitted by the model. These fluctuations are closely related to the transport service. Indeed, if there is no train at the platform, the flow is interrupted (troughs) and the passengers have to wait for the next train before boarding (peaks). Thus, although the results of this experiment illustrate the potential of the Bayesian network approach, they also highlight the need to incorporate information on the transport service.

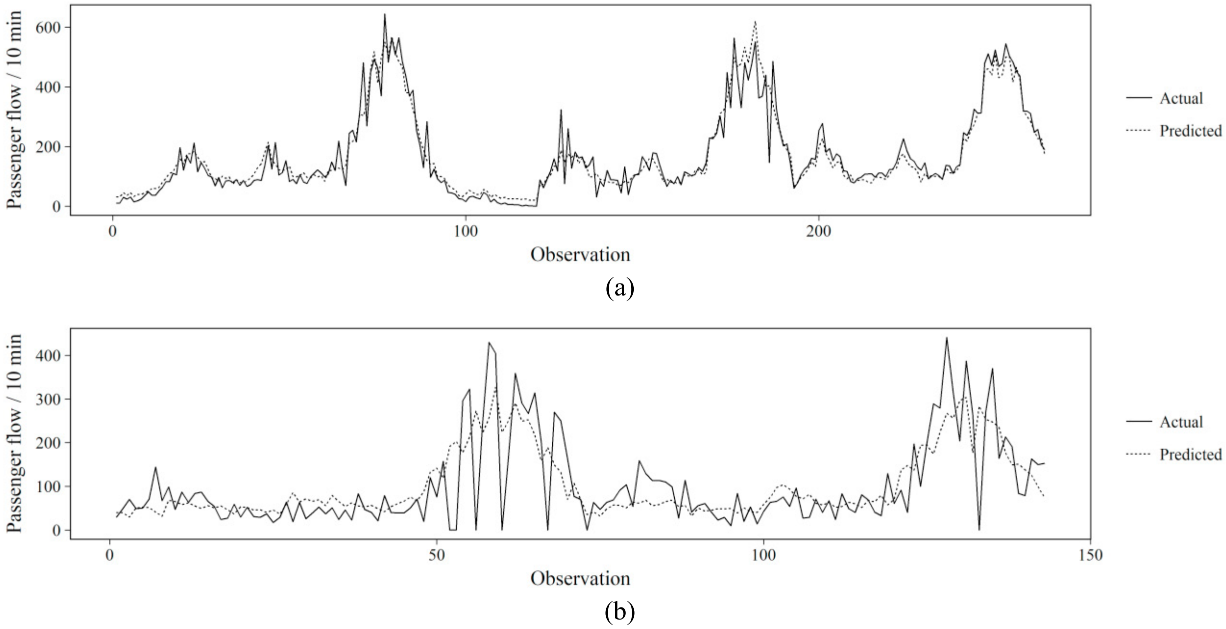


Fig. 1. Actual and predicted values of two flows in Nanterre-Préfecture: (a) passengers moving from the entrance hall to the north-east wing; (b) passengers boarding trains from Cergy – Le Haut / Poissy to Paris.

4. Dynamic Bayesian network approach

4.1. Spatiotemporal modeling

Dynamic Bayesian networks extend Bayesian networks to model systems that evolve over time (Dean and Kanazawa, 1989). Each node $X_i^{(t)}$ represents the instantiation of the variable X_i at time slice t. The parents of $X_i^{(t)}$ can belong to $t, t-1, \ldots, t-r$, where r is the order of the dynamic Bayesian network (Ghahramani, 1998).

In the previous experiment, we predict the passenger flows using their upstream flows at the same time slice. To forecast their future values, it is necessary to extend this principle to the time factor. Following Sun et al. (2006), we define a parameter r_u such that each flow at time slice t depends on its adjacent upstream flows at $t-1, \ldots, t-r_u$. Assuming that the historical values of the flows provide information on their trend, we define another parameter r_h

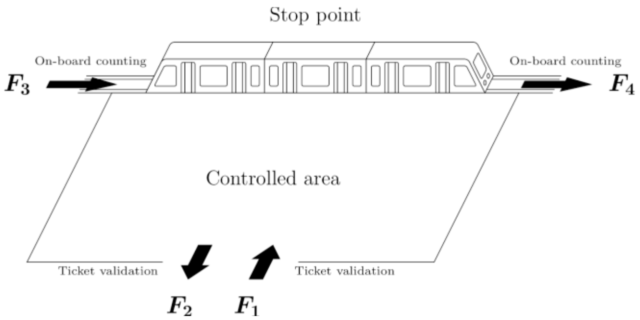
such that each flow at time slice t depends on its values at $t-1,\dots,t-r_h$. These dependencies are represented by a dynamic Bayesian network of order $r=\max(r_u,r_h)$.

As evidenced by the previous results, the passenger flows are closely related to the transport service. To incorporate this type of information, we rely on the following observation: the longer the time interval between two trains, the more passengers board the latter. Let F be a flow of passengers traveling between two successive stop points and $D_F^{(t)}$ the set of departure times from the stop point of origin of F during time slice t (several departures may occur during the same time slice). We define the "transport service variable" associated with F at t:

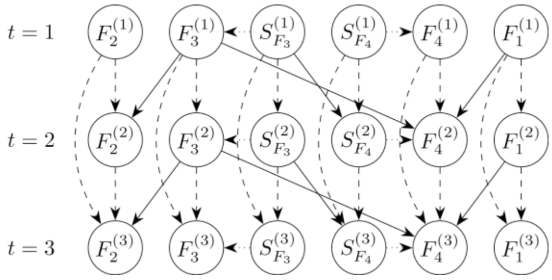
$$S_F^{(t)}=\begin{cases}\max D_F^{(t)}-\max\bigcup_{k<t} D_F^{(k)}, & \text{if } D_F^{(t)}\neq\emptyset\\ 0, & \text{otherwise}\end{cases} \tag{4}$$

where $\max D_F^{(t)}$ and $\max\bigcup_{k<t} D_F^{(k)}$ are the latest departure times during and before t respectively. $S_F^{(t)}$ is equivalent to the sum of the time intervals that precede the departures occurring during t. In the dynamic Bayesian network, it directly influences $F^{(t)}$. Similarly to a passenger flow, it also depends on its adjacent upstream transport service variables and on its historical values.

Fig. 2 provides an example of passenger flows with their corresponding dynamic Bayesian network unrolled over three time slices. For the sake of clarity, we take low order parameters: $r_u=1$ (solid arcs in Fig. 2(b)) and $r_h=2$ (dashed arcs). In this example, F_1 is upstream of F_4, and F_3 is upstream of F_2 and F_4. F_3 and F_4 are directly influenced by the transport service (dotted arcs).



(a)



(b)

Fig. 2. (a) Example of passenger flows. (b) Corresponding dynamic Bayesian network with $r_u=1$ and $r_h=2$, unrolled over three time slices.

4.2. Learning with incomplete data

With a complete dataset (like in the previous experiment), the parameters of a linear Gaussian Bayesian network can be easily estimated by maximum likelihood. However, in case of incomplete data, the traditional estimation method is no longer able to provide analytical solutions. In such a vast public transport network, the missing data are too scattered to use listwise deletion. In their paper, Sun et al. (2006) propose to replace the variables whose values are missing by their parents in the dynamic Bayesian network. Unfortunately, this method is hardly applicable because it implies that the parents are complete, which is not necessarily true in many situations.

Proposed by Dempster et al. (1977), the expectation-maximization (EM) algorithm is a method for iteratively finding the maximum likelihood estimates of the parameters when the training dataset has missing (or hidden) values. Starting from an initial guess of the parameters, this method performs two steps at each iteration. In the expectation (E) step, it completes the dataset from the observed data and the current estimates of the parameters. In the maximization (M) step, it uses this completed dataset to update the parameters by maximizing the log-likelihood.

As proven by Dempster et al. (1977), the log-likelihood increases at each iteration until convergence to a local maximum.

Depending on the order parameters r_u and r_h, the number of arcs (and thus of parameters) in the dynamic Bayesian network can be huge. In addition to increasing the computational complexity, this situation may lead to overfitting and decrease the forecasting performance. To avoid these problems, we try to reduce the dimension of the model by selecting the best subset of arcs among those described in the previous subsection. The selection of an optimal substructure in the presence of incomplete data can be performed by the structural EM algorithm. Proposed by Friedman (1997) and extended to dynamic Bayesian networks by Friedman et al. (1998), this algorithm performs similarly to the parametric version described above. The difference comes in the M step, where it improves not only the parameters but also the structure of the model. In the same way, the structural EM algorithm is guaranteed to increase a scoring function until convergence to a local maximum (Friedman, 1997). In this paper, we adopt the Bayesian information criterion (BIC) score (Schwarz, 1978), which takes into account the log-likelihood while penalizing the complexity of the structure.

4.3. Short-term forecasting

The short-term forecasting process consists in determining the passenger flows at the next time slice, which can be regarded as an inference problem in the dynamic Bayesian network. A comprehensive review of inference methods for dynamic Bayesian networks can be found in the thesis of Murphy (2002) or in the more recent book of Koller and Friedman (2009). The approximate inference methods are generally less time-consuming than the exact methods. Given the complexity of our model, they appear to be a better choice to ensure real-time predictions.

The bootstrap filter (Gordon et al., 1993), also known as survival of the fittest (Kanazawa et al., 1995), is a stochastic simulation algorithm which can perform inference in real time. It consists in generating weighted sample sequences by sampling the unobserved values. These sequences are propagated forward in time proportionally to their weight, which reflects their likelihood for the measures collected over time. Initially applied to first-order dynamic models (Gordon et al., 1993; Kanazawa et al. 1995), the bootstrap filter has been extended to higher-order models (Pan and Schonfeld, 2011).

5. Large-scale experiment

5.1. Experimental method

We apply the dynamic Bayesian network approach to Paris metro line 2. More precisely, we forecast the flows at the next time slice, harnessing three types of information:

- on-board flows of line 2 (48 flows) and connecting lines (12 flows), recorded at train departures by on-board weighing systems;
- passenger flows entering or leaving controlled areas (i.e., areas requiring a valid ticket) served by line 2 (30 flows) and connecting controlled areas (5 flows), recorded by ticket validation;
- the transport service operated for line 2 (48 variables) and connecting lines (66 variables).

The data are collected during 33 weekdays from March 2 to April 17, 2015, between 7.30 and 9.30 am (morning peak) and aggregated by 2 minutes, totalling 1980 observations. Due to failures of collection systems and to the fact that some trains are equipped with weighing systems, 80 of the 95 passenger flows are incomplete. Their overall missing data rate is 4.8%, but reaches more than 50% for some flows (this rate is only 0.2% for the transport service variables). In addition to this high spatial dispersion, the missing data cover nearly all time slices recorded (99.7%).

The dataset is divided into a training and a test set, which contain the data of the first 24 days and of the last 9 days respectively. First, we learn the structure and the parameters of the dynamic Bayesian network from the training set, using the structural EM algorithm (see subsection 4.2). In this experiment, we take as order parameters $r_u = r_h = 4$, which empirically provide good forecasting results. The performance of the model is evaluated on the test set, using the bootstrap filter (see subsection 4.3).

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 appear optimistic, since unexpected changes may occur at the last moment. To be more realistic, these values should be preset to their "expectation" at $t-1$. However, this information is not stored and thus cannot be used in this experiment.

5.2. Forecasting results

Table 1 gives the average forecasting errors for the 48 on-board passenger flows of line 2 and the 30 flows entering or leaving controlled areas served by line 2. In addition to the complete version of the model, two partial versions are tested to assess the individual contribution of each type of relationship: a version without transport service and a version without relationships between the adjacent flows. We also compare the results to those of two naïve methods: historical average, which consists in forecasting the flows by averaging their historical values at the corresponding time slices (Smith and Demetsky, 1997), and LOCF, which simply uses the last observed value (Williams, 1999).

Table 1. Average forecasting errors for the passenger flows related to Paris metro line 2 (WMAPE in %).

Passenger flows	Dynamic Bayesian network			Historical average	LOCF
	(complete)	(w/o transport service)	(w/o adj. relationships)		
On-board	17.1	36.2	19.3	39.6	63.1
Entering/leaving controlled areas	18.9	19.0	19.2	17.2	24.8
All passenger flows	17.8	29.6	19.3	31.0	48.4

Overall, the dynamic Bayesian network approach outperforms the historical average and LOCF methods, with average WMAPEs of 17.8% versus 31.0% and 48.4% respectively. The transport service largely contributes to these good forecasting results. Indeed, when incorporating this type of information, the average WMAPE for the on-board flows drops from 36.2% to 17.1%. The ability of the model to fit the large fluctuations of these flows is well illustrated by Fig. 3, which shows the actual and predicted values of the flow from Blanche to Place de Clichy (WMAPE of 17.0%) for the test days from April 7 to 9, 2015.

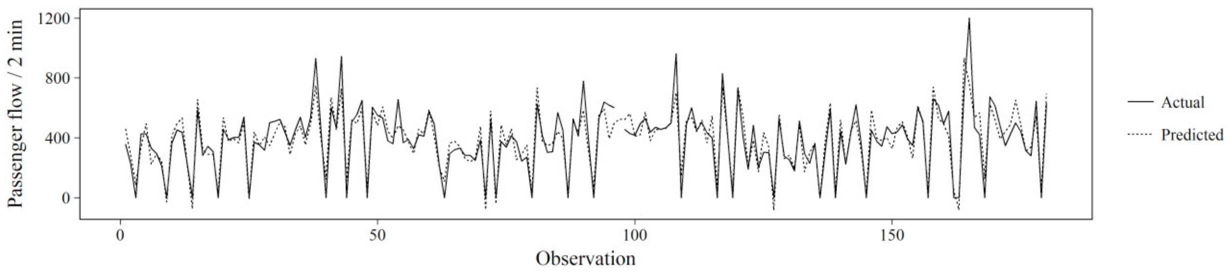


Fig. 3. Actual and predicted values of the passenger flow from Blanche to Place de Clichy, from April 7 to 9, 2015, between 7.30 and 9.30 am.

The forecasting results also highlight the important role of the relationships between the adjacent flows. Their contribution is clearly visible for the on-board flows, the average WMAPE dropping from 19.3% to 17.1%. However, we observe that our approach is less effective than historical average for the flows entering or leaving controlled areas (average WMAPE of 18.9% versus 17.2%). One possible reason lies in the relative position of these flows in the application area. Indeed, either they are on the edge and have no upstream flows, or they are in wide interchange stations where too few of their upstream flows are recorded. Hence, they cannot exploit the full potential of the model. On the other hand, their day-to-day regularity may explain the good results of historical average.

6. Conclusion

In this paper, we propose a dynamic Bayesian network approach to forecast the short-term passenger flows of the urban rail network of Paris. This approach is designed to provide real-time predictions even in case of incomplete data. Based on Sun et al. (2006), the flows are predicted from their spatiotemporal neighbourhood. As evidenced by the first test on Nanterre-Préfecture, they also depend on the transport service. Thus, we propose a structure that incorporates this information. In the presence of incomplete data, the structural EM algorithm is performed both to reduce the dimension of the structure and to find the maximum likelihood estimates of the parameters. Then the bootstrap filter is applied for short-term forecasting. The experiment carried out on Paris metro line 2 provides encouraging results and confirms the key role of the transport service.

Despite its good overall performance, our approach still has limitations. First, the choice of conditional linear Gaussian distributions simplifies the computation but implies that the interactions between the variables are linear, which is 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. In case of major disruptions, the relationships between the variables may change to patterns that have never been observed, which may lead to decreased forecasting performance.

In future work, we could improve the performance of the model by introducing new factors, such as temporal features (e.g., day of the week, month, vacation) or external conditions (e.g., weather, sporting or cultural events). As pointed out by Leray (2006), one of the strengths of Bayesian networks is their high modularity, which allows the gradual incorporation of new information sources and the construction of increasingly sophisticated models.

References

Ahmed, M. S., Cook, A. R. (1979). Analysis of freeway traffic time-series data by using Box-Jenkins techniques. *Transportation Research Record*, 722, 1–9.

Celikoglu, H. B., Cigizoglu, H. K. (2007). Public transportation trip flow modeling with generalized regression neural networks. *Advances in Engineering Software*, 38(2), 71–79.

Clark, S. (2003). Traffic Prediction Using Multivariate Nonparametric Regression. *Journal of Transportation Engineering*, 129(2), 161–168.

Dean, T., Kanazawa, K. (1989). A Model for Reasoning About Persistence and Causation. *Computational Intelligence*, 5(3), 142–150.

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorith. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38.

Friedman, N. (1997). Learning Belief Networks in the Presence of Missing Values and Hidden Variables. *Proceedings of the 14th International Conference on Machine Learning*, 125–133.

Friedman, N., Murphy, K., Russel, S. (1998). Learning the Structure of Dynamic Probabilistic Networks. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 139–147.

Ghahramani, Z. (1998). Learning Dynamic Bayesian Networks. *Adaptive Processing of Sequences and Data Structures*. Springer-Verlag, 168–197.

Gordon, N. J., Salmond, D. J., Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F - Radar and Signal Processing*, 140(2), 107–113.

Haworth, J., Cheng, T. (2012). Non-parametric regression for space-time forecasting under missing data. *Computers, Environment and Urban Systems*, 36(6), 538–550.

Haworth, J. (2014). *Spatio-temporal forecasting of network data*. Ph.D. dissertation, University College London.

Kanazawa, K., Koller, D., Russel, S. (1995). Stochastic simulation algorithms for dynamic probabilistic networks. *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, 346–351.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1137–1143.

Koller, D., Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Leblond, V., Garcia Castello, F. (2016). GLOBAL et IMPACT, deux outils complémentaires pour la planification des transports. *Rencontres de la Mobilité Intelligente 2016*.

Leray, P. (2006). *Réseaux bayésiens : Apprentissage et diagnostic de systemes complexes*. Accreditation to supervise research, Université de Rouen.

Li, Q., Qin, Y., Wang, Z., Zhao, Z., Zhan, M., Liu, Y., Li, Z. (2013). The Research of Urban Rail Transit Sectional Passenger Flow Prediction Method. *Journal of Intelligent Learning Systems and Applications*, 5(4), 227–231.

Ma, Z., Xing, J., Mesbah, M., Ferreira, L. (2014). Predicting short-term bus passenger demand using a pattern hybrid approach. *Transportation Research Part C: Emerging Technologies*, 39, 148–163.

Murphy, K. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning*. Ph.D. dissertation, University of California.

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

Pan, P., Schonfeld, D. (2011). Visual Tracking Using High-Order Particle Filtering. *IEEE Signal Processing Letters*, 18(1), 51–54.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461–464.

Shachter, R. D., Kenley, R. C. (1989). Gaussian influence diagrams. *Management Science*, 35(5), 527–550.

Smith, B. L., Demetsky, M. J. (1997). Traffic Flow Forecasting: Comparison of Modeling Approaches. *Journal of Transportation Engineering*, 123(4), 261–266.

Sun, S., Zhang, C., Yu, G. (2006). A Bayesian Network Approach to Traffic Flow Forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 7(1), 124–132.

van Hinsbergen, C. P. I. J., van Lint, J. W. C, Sanders, F. M. (2007). Short Term Traffic Prediction Models. *Proceedings of the 14th World Congress on Intelligent Transport Systems*, 5013–5030.

Vlahogianni, E. I., Karlaftis, M. G., Golias, J. C. (2005). Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C: Emerging Technologies*, 13(3), 211–234.

Wei, Y., Chen, M.-C. (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies*, 21(1), 148–162.

Whitlock, M. E., Queen, C. M. (2000). Modelling a Traffic Network with Missing Data. *Journal of Forecasting*, 19(7), 561–574.

Williams, B. M. (1999). *Modeling and forecasting vehicular traffic flow as a seasonal stochastic time series process*. Ph.D. dissertation, University of Virginia.