Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2023) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2023 Montreal 17-21 July 2023

Modeling Transportation Time Series using Bayesian Dynamic Linear Models

Saeid Amiri[a], James-A. Goulet[a], Martin Trépanier[a], Nicolas Saunier[a,*]

[a] *Civil, Geological and Mining Engineering Department, C.P. 6079, succ. Centre-Ville, Montreal H3T1J4, Canada*

Abstract

Sudden large-scale changes like the recent COVID-19 pandemic make the management and planning of transport systems difficult, despite the ever-increasing availability of data. The primary goal of this work is to model transportation data time series using a dynamic model that is interpretable and can be used for long-term forecasting. The Bayesian Dynamic Linear Model (BDLM) is chosen because it can account for complex data and can be easily adapted to the data. A component for the BDLM is introduced to recognize the underlying patterns using temporal control points. A moving-event component is also added to take into account events that do not occur on the same date every year such as sports games. The proposed model is parsimonious and can learn from the data. After providing a brief summary of the theory of the model, experimental results are shown for transport demand data obtained from smart card transaction data for the Montreal public transit system. The proposed model is compared to different time series models and shows superior accuracy.

Keywords: Dynamic linear models; Forecasting; Model uncertainty; Transportation data

1. Introduction

With the rapid growth of urban populations, transportation systems face increasing challenges to improve their sustainability and provide efficient options to all segments of the population. One key to improving transportation systems is data. Transportation professionals and decision-makers need data about travel demand, negative impacts such as congestion, accidents, and pollution, so that they can manage the transportation system and provide sustainable and integrated travel options. They need models to forecast demand in order to plan the expansion of transportation systems. In recent years, more and more data have become available from different sources such as management systems, for example from smart card validations for the transit network, from global navigation satellite system (GNSS) devices on vehicles like taxis or other personal devices like smart phones, and from social network usage.

In this paper, we propose a machine learning approach with a new extension for Bayesian Dynamic Linear Models (BDLM) to learn and predict transportation time series data subject to punctual events. This approach is demonstrated

* Corresponding author. Tel.: +1-514-340-4711, ext. 4962.
E-mail address: nicolas.saunier@polymtl.ca

on long-term passenger load forecasting obtained from smart card transaction data. In contrast to other methods proposed for transportation data, BDLMs 1) can naturally handle missing data and outliers, 2) provide an uncertainty for the forecast, 3) can be updated incrementally online, 4) can be used to simulate the transportation behavior under new conditions, and 5) can deal with events with changing dates like yearly festivals. BDLMs are a special case of State-Space Models (SSM). BDLMs have become widely used to model time series with various dynamic processes. BDLMs have a simple structure, where components can be added to model data with a complex generating process. The variables involved in these components are known as hidden states because they are not directly observed. Unlike conventional statistical models, the hidden states can be easily estimated in complex models with many components, by using the recursive procedure known as the Kalman filter.

The contribution of this paper is to introduce a novel approach for transportation time series modeling by defining a set of hidden states that vary over time, and to extract the non-harmonic changing patterns. This manuscript is structured as follows: section 2 presents a short review of past relevant work, then section 3 describes the methodology, including a brief summary of the BDLM and the method to construct a kernel regression and its calendar variant developed for transportation data. Section 4 presents and discusses an application of the proposed model to subway passenger demand data for three years in Montreal. Section 4.2 describe the event component to model the impact of events on passenger demand: the performance of the proposed model is tested using synthetic and real events and the results show that the proposed component increases the forecasting accuracy.

2. Background

Research is very active in transportation time series data modeling. A review of the literature on short-term passenger flow prediction can be found in (Vlahogianni et al., 2014). A hybrid of linear regression and seasonal autoregressive integrated moving average model (SARIMA) was proposed by Ni et al. (2016) for passenger flow prediction using social media data. Wang et al. (2016) considered a combination of Least Squares Support Vector Machine (SVM) and Radial Basis Function (RBF) neural networks for passenger flow forecasting. A new multi-scale RBF neural network was proposed and compared to SVMs and boosted regression tree for passenger flow forecasting using smart card data from the Beijing subway network in (Li et al., 2017).

Dynamic Bayesian network have also been applied to the short-term forecast of passenger flows of the Paris transit network (Roos et al., 2016). Newer deep learning methods have also been applied. Toqué et al. (2016) applied Long Short-Term Memory (LSTM), Recurrent Neural Networks and the Vector AutoRegressive (VAR) model to forecast dynamic Origin-Destination (OD) matrices in the Paris subway network using smart card data. Later, they studied the performance of gated recurrent unit (GRU) and recurrent neural networks, compared to random forests for the short-term forecasting problem in (Toqué et al., 2018).

In contrast to the aforementioned research, this work intends to forecast the passenger flow using BDLMs. Although simpler SSMs have already been applied to transportation data, for example for short-term traffic flow prediction in (Noursalehi et al., 2018), to the best of the authors' knowledge, BDLMs have not been applied to long-term passenger load forecasting. There has been a large and rapidly growing body of literature over the past decade on BDLMs, both theoretical and applied, for example in (West and Harrison, 2006), (Durbin and Koopman, 2012), (Aoki, 2013), and (Douc et al., 2014), among others.

3. Methodology

3.1. Bayesian Dynamic Linear Models

The objective is to model uni-dimensional time series $y_t \in \mathbb{R}, t \in \{t_1, \ldots, t_{n_T}\}$, where n_T is the sample size. A BDLM is defined by the following pair of equations constituting respectively the observation and transition models:

$$y_t = \mathbf{C}_t \mathbf{x}_t + v_t, \quad v_t \sim \mathcal{N}(0, R_t), \text{ observation model} \tag{1}$$

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t), \text{ transition model} \tag{2}$$

where $\mathbf{A}_t$ and $\mathbf{C}_t$ are the transition and observation matrices, v_t and $\mathbf{w}_t$ are the Gaussian measurement and model errors, with respective variance R_t and covariance matrix $\mathbf{Q}_t$, respectively. The four terms $\mathbf{A}_t$, $\mathbf{C}_t$, R_t and $\mathbf{Q}_t$ define the model and involve unknown parameters θ to be learnt from the data. The observation model links the hidden states to the observations, and the transition model governs the change in the hidden states over time. The unknown parameters θ can be learned using the maximum likelihood estimation,

$$\begin{aligned}\widehat{\theta} &= \arg\max_{\theta} \ln(f(y_{1:T}|\theta)) = \arg\max_{\theta}\left(\sum_{t=1}^{n_T}\ln(f(y_t|y_{1:t-1}))\right)\\ &= \arg\max_{\theta}\left(\sum_{t=1}^{n_T}\ln(N(y_t; C_t\mu_{t|t-1}, C_t\Sigma_{t|t-1}C_t^T + R_t))\right).\end{aligned} \tag{3}$$

where $\mathbf{y}_{1:t} = \{y_1, y_2 \ldots, y_t\}$, $\mu_{t|t-1} = E(\mathbf{X}_t|\mathbf{y}_{1:t-1})$ is the prior expected value, $\mu_{t|t} = E(\mathbf{X}_t|\mathbf{y}_{1:t})$ the posterior expected value, $\mathbf{\Sigma}_{t|t} = V(\mathbf{X}_t|\mathbf{y}_{1:t-1})$ the prior covariance matrix and $\mathbf{\Sigma}_{t|t} = V(\mathbf{X}_t|\mathbf{y}_{1:t})$ is the posterior covariance matrix. The maximum can be found using a gradient-based optimization method such as Newton-Raphson. The Kalman Filter algorithm consists of two successive steps: the prediction step which estimates the state at the current time step based on the previous one and the updating step which refines the state estimate based on the current observation. The prediction step consists of the following calculations:

$$\begin{aligned}\mu_{t|t-1} &= \mathbf{A}_t\mu_{t-1|t-1},\\ \mathbf{\Sigma}_{t|t-1} &= \mathbf{A}_t\mathbf{\Sigma}_{t-1|t-1}\mathbf{A}_t^T + \mathbf{Q}_t.\end{aligned} \tag{4}$$

The updating step follows:

$$\begin{aligned}\mu_{t|t} &= \mu_{t|t-1} + \mathbf{K}_t r_t,\\ r_t &= y_t - \widehat{y}_t,\\ \widehat{y}_t &= \mathbf{C}_t\mu_{t|t-1},\\ \mathbf{\Sigma}_{t|t} &= (\mathbf{I} - \mathbf{K}_t\mathbf{C}_t)\mathbf{\Sigma}_{t|t-1},\\ \mathbf{K}_t &= \mathbf{\Sigma}_{t|t-1}\mathbf{C}_t^T\mathbf{G}_t^{-1},\\ \mathbf{G}_t &= \mathbf{C}_t\mathbf{\Sigma}_{t|t-1}\mathbf{C}_t^T + R_t.\end{aligned} \tag{5}$$

At the initial time, $t = 0$, $\mu_0 = E(\mathbf{X}_0) =$ and $\mathbf{\Sigma}_0 = V(\mathbf{X}_0)$ constitute our prior knowledge and can be estimated using the Newton-Raphson method.

3.2. Generic Components

The linear model structure is built from generic components that are already available for different phenomena and the BDLM (Goulet, 2017). In the following, the two components used to model the transportation data are presented: the Local Level (LL) and Calendar Covariate Kernel regression (CKR). The final model using these two components can be expressed as:

$$\begin{aligned}y_t &= \begin{bmatrix}\mathbf{C}_t^{LL} & \mathbf{C}_t^{CKR}\end{bmatrix}\begin{bmatrix}\mathbf{x}_t^{LL}\\ \mathbf{x}_t^{CKR}\end{bmatrix} + v_t, \quad v_t \sim \mathcal{N}(0, R_t),\\ \begin{bmatrix}\mathbf{x}_t^{LL}\\ \mathbf{x}_t^{CKR}\end{bmatrix} &= \begin{bmatrix}\mathbf{A}_t^{LL} & \mathbf{0}\\ \mathbf{0} & \mathbf{A}_t^{CKR}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{t-1}^{LL}\\ \mathbf{x}_{t-1}^{CKR}\end{bmatrix} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{Q}_t^{LL} & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_t^{CKR}\end{bmatrix}\right).\end{aligned} \tag{6}$$

The Local Level component represents the baseline response of the structure without the effect of external solicitation. The local level generic formulation is defined by

$$\mathbf{x}_t^{LL} = x^{LL}, \mathbf{A}_t^{LL} = 1, \mathbf{C}_t^{LL} = 1, \mathbf{Q}_t^{LL} = (\sigma^{LL})^2, \tag{7}$$

and it assumes a constant quantity over time.

Here we first outline the Kernel Regression (KR) component introduced by Nguyen et al. (2019). It is developed for modeling periodic components. These components employ N hidden variables via a kernel function between the current time and N reference variables which we refer to as control points. The mathematical formulas of the kernel regression are presented below,

$$
\begin{aligned}
\mathbf{x}^{KR} &= [x_0^{KR}, x_1^{KR}, \ldots, x_N^{KR}]^T, &(8)\\
\mathbf{A}_t^{KR} &= \begin{bmatrix} \mathbf{0} \; \tilde{k}(t, t_1^{KR}), \ldots, \tilde{k}(t, t_i^{KR}), \ldots, \tilde{k}(t, t_N^{KR}) \\ \mathbf{0} \qquad\qquad\qquad \mathbf{I} \end{bmatrix},\\
\mathbf{Q}^{KR} &= \begin{bmatrix} \sigma_0^{KR} & \mathbf{0} \\ \mathbf{0} & (\sigma_W^{KR})^2 \mathbf{I}_{N\times N} \end{bmatrix},\\
\mathbf{C}^{KR} &= [1, 0, \ldots, 0],\\
\mathbf{t}^{KR} &= [t_1^{KR}, \ldots, t_N^{KR}]^T,\\
\tilde{k}(t, t_i^{KR}) &= \frac{k(t, t_i^{KR})}{\sum_{i=1}^{N} k(t, t_i^{KR})},\\
k(t, t_i^{KR}) &= \exp\left(-\frac{2}{\ell^2} \sin\left(\pi \frac{t - t_i^{KR}}{p}\right)^2\right), &(9)
\end{aligned}
$$

where k is the kernel function and $\tilde{k}$ the normalized kernel function, p and ℓ are the period and kernel length scale parameters, respectively. The model with the kernel allows to model time series with periodic phenomena. $\mathbf{t}^{KR}$ has an important role in the kernel regression, because it includes a vector of time covariates associated with a vector of hidden control point values and the normalized kernel values measure the similarity between the time t and those times recorded as the control points that enable the model to recognize the patterns observed in the training data. For an in-depth treatment of kernel function we direct the interested reader to the reviews by Duvenaud (2014), Henderson and Parmeter (2015), and Racine et al. (2008), while for an applied approach to using the kernel technique we direct the reader to (Nguyen et al., 2019) and the references therein.

Although the BDLM using equation 8 can be applied to data with a periodic pattern, it is not appropriate for transportation data, mainly because the pattern of passenger loads changes over years, hence the kernel functions need to be elaborated. In order to make the model recognize these changes, we redefine the time in terms of day of the week and week of the year: each day of the year can thus be defined by a tuple (d, w) where $d \in \{1, 2, 3, 4, 5, 6, 7\}$ and $w \in \{1, 2, \ldots, 53\}$. The kernel function in equation 9 can be adapted as:

$$k(t, t_i^{CKR}) = \exp\left(-\frac{\pi}{2\ell_d^2}(d - d_i^{CKR})^2\right). \exp\left(-\frac{\pi}{2\ell_w^2}(w - w_i^{CKR})^2\right), \tag{10}$$

where ℓ_d and ℓ_w are the new kernel length scale parameters. We refer to this component as the calendar covariate kernel regression (CKR) hereafter. The mathematical formulas, matrices $\mathbf{C}^{CKR}$, $\mathbf{A}_t^{CKR}$, $\mathbf{Q}^{CKR}$, and variance R^{CKR} of this component are the same as the kernel regression presented in equations 8. To create the timestamp for this component, we consider all the possible combinations of days and weeks observed in the training set. For years 2015 and 2016 there are 369 possible combinations, so we set $N = 369$ for this model. We can also consider fewer control points by creating a uniform 2-D grid from d and w. Assigning fewer control points however might drop the accuracy of CKR significantly, as shown in the next section.

3.3. Moving-Event Model

Events like festivals or holidays might have a significant impact on the transportation passenger loads: they cause an increase in or displacement of passenger load. If the events happen on the same days each year, a reliable model might recognize them. Some of them however change date every year: for instance Easter always falls on a Sunday between March 22 and April 25, and the Montreal International Jazz Festival was held between June 26, 2015 and July 5, 2015, June 29, 2016 and July 9, 2016, and June 28, 2017 and July 8, 2017. Such events take place on specific dates, and will take place on scheduled dates in the future, so there is a possibility of redefining the proposed model in a way to capture the behavior of data on those dates. There has been attempts to model the moving event scenarios in variants of the computerized ARIMA model: the U.S. Census Bureau developed a particular algorithm, X-12-ARIMA, to adjust ARIMA for counting the effect the moving event scenarios (Findley et al., 1998; Roberts and Holan, 2010). The proposed model CKR can generalize for modeling events: since we know the dates of the events, we can add the hidden states in the model to pick out the pattern from the event dates and use them for forecasting. The model, when taking into account events, can be expressed as below

$$y_t = \begin{bmatrix} \mathbf{C}_t^{LL} & \mathbf{C}_t^{CKR} & C_t^{EC} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t^{LL} \\ \mathbf{x}_t^{CKR} \\ x_t^{EC} \end{bmatrix} + v_t, \tag{11}$$

$$\begin{bmatrix} \mathbf{x}_t^{LL} \\ \mathbf{x}_t^{CKR} \\ x_t^{EC} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_t^{LL} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_t^{CKR} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_t^{EC} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1}^{LL} \\ \mathbf{x}_{t-1}^{CKR} \\ x_{t-1}^{EC} \end{bmatrix} + \mathbf{w}_t,$$

$$v_t \sim \mathcal{N}(0, R_t), \qquad \mathbf{w}_t \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Q}_t^{LL} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_t^{CKR} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_t^{EC} \end{bmatrix} \right),$$

where EC denotes the event component. This model is referred to as CKREC (CKR+EC) in the rest of work. Since it deals with events, we define C_t^{EC} as

$$C_t^{EC} = \begin{cases} 1 & t \in \text{event date} \\ 0 & \text{otherwise} \end{cases}. \tag{12}$$

This component allows x_t^{EC} to pick out the pattern from specific dates, and for the future occurrences of the event, the component can be used to transfer the pattern. It is an important feature of the BDLM that, unlike other models, it does not require sophisticated techniques to add a component to the body of the model to improve its performance.

4. Experimental Results

To evaluate and compare the proposed model, we worked on real data from the Montreal public transport agency "Société de Transport de Montréal" (STM) and used daily passenger loads from the Montreal metro stations. The STM has four metro lines, the orange, green, blue and yellow lines, with 31, 27, 12 and 3 stations, respectively. The Berri-UQAM station has the highest passenger traffic as it is a junction of the green, yellow, and orange lines. Figure 1 shows the daily passenger loads at the Berri-UQAM station in Montreal over a three-year period, from 2015 to 2017. The underlying pattern is a combination of weekly and monthly periodicity. There are also changes over the years, which can be attributed to different reasons, for example because each year starts on a different day of week: 2015 started with a Tuesday, 2016 started with a Friday, and 2017 started on a Monday. The passenger loads change greatly from one day of the week to the next: the average proportion of passenger loads for each day from Monday to Sunday are 0.1425, 0.1731, 0.1681, 0.1723, 0.1583, 0.1003, and 0.0854, respectively. This shows that the passenger loads are typically very high on weekdays and much lower on the weekend. Public holidays and events have a huge impact on passenger loads, and this should be properly accounted in the model. Note that the number of days in each year is different: 2015 and 2017 have 365 days, while 2016 has 366.

4.1. Generic Components

In this section, we compare the performance of the proposed technique with a few existing time series models: SARIMA, Neural Network Time Series (nnetar), and the time series linear model (tslm) (Hyndman et al., 2019). The focus of this work is not, however, to compare the performance of different time series model on a data set, but rather to explore the potential of BDLM and to develop new components to present a reliable machine learning model. The computation and modeling of BDLM rely on the open-source OpenBDLM platform[1] (Gaudot et al., 2019). Other methods such as Gaussian Processes, SVMs, and LSTMs were also tested, but are not presented because of their poor performance and our focus on models developed for time series.

[1] OpenBLDM is available at `http://github.com/CivML-PolyMtl/OpenBDLM`.

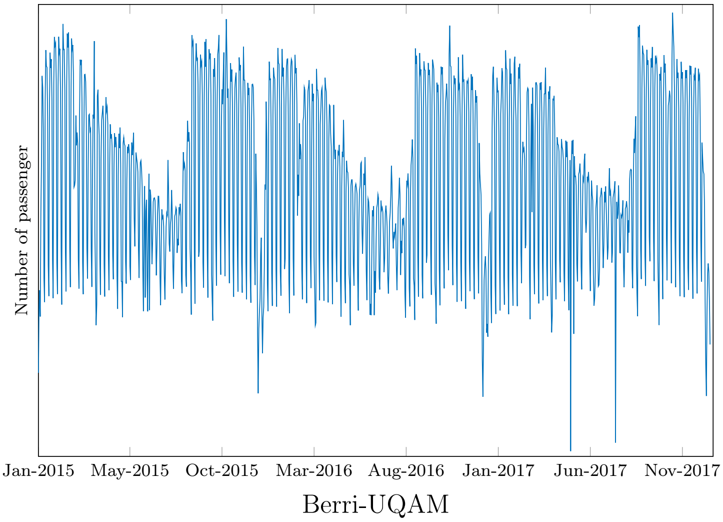

Fig. 1. Daily passenger load at the Berri-UQAM station at Montreal. (Note: the metro system was accessible for free to the public for two days for events taking place in July and August in 2017, which is why very small passenger loads are recorded on these two days.)

Table 1. Comparison of MAER for the different stations and models

Model	Selected Stations					
	Berri-UQAM	McGILL	University of Montreal	Lionel-Groulx	Snowdon	Jean-Talon
SARIMA	52.63	70.65	151.16	46.94	50.99	38.92
nnetar	78.67	119.56	288.06	109.55	76.86	76.53
tslm	78.00	121.19	227.84	65.05	81.45	61.81
CKR(100)*	947.89	99.75	532.74	212.24	249.46	208.30
CKR(200)*	35.98	70.89	279.80	209.73	251.36	212.94
CKR(369)*	**30.24**	**42.60**	**70.64**	**25.54**	**39.68**	**23.26**

* : the number inside parenthesis is the number of control points for CKR.

From this dataset, we processed the daily passenger traffic of years 2015-2016 as the training set, using 2017 as the test set: the models are fitted on the data of years 2015-2016 and the forecasting is done on 2017. To quantify the performance of the proposed method, we used the mean absolute error rate (MAER) on the test sets, defined as $MAER = \frac{1}{n}\sum\frac{|\widehat{y}_t - y_t|}{y_t}$. The MAERs of different fitted models for six metro stations are presented in Table 1. It shows that the BDLM with calendar covariate kernel regression with 369 control points, CKR(369), provides a better accuracy than the other models. In order to present a comprehensive picture of the performance of proposed method, Figure 2 depicts the forecast of passenger loads using CKR(369) vs the observed passenger loads for different stations, it shows the CKR(369) has a very good performance.

4.2. *Moving-Event model*

To evaluate the accuracy of proposed model, we added a simulated event: we assume a moving event happened on 2015/9/15, and on 2016/10/18 in the training set, and will happen on 2017/11/14 in the test set. We add a change (positive or negative) to the Berri-UQAM station passenger loads on 2015/9/15 and 2016/10/18, and observe the change that will occur on 2017/11/14. The change (positive or negative) is calculated as $\frac{New-old}{old}$. The results are presented in Table 2 and show the model can recognize the underlying patterns, enabling it to easily make a reliable forecast.

To test the proposed event model on a real moving event, we consider the games of the National Hockey League's Montreal Canadiens. Their games are held at the Bell Centre and the closest metro stations to this centre are Lucien-L'Allier and Bonaventure. The schedules and number of games change over the years. For example, there were games

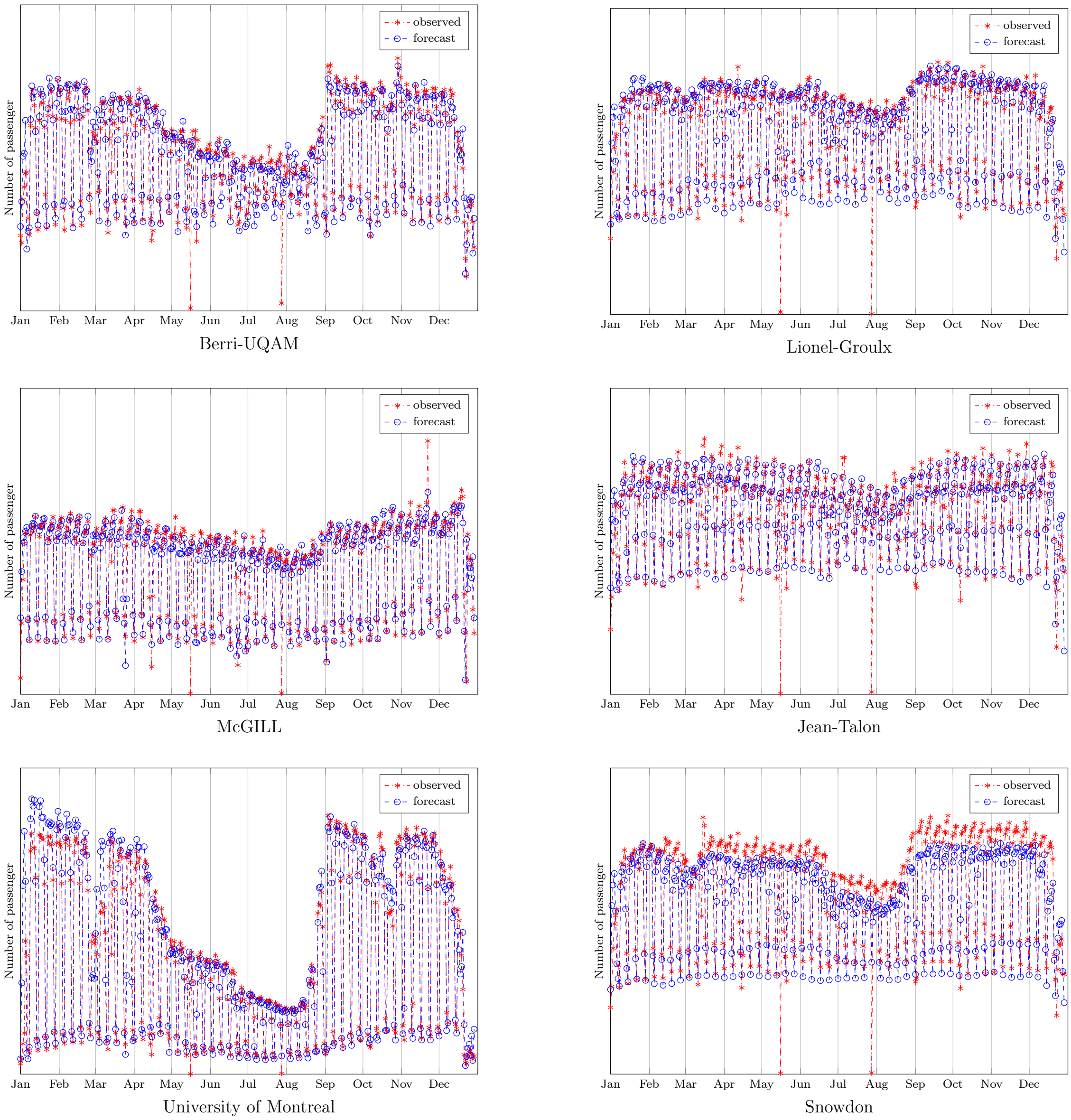

Fig. 2. Forecast vs observed passenger load using CKR(369) for different stations

in May 2015, but not in May of 2016 and 2017 To compare the model, we considered the mean absolute error (MAE) (difference between the true values and the forecasts) from the testing set and the difference between log-likelihood (LL) of model with testing test and training sets. The fitted CKR and CKREC have respective MAEs of 5,271.03 and 3,257.10. The LL of CKR and CKREC are -3,637.27 and -3,619.74, respectively. These demonstrate CKREC provides a better model in forecasting.

5. Conclusion

This work attempted to establish a new general inferential basis for modeling the transportation data using BDLMs. We presented a new component named calendar covariate kernel regression that can be used for the transportation data,

Table 2. Performance of event component on the simulated event on 2015/9/15, 2016/10/18, and 2017/11/14

	Change recognized in Model			Change recognized in Model	
True change	CKR	CKREC	True change	CKR	CKREC
-2.5%	0.5	-1.2	+2.5%	0.6	3.6
-5.0%	0.6	-3.6	+5.0%	0.6	6.0
-7.5%	0.6	-5.6	+7.5%	0.6	8.4
-10%	0.6	-8.5	+10%	0.6	10.8
-15%	0.6	-13.5	+15%	0.6	15.6
-20%	0.6	-28.3	+20%	0.6	20.4
-30%	0.6	-29.5	+30%	0.6	30.6
-40%	0.6	-37.1	+40%	0.6	40.2

and our model accounted for time by breaking it down into the day of week and the week of year instead of defining time as the existing BDLM model does. We compared the performance of the proposed method with other time series models and found that its forecast accuracy is promising. Although forecast accuracy is an important criterion, we prefer to obtain a model that is more interpretative in practice.

The disadvantages of this method are its computation time for a large number of control points: the calendar covariate kernel regression demands a large number of control points, N, to learn the underlying pattern, but increasing the number of control points increases the computational burden. We also presented the moving event component, which enables the model to pick out the patterns of an event or phenomena observed in the training sets and use them for forecasting.

Changes or disruption of service are issues for transport demand forecasting: the moving event model can be applied to learn the impact of such changes. The BDLM components presented in this work pave the way for more advanced modeling of transport data. Future work will involve a unified modeling perspective 1) to add spatio-temporal covariates such as weather in the model, and 2) to develop algorithms for online learning.

6. Acknowledgements

The authors wish to acknowledge the financial support of IVADO through its fundamental research funding program. They also would like the STM for sharing the data used for this research.

References

Aoki, M. (2013). *State space modeling of time series*. Springer Science & Business Media.

Douc, R., Moulines, E., and Stoffer, D. (2014). *Nonlinear time series: Theory, methods and applications with R examples*. Chapman and Hall/CRC.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford university press.

Duvenaud, D. (2014). *Automatic model construction with Gaussian processes*. PhD thesis, University of Cambridge.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., and Chen, B.-C. (1998). New capabilities and methods of the x-12-arima seasonal-adjustment program. *Journal of Business & Economic Statistics*, 16(2):127–152.

Gaudot, I., Nguyen, L. H., Khazaeli, S., and Goulet, J.-A. . (2019). Openbdlm, an open-source software for structural health monitoring using bayesian dynamic linear models. In *In Proceedings from the 13th International Conference on Applications of Statistics and Probability in Civil Engineering*.

Goulet, J.-A. (2017). Bayesian dynamic linear models for structural health monitoring. *Structural Control and Health Monitoring*, 24(12):e2035.

Henderson, D. J. and Parmeter, C. F. (2015). *Applied nonparametric econometrics*. Cambridge University Press.

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., and Razbash, S. (2019). Package forecast.

Li, Y., Wang, X., Sun, S., Ma, X., and Lu, G. (2017). Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. *Transportation Research Part C: Emerging Technologies*, 77:306–328.

Nguyen, L. H., Gaudot, I., Khazaeli, S., and Goulet, J.-A. (2019). A kernel-based method for modeling non-harmonic periodic phenomena in bayesian dynamic linear models. *Frontiers in Built Environment*, 5:8.

Ni, M., He, Q., and Gao, J. (2016). Forecasting the subway passenger flow under event occurrences with social media. *IEEE Transactions on Intelligent Transportation Systems*, 18(6):1623–1632.

Noursalehi, P., Koutsopoulos, H. N., and Zhao, J. (2018). Real time transit demand prediction capturing station interactions and impact of special events. *Transportation Research Part C: Emerging Technologies*, 97:277–300.
Racine, J. S. et al. (2008). Nonparametric econometrics: A primer. *Foundations and Trends in Econometrics*, 3(1):1–88.
Roberts, C. G. and Holan, S. H. (2010). Comparison of x-12-arima trading day and holiday regressors with country specific regressors. *Journal of Official Statistics*, 26(2):371.
Roos, J., Bonnevay, S., and Gavin, G. (2016). Short-term urban rail passenger flow forecasting: A dynamic bayesian network approach. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1034–1039. IEEE.
Toqué, F., Côme, E., El Mahrsi, M. K., and Oukhellou, L. (2016). Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1071–1076. IEEE.
Toqué, F., Côme, E., Oukhellou, L., and Trépanier, M. (2018). Short-term multi-step ahead forecasting of railway passenger flows during special events with machine learning methods.
Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43:3–19.
Wang, P., Wu, C., and Gao, X. (2016). Research on subway passenger flow combination prediction model based on rbf neural networks and lssvm. In *2016 Chinese Control and Decision Conference (CCDC)*, pages 6064–6068. IEEE.
West, M. and Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.