Wireless Airtime Traffic Estimation Using a State Space Model

Farzaneh Kohandani
Elect. and Comp. Eng.
University of Waterloo
Waterloo, Ontario, Canada
farzaneh@cst.uwaterloo.ca

Derek W. McAvoy
Bell Mobility
Mississauga, Ontario, Canada
derek.mcavoy@bell.ca

Amir K. Khandani
Elect. and Comp. Eng.
University of Waterloo
Waterloo, Ontario, Canada
khandani@cst.uwaterloo.ca

Abstract

A new forecasting technique called the extended structural model (ESM) is presented. This technique is derived from the basic structural model (BSM) by the introduction of extra parameters that were assumed to be 1 in the BSM. The ESM model is constructed from the training sequence using the standard Kalman filter recursions, and then the extra parameters are estimated to minimize the mean absolute percentage error (MAPE) of the validation sequence. The model is evaluated by prediction of the total number of minutes of wireless airtime per month on the Bell Canada network. The ESM model shows an improvement in MAPE of the test sequence over both the BSM and seasonal autoregressive integrated moving average. The improved prediction can significantly reduce the cost for wireless service providers, who need to accurately predict future wireless spectrum requirements.

Keywords: basic structural model; autoregressive integrated moving average; Kalman filter; mean absolute percentage error.

1 Introduction

Accurate forecasting is crucial in many organizations and plays an important role in most decision-making processes [1]. Companies rely on the demand predictions to justify the investment, and to ensure the resources are correctly allocated [4]. To generate predictions, information about past events, called "time series data", is collected. There are four general pattern types for time series data: stationary, trend, seasonal, and cyclical [9].

A stationary series is one, whose basic statistical properties is constant over time. The trend pattern is the long-term component which represents the growth or decline of the time series over a long period of time. A seasonal pattern corresponds to a series that repeats every year, while a cyclical pattern repeats every two, three, or more years.

A number of naive forecasting techniques use a very simple model to predict the future values from the past. Some of these methods simply use the last observed value to predict the future value [9]. For stationary data, the moving average model is a suitable forecasting method, which makes the prediction based on the mean computed from the most recent data [5]. Exponential smoothing is a simple forecasting method, where the forecast is constructed from an exponentially weighted average of all the past observations [31]. Exponential smoothing model can be used to forecast stationary data; however, it does not work well for data with trend and seasonality [31].

Autoregressive integrated moving average (ARIMA) methods are more advanced models that integrate an autoregressive model and a moving average model [24, 37]. This modelling is a very powerful tool that provides accurate forecasts and compared to the previous models does not assume any particular pattern in time series data. However, the disadvantage of this model is that it needs a relatively large amount of data to develop the model. In addition, there is no easy way to update the parameters as new data becomes available, because the model has to be completely refitted.

Another advanced class of methods use the structural state space models, which captures the observations as a sum of separate components (such as trend and seasonality) and can be used to forecast the stationary, trend, seasonal, and cyclical data [22]. This is a very powerful technique in forecasting that can easily update the parameters of the model as new data becomes available, which is better than the ARIMA technique that needs to completely refit the model.

In this paper, we present a state space model called extended structural model (ESM) as an extension of the basic structural model (BSM) [16] to predict the total number of minutes of airtime per month on the Bell Canada network. The main objective of this work is to predict the network load and to justify future allocation of the available spectrum. Even a small improvement in the prediction of the

required spectrum can result in a substantial cost saving for the service provider. Underestimation in the required spectrum demand results in a network that fail to handle the load, and overestimation results in unnecessary and wasteful spending to acquire the spectrum.

A plot of the data traffic against time easily reveals the existence of a trend and a seasonal component. Seasonality component which is equal to 12 months (one year period) comes from the fact that special occasions like mother's day in the month of May will result in an increase in the wireless data traffic at the same time every year. Both ARIMA and state space modelling are able to capture the behavior of this traffic. Because of the advantages that state space modelling offers over ARIMA, such as capability to update the parameters as the new data becomes available, state space modelling has been selected.

To estimate the unknown parameters of the model, maximum log-likelihood and minimum mean absolute percentage error (MMAPE) criteria are adopted. Kalman filter is used to compute the log-likelihood function recursively and optimization techniques are applied to maximize the log-likelihood function to estimate the underlying parameters. Table 1 presents the notations used in this paper.

This paper is organized as follows. After the introduction in Section II, some of the related work is discussed. Section III describes the data traffic model, and the extended structural model is described in Section IV. Numerical results and comparisons are provided in Section V, and paper concludes with summary remarks in Section VI.

2 Related Work

A number of other researchers investigated the problem of predicting the seasonal time series. Magalhaes et al. proposed an approach to combine three different forecasting methods to predict the seasonal stream flow time series in southeast region in Brazil [23]. The three different combined techniques are periodic autoregressive moving average (PARMA) model and two fuzzy clustering-based forecasting models. They are combined by a simple neural network trained with the steepest descent algorithm. The work provides results for a four-year prediction using 55 years (660 points) as the training set.

Guang et al. presented a seasonal neural network model to forecast the seasonal internet traffic of eastern China, CERNET [8]. This model consists of three layers, input layer, hidden layer and output layer. The data samples are divided to different sets such that every set contains one period of data and are used to train the neural network using back propagation technique. The length of the training data set is 30 days and the calibration period is 10 days, and since the traffic is collected hourly, the total length of the training data is 720.

Young presented an unobserved component model that captures the effect of a vector of exogenous variables [34, 36]. This model is used to restore non-stationary audio signal of a recorded piano sample. The stochastic varying parameters of the model are described by generalized random walk process, and the estimation of the parameters are performed by maximum log-likelihood estimation using the Kalman filter. The parameter optimization is performed in the frequency domain. An autoregressive spectrum of the observation process is estimated to find the linear least square estimate of the objective function. Kalman filter recursions are performed to restore the signal, where the total number of the samples were 3000.

ARIMA technique is used to forecast NSFNET internet traffic, where the traffic is seasonal and non-stationary [7]. A non-stationary series can often be transformed into a stationary series by subtracting the data from the data in the previous period. To capture the seasonality, a seasonal form of ARIMA (p, d, q) technique is used, where p is the order of the autoregressive part, d is the order of the differencing and q is the order of the moving average part, respectively. All the data analysis is done by S-PLUS [30] statistical package, where the appropriate values of p and q are determined by inspecting both autocorrelation and partial autocorrelation functions. There are three diagnostic tests described in [2], defined as standardized residuals, autocorrelation plot of the residuals and goodness of fit statistics. The training data is from August 1988 through June 1992 and the forecasting is from July 1992 through June 1993.

Hansen and Nelson proposed a neural network model to forecast revenues of Utah legislative session in 1997 [10]. There are two economic measurements which play significant role in revenue forecast: the growth rate for nonagricultural employment and the total taxable sales. The first measurement is a cyclical pattern of variable period, and the second measurement is a quarterly seasonal pattern. The prediction is done by using time-delay neural network and the application of genetic algorithm. The growth rate of nonagricultural employment measurement has the training data from 1981-96 and the forecast is performed for a period of one year. The taxable sale measurement has the training data from 1978 to 1996 and a prediction of one year.

Tych et al. present a customized package, CRESFOR for forecasting and signal extraction [32]. This technique was developed at Lancaster University to forecast up to three weeks in advance of the hourly telephone calls made to Barclaycard place. The model is an enhancement of dynamic harmonic regression (DHR) model [35]. DHR model is the extension of the BSM model [11] but is different in regards of the nature of the unobserved components, their state space representation and optimization of hyper-parameters. The training data is over a period of one year 98/99 and the prediction period is 70 realization points in one week.

Table 1. Table of notations

Symbol	Description
X_t	observation of the data traffic
μ_t	trend component of the data traffic
γ_t	seasonal component of the data traffic
ϵ_t	measurement noise of the data traffic
$\boldsymbol{\alpha}_t$	state vector $(r \times 1)$
$\mathbf{H}$	observation matrix $(1 \times r)$
$\boldsymbol{\Phi}$	transition matrix $(r \times r)$
$\mathbf{K}$	state noise coefficient matrix $(r \times n)$
$\boldsymbol{\theta}_t$	state noise vector $(n \times 1)$
η_t	noise of the trend component
ω_t	noise of the seasonality component
ζ_t	noise of the trend slope
β_t	trend slope
L	likelihood function of the time series X_t
N	total number of observations
$\overline{X}_t$	mean of the conditional distribution $X_t\|X_1, X_2, \dots, X_{t-1}$
V_t	prediction error of the conditional distribution
F_t	variance of V_t
$\mathbf{a}_{t\|t-1}$	MMSE estimate of state $\boldsymbol{\alpha}_t$ at time $t-1$ $(r \times 1)$
$\mathbf{C}_{t\|t-1}$	covariance matrix of the estimate of $\boldsymbol{\alpha}_t$ at time $t-1$ $(r \times r)$
$\mathbf{a}_t$	MMSE estimate of state $\boldsymbol{\alpha}_t$ at time t $(r \times 1)$
$\mathbf{C}_t$	covariance matrix of the estimate of $\boldsymbol{\alpha}_t$ at time t $(r \times r)$
$\boldsymbol{\Sigma}$	covariance matrix of the state noise $\boldsymbol{\theta}_t$ $(n \times n)$
$\sigma_\eta^2, \sigma_\omega^2, \sigma_\epsilon^2$	noise variances
b, c, d, e, f	extra parameters defined for the ESM model

Nelson et al. investigated if the neural network is capable of modelling the data with seasonality [25]. They determined that deseasonalising the data before applying the neural network model is better than using the data directly. Later, Prochazka [27] used a neural network to construct a forecasting method without the need to deseasonalise the data. To process the data, [27] first applied a wavelet transform and then used the neural network to forecast in the wavelet domain.

Econometric models which are completely different from previous models can forecast economic parameters and are suitable for cyclical data. Since economic data may exist in both a time series and cross-sectional form (data between two countries at the same time), we can also combine these two types of information in the statistical models [17]. There are four basic models: fixed effect model, random effect model, random coefficient model and coefficients as functions of exogenous variables [6, 14, 15, 18]. These techniques are suitable for applications with short time series and combined with cross-sectional data. Several other forecasting techniques have also been proposed to forecast time series data [3, 12, 13, 21, 28, 29].

3 Data Traffic Model

3.1 Time Series Definition

Time series is a sequence of observations $X_1, X_2, \dots, X_{t-1}, X_t$ of a random process X at discrete time intervals, where an observation at time t is given by X_t. Time series are often described as having trends and seasonal terms. Each observation X_t is defined as the summation of three components: trend, seasonal, and measurement noise, and is given as [16]:

$$X_t = \mu_t + \gamma_t + \epsilon_t, \tag{1}$$

where μ_t is the trend component at time t, γ_t is the seasonal component at time t, and ϵ_t is the measurement noise at time t.

3.2 State Space Model

The observation sequence $X_1, \dots, X_t$ can be described by a state space model:

$$X_t = \mathbf{H}\boldsymbol{\alpha}_t + \epsilon_t \tag{2}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\Phi}\boldsymbol{\alpha}_{t-1} + \mathbf{K}\boldsymbol{\theta}_t. \tag{3}$$

The state space model is described in terms of three matrices $\mathbf{H}$, $\boldsymbol{\Phi}$ and $\mathbf{K}$. At time t, the past observations of X are encapsulated by a state vector $\boldsymbol{\alpha}_t$, and any future observations are assumed to be independent of the past observations, given this state vector. The state vector is composed of r state variables $\alpha_t^{(1)}, \dots, \alpha_t^{(r)}$, such that $\boldsymbol{\alpha}_t = (\alpha_t^{(1)}, \dots, \alpha_t^{(r)})^T$.

The current state $\boldsymbol{\alpha}_t$ is calculated from the previous state $\boldsymbol{\alpha}_{t-1}$ by the state equation (3), where $\boldsymbol{\Phi}$ is a $r \times r$ matrix of the parameters, called the transition matrix, and $\mathbf{K}$ is a $r \times n$ matrix, called the state noise coefficient matrix. The state noise $\boldsymbol{\theta}_t$ is an n-dimensional zero-mean Gaussian noise vector with the covariance matrix

$$\boldsymbol{\Sigma} = E\{\boldsymbol{\theta}_t\boldsymbol{\theta}_t^T\} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2). \tag{4}$$

The observation X_t is calculated from the state vector $\boldsymbol{\alpha}_t$ by the observation equation (2). Here $\mathbf{H}$ is a $1 \times r$ observation vector and ϵ_t is the measurement noise, defined as a white Gaussian noise with zero mean and variance of σ_ϵ^2.

3.3 Basic Structural Model

The time series defined in (1) leads to the simplest (with fewest state variables) corresponding state space model as

[16]:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$
$$\beta_t = \beta_{t-1} + \zeta_t$$
$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t, \qquad (5)$$

with the state vector defined as

$$\boldsymbol{\alpha}_t = (\mu_t, \beta_t, \gamma_t, \gamma_{t-1}, \ldots, \gamma_{t-s+1})^T. \qquad (6)$$

Given the equations in (5) and the state vector in (6), the corresponding state space matrices $\boldsymbol{\Phi}$, $\mathbf{K}$ given in (3) can be derived. In this model, s is the season period, β_t is the trend slope. The parameters η_t, ζ_t, and ω_t are noise of the trend component, noise of the trend slope, and noise of the seasonality component, respectively. These parameters are mutually uncorrelated white Gaussian random variables with a zero mean and variances of σ_η^2, σ_ζ^2, and σ_ω^2.

This is a well-known model in statistical forecasting and has not been used in the area of telecommunications before. Our data traffic shows an increasing linear trend with seasonality factor equal to a period of 12 months. This model is selected by inspection since it can capture the linear trend and seasonality of the data in a simple way. Other models with higher degree of the recursive equations and larger number of states have also been studied; however, there was not much improvement in the performance, although the models were more complicated.

3.4 Model Parameter Estimation

The state space model is defined by the equations (2) and (3) and one needs to estimate some or all of the unknown parameters $\mathbf{H}$, σ_ϵ^2, $\boldsymbol{\Phi}$, and $\mathbf{K}$. To estimate the unknown parameters, the maximum log-likelihood criterion is used. The likelihood of a time series X can be decomposed into the product of the conditional distributions of each X_t on its predecessors, i.e.

$$L = \prod_{t=1}^{N} P(X_t | X_1, X_2, \ldots, X_{t-1}), \qquad (7)$$

where N is the total number of observations. Equivalently, we can also calculate the log-likelihood function as:

$$\log L = \sum_{t=1}^{N} \log P(X_t | X_1, X_2, \ldots, X_{t-1}). \qquad (8)$$

Since the joint distribution of the series X is multivariate normal, the distribution of any observation X_t conditioned on any other observation must be normal. Noting that the conditional probability density function P is Gaussian, the log-likelihood function of the time series X is given by

$$\log P(X_t | X_1, X_2, \ldots, X_{t-1}) =$$
$$= -\frac{1}{2} \log 2\pi - \frac{1}{2} \log F_t - \frac{1}{2F_t}(X_t - \overline{X}_t)^2$$
$$= -\frac{1}{2} \log 2\pi - \frac{1}{2} \log F_t - \frac{V_t^2}{2F_t}, \qquad (9)$$

where $\overline{X}_t$ is the mean of the distribution, F_t is the variance, and $V_t = X_t - \overline{X}_t$ is the prediction error. The log-likelihood of the series is then given by

$$\log L = -\frac{N}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{N} \log F_t - \frac{1}{2} \sum_{t=1}^{N} \frac{V_t^2}{F_t}. \qquad (10)$$

To compute this log-likelihood function, the $\overline{X}_t$ and F_t must be calculated for all $t = 1, \ldots, N$. This is achieved by a set of recursions, called the Kalman filter recursions.

These recursive computations are divided to two sets of equations: prediction equations and update equations [19]. From (3), we can derive the prediction equations given as

$$\mathbf{a}_{t|t-1} = \boldsymbol{\Phi}\mathbf{a}_{t-1}$$
$$\mathbf{C}_{t|t-1} = \boldsymbol{\Phi}\mathbf{C}_{t-1}\boldsymbol{\Phi}^T + \mathbf{K}\boldsymbol{\Sigma}\mathbf{K}^T, \qquad (11)$$

where $\mathbf{a}_{t|t-1}$ is the minimum mean square estimate of the state $\boldsymbol{\alpha}_t$, given all the previous observations at time $t-1$. The corresponding covariance matrix of this estimate is given by $\mathbf{C}_{t|t-1}$.

The update equations for the Kalman filter are given as [16]:

$$\overline{X}_t = \mathbf{H}\mathbf{a}_{t|t-1}$$
$$F_t = \mathbf{H}\mathbf{C}_{t|t-1}\mathbf{H}^T + \sigma_\epsilon^2$$
$$\mathbf{C}_t = \mathbf{C}_{t|t-1} - \mathbf{C}_{t|t-1}\mathbf{H}^T\mathbf{H}\mathbf{C}_{t|t-1}/F_t \qquad (12)$$
$$\mathbf{a}_t = \mathbf{a}_{t|t-1} + \mathbf{C}_{t|t-1}\mathbf{H}^T(X_t - \overline{X}_t)/F_t,$$

where $\mathbf{a}_t$ is the minimum mean square estimate of the current state $\boldsymbol{\alpha}_t$ at time t, and $\mathbf{C}_t$ is the corresponding covariance matrix of this estimate.

In practice, the recursions are started at $t = 1$ which requires the estimator of the state at $t = 0$, $\mathbf{a}_0$ and the corresponding error covariance matrix $\mathbf{C}_0$. It can be shown that for models that reach a steady state, the state estimates for large t are unaffected by the initial choice of $\mathbf{a}_0$ and $\mathbf{C}_0$. The important advantage of this recursive algorithm is the fact that $\mathbf{a}_{t|t-1}$, $\mathbf{a}_t$ and $\overline{X}_t$ are updated for new available data without the need to restart the computation for all the previous data.

Kalman filter is first used to compute the log-likelihood function, which must then be maximized to find the unknown parameters. This is an optimization problem and for our case, $\Sigma = \text{diag}(\sigma_\eta^2, \sigma_\omega^2)$ and σ_ϵ^2 include the unknown variables in log-likelihood function. In this work, we have implemented optimization techniques such as steepest descent [26], quasi-Newton and simulated annealing [20] to optimize the log-likelihood function. The implementation results are similar while steepest descent is a simple method and obtains the minimum at a fast convergence rate.

4 Extended Structural Model

To improve the performance in traffic prediction, we define the extended structural model by adding extra degrees of freedom to the basic structural model in (5). The extended traffic model is given as

$$
\begin{aligned}
\mu_t &= b\mu_{t-1} + c\beta_{t-1} + \eta_t \\
\beta_t &= d\beta_{t-1} + e\zeta_{t-1} \\
\zeta_t &= f\zeta_{t-1} \\
\gamma_t &= -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t.
\end{aligned}
\tag{13}
$$

The parameter ζ_t in the model defined in (13) is changed from a random variable in the basic structural model in (5) to a state variable. The extended model also contains additional variables b, c, d, e, and f. In this model, the state vector is defined as:

$$
\boldsymbol{\alpha}_t = (\mu_t, \beta_t, \zeta_t, \gamma_t, \gamma_{t-1}, \ldots, \gamma_{t-s+1})^T.
\tag{14}
$$

During the training phase, both the basic structural model parameters Σ and σ_ϵ^2, as well as the extended structural model parameters b, c, d, e, and f must be estimated. All the data available to build the model (training sequence) is split into a *model training sequence* and a *model validation sequence*. Note that the model training sequence must be followed by the model validation sequence.

The algorithm first selects initial values for b, c, d, e, and f. Using these values and the model training sequence, the Kalman filter prediction and update equations in (11) and (12) are utilized to estimate the noise variances $\Sigma = \text{diag}(\sigma_\eta^2, \sigma_\omega^2)$ and σ_ϵ^2 by maximizing the log-likelihood function (8).

The estimated noise variances are utilized by the model to predict future values that are compared to the model validation sequence. The quality of the model is calculated as the mean absolute percentage error (MAPE) between the predicted values and the model validation sequence. The values for the extended model parameters b, c, d, e, and f are changed according to an optimization search strategy
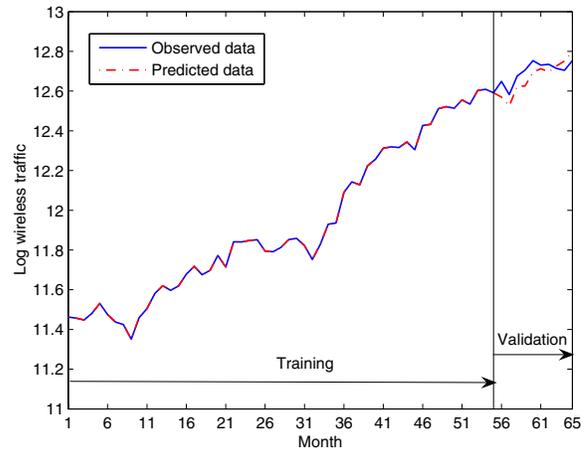

Figure 1. Model training and model validation sequences for the ESM model

and the entire process is repeated leading to a new set of parameters and a new value for the MAPE. In this work, the search follows a steepest descent algorithm; however, other search techniques could be also used.

The parameters b, c, d, e, and f that resulted in the lowest MAPE are used by the Kalman filter to build the final model. However, in this case the entire training sequence, i.e. both the model training and model validation, is used. Fig. 1 shows the data sequence of 65 points split up into a model training sequence with 55 points and a model validation sequence with 10 points and the entire training algorithm is shown in Fig. 2.

5 Numerical Results and Comparison

The effectiveness of the forecasting models is tested by predicting the number of airtime minutes per month on the Bell Canada network. To collect this data, the switch measures the call times for every call, which are then summed for a given month. Since there are 12 months in a year and the traffic follows a yearly seasonal behavior, the seasonality of the model was selected as 12. Using this data, the performance of the ESM is compared to that of both the BSM and the X12-ARIMA [33] software. X12-ARIMA is a seasonal ARIMA package designed by US census bureau that needs at least three complete cycle of data for prediction to have a small prediction error.

The available data sequence is split into training and test sequences, where the training sequence is used to build the models and the test sequence is used to measure the performance of the models. Both the BSM and ARIMA use the entire training sequence to build the model, while ESM

{*initialization*}

Split the training data into model training and model validation sequences

Select initial values of b, c, d, e, and f

$minError \leftarrow \infty$

{*main loop*}

while

  Train the Kalman filter by the model training sequence using equations (11) and (12)

  Maximize the log-likelihood function by steepest descent to find Σ and σ_ϵ^2

  Use the Kalman filter on the model validation sequence to generate the predictions

  $Error \leftarrow$ MAPE over the model validation sequence between the observed and predicted data

  if $Error < minError$ **then**

    $minError \leftarrow Error$

    $b_{opt} \leftarrow b, c_{opt} \leftarrow c, d_{opt} \leftarrow d, e_{opt} \leftarrow e, f_{opt} \leftarrow f$

  end if

  if not done

    Select new values for b, c, d, e, and f

  else

    break

  end if

end while

{*final model*}

Train the Kalman filter over the model training and model validation sequence using $b_{opt}, c_{opt}, d_{opt}, e_{opt}, f_{opt}$

Maximize the log-likelihood function by steepest descent to find Σ and σ_ϵ^2

Output the model

Figure 2. Extended structural model training algorithm

further splits the training sequence into model training and model validation sequences.

Fig. 3 shows the wireless traffic prediction using X12-ARIMA, BSM and ESM models compared to the actual observed values. In Fig. 3(a), the first 55 points (training sequence) were used to train each model. In the case of ESM, the training sequence was further divided into first 45 points as the model training sequence followed by 10 points of model validation sequence. Then, the models were used to predict points 56 to 65 (test sequence) and absolute percentage error between the predictions and the actual observations was calculated.

Fig. 3(b) shows similar results for 75 points, where the first 65 points are the training sequence, and the points 66 to 75 are the test sequence. In the case of ESM, the training sequence was divided into 55 point model training sequence and 10 point model validation sequence. Fig. 4 shows the corresponding absolute percentage error of the different model predictions.

In both of the figures, we see that the ESM model follows

Table 2. mean absolute percentage error and maximum absolute percentage error for different prediction models

Training	Test	X12-ARIMA	BSM	ESM
mean absolute percentage error				
points 1-55	points 56-65	5.81 %	7.69 %	**4.83 %**
points 1-65	points 66-75	3.70 %	7.67 %	**2.29 %**
maximum absolute percentage error				
points 1-55	points 56-65	13.27 %	17.82 %	**9.64 %**
points 1-65	points 66-75	8.54 %	12.76 %	**5.02 %**

the observed data closer than either the BSM or the X12-ARIMA models. The improvement of the ESM over the BSM and the X12-ARIMA is better as the prediction period increases. We can also observe an improvement of each model as extra 10 training points are added.

The absolute percentage error results in Fig. 4 are summarized in Table 2. In this table, we report the mean absolute percentage error, which corresponds to the error values in Fig. 4 averaged over the prediction period. The maximum absolute percentage error is the maximum error value reported in Fig. 4. As Table 2 shows, the ESM model has the smallest value of mean absolute percentage error and maximum absolute percentage error than that of the BSM and X12-ARIMA models.

6 Conclusion

We presented prediction model of the total number of minutes of wireless airtime per month on the Bell Canada network. The presented model is the new extended structural model (ESM) that was derived from the basic structural model (BSM). The presented ESM model has an improved mean absolute percentage error (MAPE) over the BSM and seasonal ARIMA model. In addition, the maximum of MAPE is also improved, which ensures a more accurate forecasting for longer period of prediction. The improved prediction can significantly reduce the cost for the wireless service providers, who need to accurately predict the future wireless spectrum requirements.
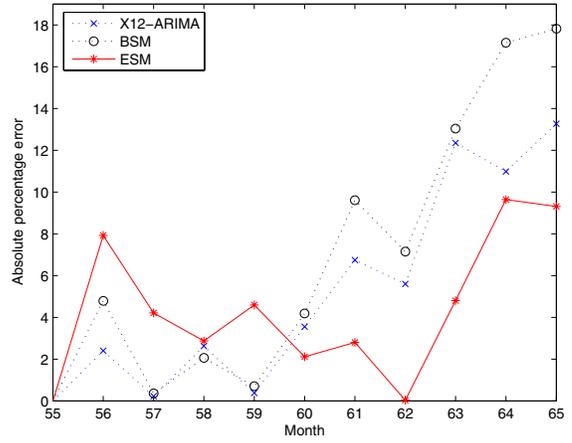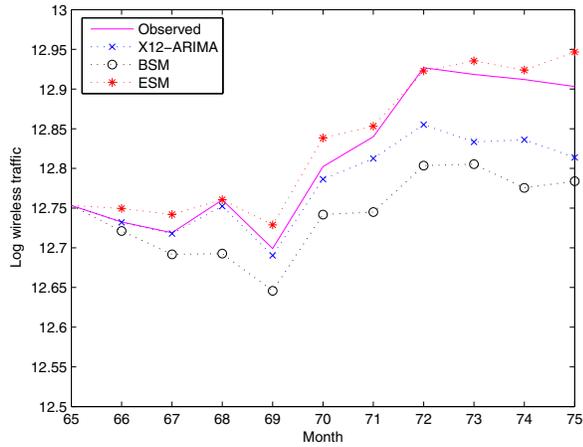
7 Acknowledgment

References

[1] B. L. Bowerman. *Time series and forecasting: an applied approach*. Duxbury Press, 1979.
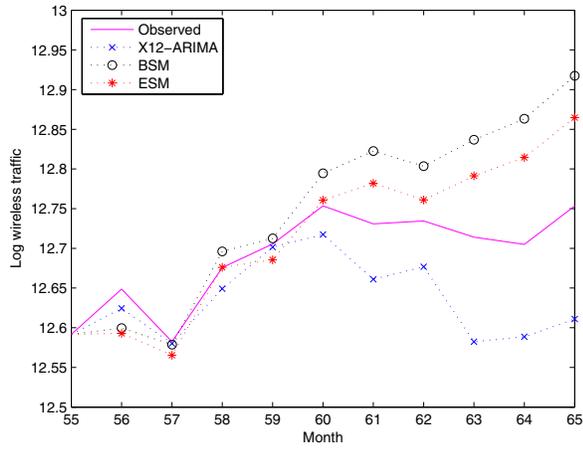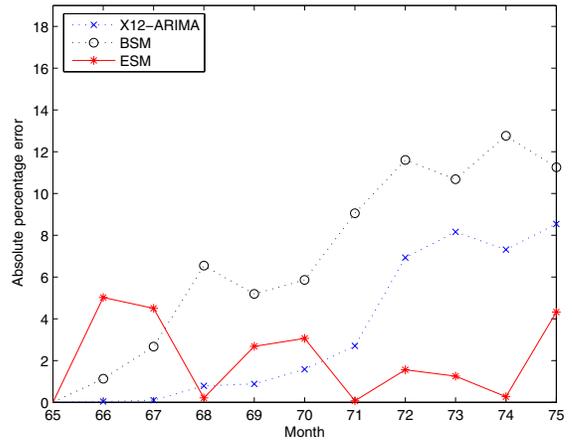
(a) 55 training points

(a) 55 training points

(b) 65 training points

(b) 65 training points

Figure 3. Prediction of 10 points for the wireless traffic using different models

Figure 4. Absolute percentage error for 10 point prediction of the wireless traffic using different models

[2] G. Box and G. Jenkins. *Time series analysis forecasting and control*. Holden-Day, San Francisco, CA, 1970.

[3] O. Darne, V. Guiraud, and M. Terraza. Forecasts of the seasonal fractional integrated series. *Journal of Forecasting*, 23, issue 1:1–17, 2004.

[4] R. Files and V. Kumar. Telecommunications demand forecasting-a review. *International Journal of Forecasting*, 18, issue 4:489–522, 2002.

[5] A. G. Gray and P. J. Thomson. On a family of finite moving-average trend filters for the ends of series. *Journal of Forecasting*, 21, issue 2:125–149, 2002.

[6] W. H. Greene. *Econometric Analysis*. Prentice Hall, Inc., 3rd edition, 1997.

[7] N. K. Groschwitz and G. C. Polyzos. A time series model of long-term NSFNET backbone traffic. In *IEEE international conference on communications*, volume 3, pages 1400–1404, 1994.

[8] C. Guang, G. Jian, and D. Wei. Nonlinear-periodical network traffic behavioral forecast based on seasonal neural network model. In *IEEE international conference on communications, circuits and systems*, volume 1, pages 683–687, 2004.

[9] J. E. Hanke. *Business Forecasting*. Prentice Hall, 7th edition, 2001.

[10] J. W. Hansen and R. D. Nelson. Neural networks and traditional time series methods: a synergistic combination in state economic forecasts. *IEEE Transaction on neural networks*, 8, issue 4:863–873, 1997.

[11] A. C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, Cambridge, U.K., 1989.

[12] D. I. Harvey and T. C. Mills. Modelling trends in central England temperatures. *Journal of Forecasting*, 22, issue 1:35–47, 2003.

[13] N. Hyung and P. H. Franses. Forecasting time series with long memory and level shifts. *Journal of Forecasting*, 24, issue 1:1–16, 2005.

[14] T. Islam and D. G. Fiebig. Modelling the development of supply-restricted telecommunications markets. *Journal of Forecasting*, 20, issue 4:249–264, 2001.

[15] T. Islam, D. G. Fiebig, and N. Meade. Modelling multinational telecommunications demand with limited data. *International Journal of Forecasting*, 18, issue 4:605–624, 2002.

[16] G. Janacek and L. Swift. *Time series: forecasting, simulation, applications*. Ellis Horwood Limited, 1993.

[17] G. G. Judge, W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T.-C. Lee. *The Theory and Practice of Econometrics*. John Wiley & Sons, 2nd edition, 1980.

[18] D. B. Jun, S. K. Kim, Y. S. Park, M. H. Park, and A. R. Wilson. Forecasting telecommunication service subscribers in substitutive and competitive environments. *International Journal of Forecasting*, 18, issue 4:561–581, 2002.

[19] R. Kalman. A new approach to linear filtering and prediction problems. *Transaction of the ASME-Journal of Basic Engineering*, 82:35–45, 1960.

[20] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[21] S. G. Koreisha and Y. Fang. Updating ARMA predictions for temporal aggregates. *Journal of Forecasting*, 23, issue 4:275–296, 2004.

[22] H. Liu and S. G. Hall. Creating high frequency national accounts with state space modelling: A monte carlo experiment. *Journal of Forecasting*, 20, issue 6:441–449, 2001.

[23] M. H. Magalhaes, R. Ballini, P. Molck, and F. Gomide. Combining forecasts for natural streamflow prediction. In *IEEE annual meeting of the fuzzy information*, volume 1, pages 390–394, 2004.

[24] K. S. Man. Long memory time series and short term forecasts. *International Journal of Forecasting*, 19, issue 3:477–491, 2003.

[25] M. Nelson, T. Hill, B. Remus, and M. O'Connor. Can neural networks applied to time series forecasting learn seasonal patterns: an empirical investigation. In *IEEE international conference on system sciences*, volume 3, pages 649–655, 1994.

[26] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.

[27] A. Prochazka. Neural networks and seasonal time-series prediction. In *IEE international conference on artificial neural networks*, volume 440, pages 36–41, 1997.

[28] M. Salvador, J. L. Gallizo, and P. Gargallo. A dynamic principal components analysis based on multivariate matrix normal dynamic linear models. *Journal of Forecasting*, 22, issue 6-7:457–478, 2003.

[29] L. Sarno and G. Valente. Comparing the accuracy of density forecasts from competing models. *Journal of Forecasting*, 23, issue 8:541–557, 2004.

[30] Statistical Sciences, Inc., Seattle, WA. *S-PLUS User's Manual*, 1991.

[31] G. W. Taylor. Smooth transition exponential smoothing. *Journal of Forecasting*, 23, issue 6:385–404, 2004.

[32] W. Tych, D. J. Pedregal, P. C. Young, and J. Davies. An unobserved component model for multi-rate forecasting of telephone call demand: the design of a forecasting support system. *International Journal of forecasting*, 18, issue 4:673–695, 2002.

[33] X12-ARIMA. http://www.census.gov/srd/www/x12a/.

[34] P. C. Young. Nonstationary time series analysis and forecasting. *Journal of Progress in Environmental Science*, 1:3–48, 1999.

[35] P. C. Young, D. J. Pedregal, and W. Tych. Dynamic harmonic regression. *Journal of forecasting*, 18, issue 6:369–394, 1999.

[36] P. C. Young, W. Tych, and D. J. Pedregal. Stochastic unobserved component models for adaptive signal extraction and forecasting. In *IEEE signal processing society workshop*, volume 8, pages 234–243, 1998.

[37] H. Zou and Y. Yang. Combining time series models for forecasting. *International Journal of Forecasting*, 20, issue 1:69–84, 2004.