Automatic Forecasting of Double Seasonal Time Series with Applications on Mobility Network Traffic Prediction

S. Tom Au, Guang-Qin Ma, and Shu-Ngai Yeung
AT&T Labs, Inc.-Research, USA

Abstract
Automatic forecasting procedures are common in business practices where large number of time series are needed for forecast. One of such applications is on mobility network resource planning which requires accurate prediction of future peak usage at each cell tower location within the network. In this paper, we developed an automatic procedure based on univariate double seasonal ARIMA models (DSARIMA) to forecast time series database with multiple seasonal patterns. A large scale empirical study comparing automatic DSARIMA with double seasonal Exponential Smoothing (DSEXP) is performed using real mobile phone network data. We also considered the performance of combined forecasts of the two models based on OLS and variations. The results show that automatic DSARIMA models and combined forecast outperform DSEXP, especially in the forecasting horizon beyond one day ahead.

1 Introduction

Real life business forecasting exercise often involves a large amount of time series. For example, a retail cooperation may need to project sales in hundreds of store locations or sales for thousands of products. In telecoms, it is critical for mobile phone companies to understand and predict the traffic volume in their network, as mobile phone technologies advance and customers increase their usage on mobile devices, [13]. The most interested traffic measurement is the busy hour load or peak load, which is defined as the highest hourly volume in the network for each day, such as busy hour voice connection or peak data bytes downloaded, depending on the type of traffic. Mobile phone traffic patterns, and hence busy hour load is location specific, depends on many factors, to name a few, such as population in the area, location (city vs rural), surrounding facilities (business vs recreation) and device mix (smart phone vs voice phone), etc. As traffic demand and equipment deployed (such as antenna and black haul cable) vary at different cell tower sites, it is important to characterize and predict the hourly traffic demand of all its equipment locations, which may come in hundreds of thousands in a nation wide mobile network. As a result, only an automatic forecasting procedure which does not need human intervention is practical.

As for other business and economics low level time series, hourly network traffic volume exhibits multiple seasonality patterns, namely the hour-of-day, day-of-week and month-of-year seasonal patterns, due to the cyclical nature of the user activities. For forecasting purpose, only models with the feature to capture the multiple seasonal pattern is appropriate. Classical univariate time series models such as ARIMA are able to model time series with double seasonal patterns [2]. Taylor [11] proposed using double seasonal exponential smoothing for short-term electricity demand forecasting. Later on, Taylor et al. [12] compared various univariate double seasonal models, including double seasonal exponential smoothing and double seasonal ARIMA for electricity load prediction. As pointed out by Taylor [11], double seasonal exponential smoothing is automatic by nature as it does not require model specification.

Numeric software package has been developed for automatic time series forecasting based on ARIMA. The most noticeable includes AutoBox and Forecast Pro, and it has been shown that commercial software preformed very well in forecasting competition [10]. Most recently, Hydman & Khandakar [7] developed the "forecast" package in R, which selects the best model from ARIMA and Exponential Smoothing families according to AIC.

However, no automatic forecasting package has been developed for ARIMA models to handle time series with multiple seasonality, which limits the comparison of the two methods in forecasting large amount of time series, such applications is becoming very common in business practices.

In this paper, we implemented automatic forecasting procedure based on univariate time series models for predicting short term hourly mobile phone network load at the cell site level. We developed an automatic double seasonal ARIMA forecasting procedure and also implemented the double seasonal exponential model proposed by Taylor [11]. By comparing the two automatic models and their combined forecasts, an empirical performance evaluation using real network data is conducted. Although, in this paper, the application of the automatic forecasting procedure focuses on mobile phone traffic, we believe such procedure is ready to apply on other business forecasting exercise with huge amount of daily or hourly time series.

The paper is organized as follows. In section 2, we describe the feature of the dataset used in the forecasting evaluation. Section 3 details the automatic forecasting models. Section 4 gives the forecasting performance and section 5 is the conclusion.

2 Data Set

Typically in network capacity planning, operators have to ensure sufficient investment on network infrastructure to meet the demand of mobile phone traffic in the upcoming years. Hence the traffic demand prediction is focused on longer term, say 36 months ahead. The most critical measurement is daily peak usage or busy hour load at each cell tower site. A simple way to forecast is to apply univariate time series methods to the daily busy hour load time series. However, currently real daily network data with long enough history for proper time series modeling is not available to the authors.

To avoid using simulated data, we turned to hourly network data and focused on short term forecasting, say two weeks ahead. The double seasonal time series methods proposed in this paper can be applied to daily time series for long term planning when data is available. While for the short term, a detailed characterization of low level network traffic provides tactical solution for network operators to manage their

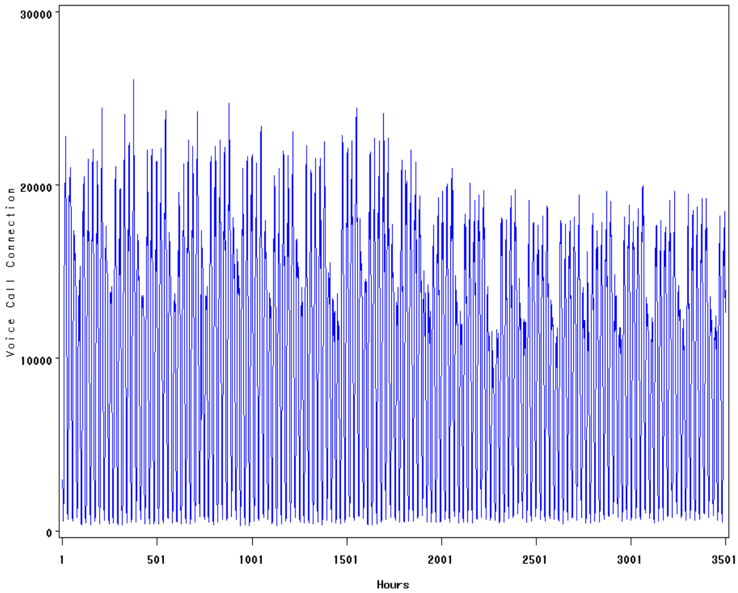

Figure 1: Example of hourly mobile phone call volume

network. For example, moving in mobile cell towers to temporarily ease congested traffic.

To evaluate the performance of the automatic forecasting procedures in practice, we used real data set in the study. The date set is a random sample of around one thousand time series from a mobile phone network database. Each time series is an hourly measurement of mobile phone calls at a particular cell tower sites. To protect the proprietary information, each time series is scaled by a random factor.

Real life data contains huge amount of noise and often has quality issues. Such volatility does not help in the modeling and comparison of performance. A pre-processing is conducted to exclude time series with structural changes and the original data is modified to smooth out extreme values. Structural changes such as level shift may occur due to new add or decommission of equipments in the tower site location, equipments failure, or data corruption in the data collection process. Extreme measurements may result due to holidays, special days (such as Mother's Day), special events (e.g. Super Bowl) or other incidents which is location specific. After the pre-processing, the total number of time series in the study is 973. The total length of each time series is 22 weeks from April 2010 to September 2010, i.e. 3696 hours. This time period is chosen to avoid the hour shift by the daylight saving in some locations in the country. The first 20 weeks of modified data is input to the models and the last 2 weeks of measurements is held out for forecasting accuracy evaluation.

A sample of the time series is shown in figure 1. It is obvious that such time series exhibit double seasonal pattern. The hour-of-week effect and day-of-week effect correspond to cycle lengths of 24 and 168 respectively.

3 Automatic Forecasting

While Taylor et al. [12] compared various univariate double seasonal models including Neural Network, this paper focuses on the classical ARIMA and exponential smoothing. The first reason is that it is not well known how to implement neural network in an automatic fashion. And the second reason is that from Taylor et al. [12] and the M3 competition [8], neural network was not outperforming ARIMA and exponential smoothing.

The advantage of ARIMA and exponential smoothing is their well proven popularity in business forecasting and their relatively straightforward automatic implementation. In the following section, we give the details of the automatic double seasonal version of ARIMA and exponential smoothing.

3.1 Automatic Double Seasonal ARIMA

The multiplicative seasonal ARIMA modeling is the most common method in business forecasting practice over the years, and often appears as a benchmark approach. The multiplicative double seasonal ARIMA model can be written as

$$\phi_p(L)\Phi_{P_1}(L^{s_1})\Omega_{P_2}(L^{S_2})\nabla^d\nabla_{s_1}^{D_1}\nabla_{s_2}^{D_2}(y_t - c) = \theta_q(L)\Theta_{Q_1}(L^{s_1})\Phi_{Q_2}(L^{s_2})\varepsilon_t \qquad (3.1)$$

where y_t is measurement in period t; c is a constant term; s_1 and s_2 are the number of periods in the different seasonal cycles; L is the lag operator; ∇ is the difference operator; $\nabla_{s_1}^{D_1}$ and $\nabla_{s_2}^{D_2}$ are seasonal difference operators; d,D_1 and D_2 are the orders of differencing; ε_t is a white noise error term; and $\phi_p, \Phi_{P_1}, \Omega_{P_2}, \theta_q, \Theta_{Q_1}$, and Φ_{Q_2} are polynomial functions of orders p, P_1, P_2, q, Q_1, and Q_2, respectively. This model can be expressed as ARIMA $(p, d, q) \times (P_1, D_1, Q_1)_{s_1} \times (P_2, D_2, Q_2)_{s_2}$.

To specify the model for each time series, one needs to identify the values of $(p, d, q, P_1, D_1, Q_1, P_2, D_2, Q_2)$. Only for lag operators with orders less than or equal to two, there are $3^6 = 729$ combinations. Following the ideas of Hydman & Khandakar [7], the automatic ARIMA performs a neighborhood search in the parameter space for models that converge in parameter estimation and minimize the AIC. We implemented the procedure under SAS using macro language and utilized the SAS PROC ARIMA for parameter estimation and AIC computation. The following algorithm is applied to each time series $\{X_t\}$:

Pre-processing, we perform log transformation and ordinal and seasonal difference of the original time series

- log transformation for the time series $Y_t = log(X_t)$
- take seasonal difference at lag s_1 and s_2, $Z_t = \nabla_{s_1}\nabla_{s_2}Y_t$. In our case, $s_1 = 24$ and $s_2 = 168$.
- preform a Dicker-Fuller test on Z_t, if the test is significant, take a difference at lag 1, $W_t = \nabla Z_t$. Otherwise, $W_t = Z_t$.

The following algorithm describes the neighborhood search for optimal model for the transformed series $\{W_t\}$. The search is conducted in the product of non-negative integer space $\aleph^6 = (p, q, P_1, Q_1, P_2, Q_2)$.

1. Initialization: set the initial optimal model set as ($opt\_p, opt\_q, opt\_P_1, opt\_Q_1$, opt_P_2, opt_Q_2)$= (1, 1, 1, 1, 1, 1)$, and opt_AIC to a very large number
2. for each $(p, q, P_1, Q_1, P_2, Q_2)$ in $(opt\_p \pm 1, opt\_q \pm 1, opt\_P_1 \pm 1, opt\_Q_1 \pm 1, opt\_P_2 \pm 1, opt\_Q_2 \pm 1)$, compute the corresponding AIC from the ARIMA model
3. replace $(opt\_p, opt\_q, opt\_P_1, opt\_Q_1, opt\_P_2, opt\_Q_2)$ by $(p, q, P_1, Q_1, P_2, Q_2)$ if the model converges and the AIC of the corresponding ARIMA model is less than opt_AIC
4. repeat step 2 until opt_AIC does not improve or the number of iteration greater than a threshold (10 in our cases)

After identifying the optimal model specification, forecasting is produced for the individual time series using the optimal model in SAS PROC ARIMA.

3.2 Double Seasonal Exponential Smoothing

Taylor [11] extended the standard Holt-Winters exponential smoothing formulation to accommodate the multiple seasonal cycles in the electricity demand series. This involves the introduction of an additional seasonal index and an extra smoothing equation for the new seasonal index. In our case, since all the values in the time series is non-negative, we take log-transformation of the original data, $y_t = log(X_t)$, and apply the additive trend and additive seasonality model. The formulation for double additive seasonality is given in the following expressions:

$$\begin{aligned}
S_t &= \alpha(y_t - D_{t-s_1} - W_{t-s_2}) + (1-\alpha)(S_{t-1} + T_{t-1}) && (3.2)\\
T_t &= \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} && (3.3)\\
D_t &= \delta(y_t - S_t - W_{t-s_2}) + (1-\delta)D_{t-s_1} && (3.4)\\
W_t &= \omega(y_t - S_t - D_{t-s_1}) + (1-\omega)W_{t-s_2} && (3.5)\\
\hat{y}_t(k) &= S_t + kT_t + D_{t-s_1+k} + W_{t-s_2+k} && (3.6)\\
&\quad + \phi^k(y_t - (S_{t-1} + T_{t-1} + D_{t-s_1} + W_{t-s_2}). && (3.7)
\end{aligned}$$

S_t and T_t, are the smoothed level and trend; D_t and W_t are the seasonal indices for the two seasonal cycles, respectively; and $\hat{y}_t(k)$ is the k step-ahead forecast made from forecast origin t. The term involving the parameter ϕ, in the forecasting expression, is a simple adjustment for first-order autocorrelation. All the parameters in the method, $\alpha, \gamma, \delta, \omega$, and ϕ, are estimated in a single procedure by minimizing the sum of squared one step-ahead in sample errors. The initial smoothed values for the level, trend and seasonal components are estimated by averaging the early observations.

3.3 Combined Forecasts

Past study has shown combined forecasts outperformed single method in forecasting accuracy, see for example Clement [3]. Such technique is very appealing in cooperate setting when one needs to forecast large amount of time series, as long as the combined forecast is able to extract features from available models feed into the system to improve the performance. The automatic ARIMA procedure developed in this paper allows the

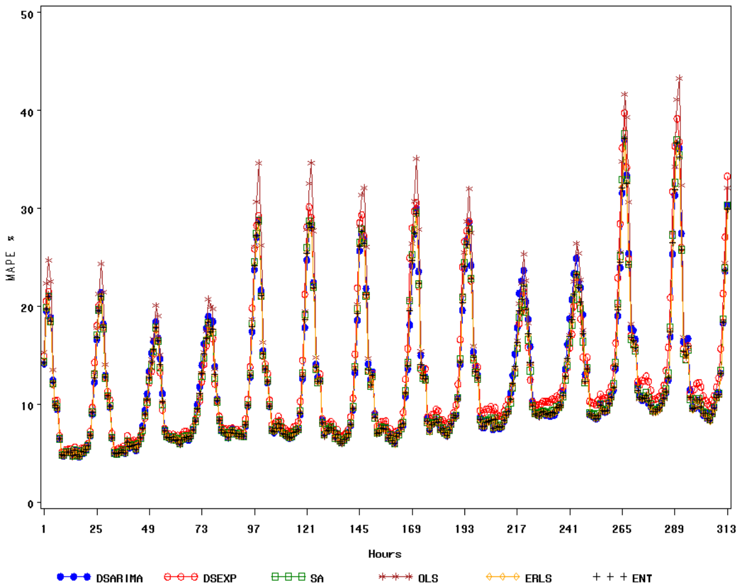

Figure 2: Out-of-Sample MAPE Up to 336 Hours Ahead

users to consider combining forecast with exponential smoothing in double seasonal time series database. Granger & Ramanathan [4] introduced the combined forecasts based on Ordinary Least Square (OLS). Aksu & Bunter [1] compared variations of OLS combine methods, while Hansen & Nelson [6] considered non-linear combined forecast based on neural network.

In this paper, we investigated the benefit of combined forecast based on double seasonal exponential smoothing and automatic double seasonal ARIMA in an empirical study. Due to the simplicity and fast computation, we consider the following 4 basic combined forecasts: simple average, Ordinary Least Square (OLS), Equality Restricted Least Squares (ERLS) and Maximum Entropy (ENT).

Suppose for each individual time series, the forecast from exponential smoothing and ARIMA is $F_{exp,t}$ and $F_{ARIMA,t}$ respectively. The form of the OLS combined forecast can be expressed as

$$F_{OLS,t} = a + b \cdot F_{exp,t} + c \cdot F_{ARIMA,t} \tag{3.8}$$

The ERLS combined forecast is obtained by imposing the $a = 0$ and $b+c = 1$ conditions on the coefficient in the regression estimation.

$$F_{ERLS,t} = b \cdot F_{exp,t} + (1 - b) \cdot F_{ARIMA,t} \tag{3.9}$$

In the OLS combined forecasting setting, the issues of multicollinearity may exist as the forecasts produced by DSARIMA and DSEXP may be highly correlated. We also tested combined forecasts by the linear estimation method based on generalized maximum entropy to see if it has advantage over traditional OLS combining. The generalized maximum entropy is designed for estimation problem with outliers in data, or in regression problems where high degrees of collinearity exist among explanatory variables or that there are more parameters to estimate than observations available to

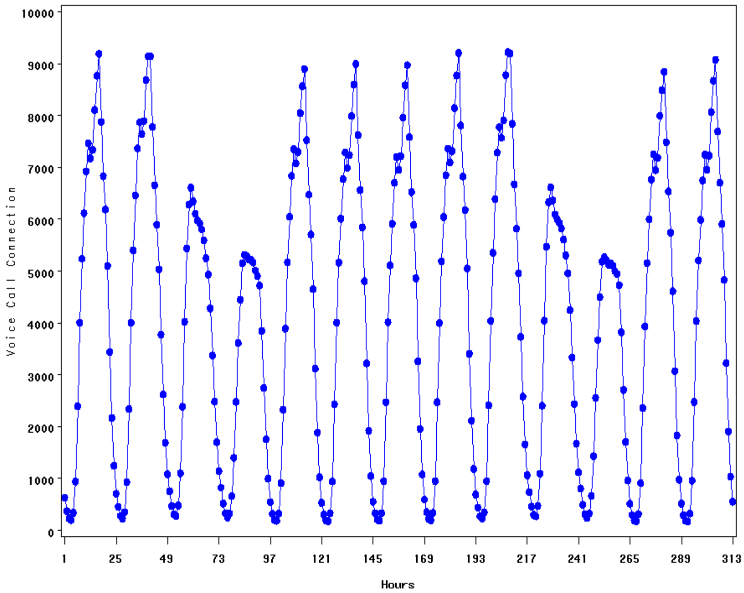

Figure 3: Average of Post Sample Hourly Phone Voice Connections

estimate them. The procedure re-parameterized the regression coefficient b and c and solved it by maximizing the entropy function. See Mittelhammer, Judge, and Miller [9] and Golan, Judge, and Miller [5] for a discussion of Shannons maximum entropy measure and the related Kullback-Leibler information. The implementation is carried out by the experimental procedure SAS PROC ENTROPY.

4 Forecasting Comparison

In this section, we compared the out-of-sample forecasting performance of the methods described in previous section. We simulated the situation where short term hourly site level forecast output is being fed to another network management system. In such situation, fixed forecast origin is used, when is normally the last hourly data point available in the system. We compared Mean Absolute Percentage Error (MAPE) and Median Absolute Percentage Error for the out-of-sample prediction versus actual. The average is taken over for the 973 sites. In real situation, one might assigned more weights to more important site location and less weight to others. In this study, we used simple average for all sites.

Figure 2 shows the MAPE for all the 336 hours lead time. The double seasonal pattern of the MAPE chart reflects the natural of the variation of the actual measurement. Low hour volume produces high MAPE due to the smaller base actual value while busy hour volume produces low MAPE. Although one can compare accuracy at each time point of the day in the forecasting horizon, busy hour or peak load accuracy is the most important criteria for network engineering operation. In this study, we compared the busy hour accuracy from day one to two weeks ahead for each site location. In the hourly time series setting, this translates to lead time ranging from 1 up to 336. Busy hour normally varies at different site location and as well as between weekdays and weekends. But for simplicity, by visually inspecting the post sample average actual

	One Day				First 7 Day			
	Mean	Median	S.D.	Max.	Mean	Median	S.D.	Max.
DSARIMA	4.78	3.74	4.34	33.75	6.16	5.08	5.12	83.3
DSEXP	5.15	4.17	4.37	32.66	6.65	5.37	6.04	81.72
SA	4.85	3.81	4.24	31.49	6.21	5.12	5.41	82.51
OLS	4.82	3.77	4.26	32.76	6.13	5.03	5.36	83.01
ERLS	4.82	3.79	4.27	32.62	6.15	5.06	5.34	82.9
ENT	4.83	3.75	4.30	32.96	6.16	5.06	5.33	83.03

Table 1: MAPE Summary for the first day and the first 7 Day Busy Hour voice call prediction. The summary statistics is based on the MAPE distribution over the 973 time series.

	Mean	Median	S.D.	Max.
DSARIMA	7.36	5.67	6.94	106.91
DSEXP	8.25	6.17	8.59	112.36
SA	7.44	5.66	7.49	109.63
OLS	7.32	5.60	7.36	108.39
ERLS	7.35	5.63	7.35	108.28
ENT	7.36	5.63	7.32	107.86

Table 2: MAPE Summary of the first 14 Day Busy Hour voice call prediction. The summary statistics is based on the MAPE distribution over the 973 time series.

measurement as shown in figure 3, we defined the busy hour as the 24 hour increments starting from lead time 16.

To evaluate the forecasting performance Table 1 and 2 show the MAPE for the first day, first 7 days and first 14 days respectively. Between DSARIMA and DSEXP, it is clear that DSARIMA outperformed DSEXP in all the 3 time horizons, and the difference between the two widens as the forecasting horizon goes further away from the origin. One reason may be due to the disadvantage of exponential smoothing, which extrapolated out based on most recent smoothed components, in the longer forecasting horizon (at least 24 hours ahead). Also, the high standard deviation of MAPE of DSEXP reflects that larger portion of the predictions is missed by large error. Figure 4 and 5 show the Mean APE and Median APE for the 973 series at each daily busy hour traffic prediction respectively. It also show that the automatic ARIMA and the combined forecasts dominated exponential smoothing after the first day. For the up to 24 hours ahead prediction, all the methods are comparable except for OLS in the first couple of hours, see Figure 6.

In automatic forecasting of large amount of times series, combined forecast has the benefit to weight the available models to best fit the individual time series. In this study, all the combined forecasts perform similarly to the DSARIMA, while simple averaging is slightly wore than DSARIMA and all the 3 OLS based combined forecasts are slightly better than DSARIMA alone. Overall, unrestricted OLS has the best performance when lead time is greater than one day.

Under the philosophy of "no single model best fits all the data', in the real life automated forecasting setting, one might prefer to implement multiple models and utilize

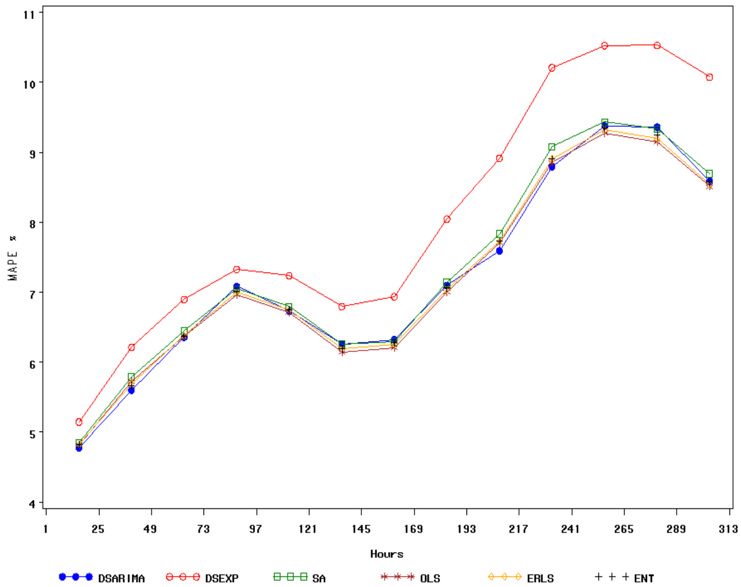

Figure 4: Out-of-Sample MAPE for Busy Hour Prediction

	Mean	Median	Mean	Median
DSARIMA	5.77	4.74	6.68	5.77
DSEXP	7.54	5.97	5.46	4.73
SA	6.48	5.13	5.86	5.04
OLS	6.20	5.00	6.05	5.18
ERLS	6.24	5.00	6.04	5.17
ENT	6.21	5.00	6.08	5.16

Table 3: Mean and Median of 7 Day Busy Hour volume prediction MAPE group by model performance. Number of time series is 556 and 417 in DSARIMA group and DSEXP group respectively.

combine forecasts to produce a more robust prediction. However, it is interesting to see if the combined forecasts is able to assign more weight to the model that best predicts the out-of-sample value. We separated the 973 time series into 2 groups according to the model performance of 7 day Busy Hour traffic prediction. One group is better predicted by DSARIMA and the complement is better by DSEXP. Table 3 summarizes the mean and median MAPE of the two groups. Notice that for each group, the combined forecasts cannot beat the best predicted model. Then we compute the average of the model weight assigned to each model for these two group, as shown in Table 4. For both ERLS and ENT methods, their average weight assigned to each model is very similar, around 40% to DSEXP and 60% to DSARIMA, while ENT in general assigns more weight to DSARIMA than ERLS. Both combined method are able to shift more weight to DSEXP when the time series is actually best predicted by DSEXP, though the shift is not as significant as one would expect. This explains why ERLS performs slightly better than ENT because it is able to shift more weight to DSEXP when the series is best predicted by DSEXP.

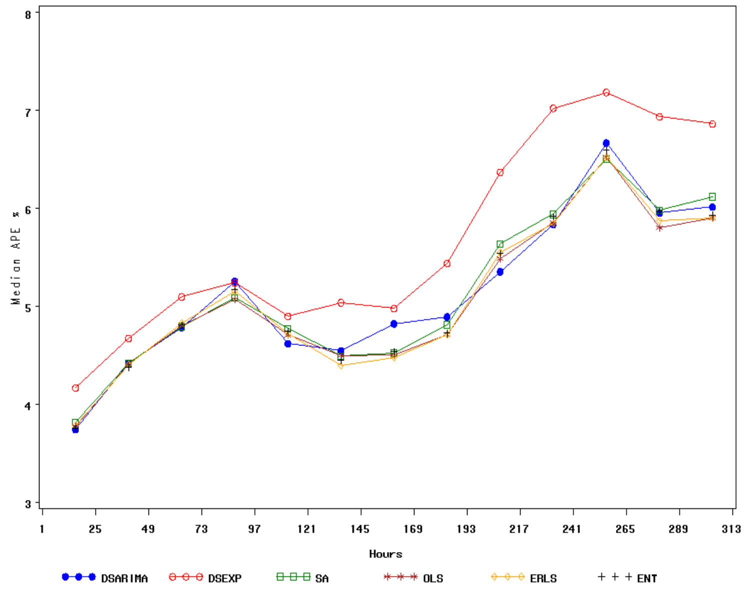

Figure 5: Out-of-Sample Median APE for Busy Hour Prediction

Combined Method	Models	Best Predicted by DSARIMA	Best Predicted by DSEXP
ERLS	DSEXP	38%	40%
ERLS	DSARIMA	62%	60%
ENT	DSEXP	36%	37%
ENT	DSAMIMA	64%	63%

Table 4: Average Model weight for DSEXP and DSARIMA by combined forecast methods and time series groups according to prediction performance.

5 Conclusion

This paper demonstrates the feasibility of automatic forecasting a large amount of time series exhibiting double seasonal patterns.The main contribution is to develop the automatic forecasting based on double seasonal ARIMA models. Such automatic procedure combined with exponential smoothing allows one to produce more robust and more accurate forecasting on double seasonal time series data base. The empirical study using real mobile network data has shown the advantage of ARIMA models and combine forecast over double seasonal exponential smoothing, especially in the forecasting horizon of one day ahead. However, the trade off of this extra accuracy is additional computation time, as the automatic ARIMA implemented in this study requires about 5 times more of CPU time than the exponential smoothing for processing the same amount of time series. So in practice double seasonal exponential smoothing may be more suitable for up to 24 hours ahead forecasting while automatic ARIMA and combine forecasts are the candidates for one day ahead and more.

A natural extension of current work is on multivariate time series models for longer term, say 36 months ahead, site level traffic prediction. Such information is critical for long term network infrastructure investment. Long term site level traffic demand

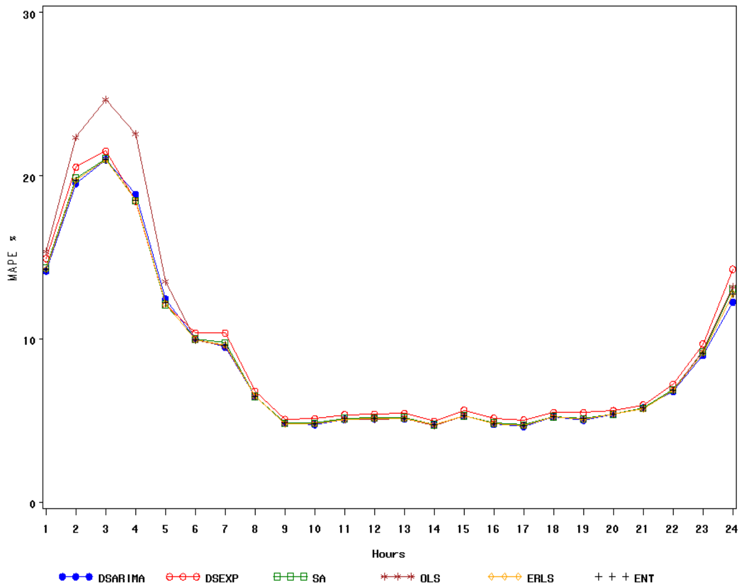

Figure 6: Out-of-Sample MAPE for the First 24 Hours Ahead

normally depends on monthly seasonal factors, population trend, network technology evolution and handset device growth. Those factors might be forecasted first and be fed into the hourly time series model for long term hourly traffic projection.

The second extension is to generalize the univariate time series models for each tower site to temporal-spatial models incorporating the traffic pattern of neighborhood sites collectively. While individual site level data may be noisy and univariate forecasts tend to have large variations, much of the information of all the site locations has not been utilized fully. It is believed that one can obtain a more stable and accurate traffic forecasting by leveraging temporal-spatial correlations of traffic patterns among neighborhood sites.

Also, it is interesting to compare with other non-linear models such as neural network, see Zhang & Qi [14], in this automatic forecasting setting.

References

[1] Aksu, C. and Bunter, S., (1992). An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts, International Journal of Forecasting, 8, 27-43

[2] Box, G.E.P., Jenkins, G.M., & Reinsel, G.C. (1994) Time series analysis: Forecasting and control (3rd ed). Englewood Cliffs, New Jersey: Prentice Hall.

[3] Clemen, R.T., (1989) Combining forecasts: A review and annotated bibliography, International Journal of Forecasting, 5, 559-583

[4] Granger, C.W.J. and Ramanathan, R., (1984). Improved methods of combining forecasts, Journal of Forecasting, 3, 197-204

[5] Golan, A., Judge, G., and Miller, D. (1996), Maximum Entropy Econometrics: Robust Estimation with Limited Data, Chichester, England: John Wiley & Sons.

[6] Hansen, J.V. and Nelson, R.D., (2003). Forecasting and recombining time-series components by using neural networks. Journal of the Operational Research Society 54 (3), 307-317

[7] Hyndman, R. and Khandakar, Y., (2008). Automatic Time Series Forecasting: The forecast Package for R, Journal of Statistical Software 27 (3), 1-22

[8] Makridakis, S., and Hibon, M. (2000). The M3-Competition: results, conclusions and implications, International Journal of Forecasting, 16, 451-476

[9] Mittelhammer, R. C., Judge, G. G., and Miller, D. J. (2000), Econometric Foundations, Cambridge: Cambridge University Press.

[10] Ord, K. (2000), Commerically available software and the M3-Competition, International Journal of Forecasting, 16, 531-537

[11] Taylor, J.W., (2003) Short-term electricity demand forecasting using double seasonal exponential smoothing, Journal of the Operational Research Society, 54, 799-805

[12] Taylor, J.W., de Menezes, L.M., and McSharry P.E. (2006), A comparison of univariate methods for forecasting electricity demand up to a day ahead, International Journal of Forecasting, 22, 1-16

[13] Forecasting Network Traffic to Avoid Meltdowns, Technology Review, DECEMBER 16, 2010, http://www.technologyreview.com/business/26807/

[14] Zhang, G.P. and Qi, M., (2005). Neural network forecasting for seasonal and trend time series, European Journal of Operational Research, 160, 501-514