Zhan-jun Qiao, Fu-ling Li, Yong Li

North China Institute of Science and Technology

Mid-long-Term Regional Load Forecasting based on Census X12-SARIMA Model

***Abstract**.: Regional power load time series has obviously trend circulation and seasonal cycle etc characteristics. In addition, using Census X12, such time series can be decomposed into trend circulation element, season element, irregular element etc. The paper attempts to establish a Census X12-SARIMA season adjustment model for mid-long-term regional power load analysis and prediction.Through empirical test for 92 months power load of Guangzhou and Suzhou area, 12 monthly power load from 2011.9 to 2012.8 was predicted. The results proved that Census X12-SARIMA model is effective in mid-long-term regional power load analysis and prediction.*

***Streszczenie.** W artykule podjęto próbę dopasowania modelu energetycznego Census X12-SARIMA na potrzeby średnio-okresowych analiz i predykcji obciążenia energetycznego. Na podstawie testów empirycznych, opartych na danych z 92 miesięcy obciążenia energetycznego regionów Guangshou i Suzhou, stworzony został 12 miesięczny profil– 09.2011-08.2012. Wyniki dowodzą ze Census X12-SARIMA jest efektywny w analizie średnio-okresowej (**Przewidywanie obciążenia średnio-okresowego, na podstawie modelu Census X12_SARIMA**).*

Keywords: seasonal adjustment; Census X12-SARIMA model; load forecasting
Słowa kluczowe: dopasowanie sezonowe, model Census X12-SARIMA, przewidywanie obciążenia.

1 Introduction

Statistics shows that as a time series, regional power load often present obvious tendency, cyclical, seasonal and other non-stationary characteristics [1]. Research literature suggests that SARIMA [2,3] model has unique advantage in the analysis of such obviously season periodic series. In addition, using Census X12 [4], such time series can be decomposed into trend circulation element, season element, irregular element etc and researched. Therefore, based on the SARIMA model and Census X12 season adjustment method, this paper attempts to establish a Census X12-SARIMA season adjustment model for mid-long-term regional power load analysis and prediction.

2 SARIMA(p,d,q)×(P,D,Q)$_s$ model

Box-Jenkins research shows that stationary time series can be analyzed by AR(p)、MA(q)、ARMA(p,q) etc model. For non-stationary time series, after "D"-order difference, it can be translated into smooth and reversible stochastic process, and it can be analyzed by ARIMA(p,d,q)[5] model. Statistics show that, the regional social electricity consumption often present seasonal variation with quarterly or monthly cycle etc. Therefore, the time series " Y_t " (quarterly or monthly load) change cycle can be set to "s". Here, seasonal difference operator can be defined as" $\Delta_s = 1 - L^s$ ". Then, the series " Y_t " "1"-order seasonal difference can be expressed as " $\Delta_s y_t = y_t - y_{t-s} = (1 - L^s) y_t$ ". For non-stationary seasonal time series, sometimes it needs to be "D" time seasonal difference, then it can be converted to a stationary series. Based on this, a cycle for " s ", "P"-order auto-regressive, "Q"-order moving average time series models can be established as:

$$A_P(L^s)\Delta_s^D y_t = B_Q(L^s)u_t \tag{1}$$

When " u_t " is non-stationary and contains ARMA components, it can be described as

$$\Phi_p(L)\Delta^d u_t = \Theta_q(L)v_t \tag{2}$$

Through the formula (1) and (2) , a general expression of the SARIMA(p,d,q)×(P,D,Q)$_s$ model can be established as:

$$\Phi_p(L)A_P(L^s)(\Delta^d \Delta_s^D y_t) = \Theta_q(L)B_Q(L^s)v_t \tag{3}$$

In which " $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ " is auto-regressive operator, " $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ " is moving average operator, "p"、"q" and "P"、"Q" represent max-lag-order corresponding the seasonal、non-seasonal for auto-regressive and moving average operator respectively. "d" and "D" represent the difference frequency corresponding the seasonal and non-seasonal respectively, " ϕ_p " represents "p"-order auto-regressive model coefficient, " θ_q " represents "q"-order moving average model coefficient, " L " is Lag operator, " v_t " is white noise.

The establishment of SARIMA(p,d,q)×(P,D,Q)$_s$ model usually includes identification order, parameter estimation and diagnostic process. The identification order means to determine the possible values of "p"、"q"、"P" and "Q". Diagnostic process is to compare the fitting results of model and choose the most suitable model order by "AIC" criterion, "SC" criterion and " $\bar{R}^2$ ". Evaluating the effectiveness of model fitting to the actual values, "AIC" and "SC" need as small as possible but" $\bar{R}^2$ " bigger is better[6].

3 Census X12 seasonal adjustment

Seasonal adjustment [6] is to eliminate seasonal and irregular fluctuant element from the quarterly or monthly time series influenced by nonlinear factors, and get potential trend circulation element of series, until revealing the series basic characteristics and basic trend. Census X12 method (USCB [7]) improved based on the X-11-ARIMA method, increases decomposition function of trend cycle, seasonal and irregular etc factor, and increases modeling function of X12-ARIMA. In addition, because of standard seasonal adjustment program embedded in the "Eviews", Census X12 method can be realized through the Windows operating.

3.1 Model selection of seasonal adjustment

Affected by uncertain factors, such as regional economic development, industry structure adjustment, seasonal variation etc, quarterly or monthly non-stationary regional power load time series can be decomposed into trend factor " T_t ", cyclic element " C_t ", seasonal element " S_t ", and irregular element " I_t ". Among them, the trend cycle composition reflects long-term change rule of the regional power load affected by the regional economic development, industrial structure adjustment and so on. Therefore, usually trend " T_t " and cycle " C_t " factors will be combined together, and denoted as" TC_t ". Seasonal component " S_t " refers to periodic variation of quarterly or monthly regional power load time series affected by climate change at same season in different years, and also refers to a regular fluctuations which power load time series appeared repeatedly around " TC_t " year after year. Irregular component " I_t " refers to random variations of power load time series affected by emergencies or natural disasters. Based on these, paper chooses multiplication model of Census X12 seasonal adjustment , and that is " $Y_t = TC_t \times S_t \times I_t$ "[7].

3.2 Algorithm of Census X12 seasonal adjustment

Step one: Initial estimate of seasonal adjustment

After centralization of 12 items moving average,

$$TC_t^{(1)} = (\frac{1}{2}Y_{t-6} + Y_{t-5} + \cdots + Y_t + \cdots + Y_{t+5} + \frac{1}{2}Y_{t+6})/12$$

$$(SI)_t^{(1)} = \frac{Y_t}{TC_t^{(1)}}$$

By "3 × 3 "moving average process, the initial estimation of seasonal factor " S_t " has

$$\hat{S}_t^{(1)} = [(SI)_{t-24}^{(1)} + 2(SI)_{t-12}^{(1)} + 3(SI)_t^{(1)} + 2(SI)_{t+12}^{(1)} + (SI)_{t+24}^{(1)}]/9$$

After eliminating the residual trend of season factor,

$$S_t^{(1)} = \hat{S}_t^{(1)} - (\hat{S}_{t-6}^{(1)} + 2\hat{S}_{t-5}^{(1)} + \cdots + 2\hat{S}_{t+5}^{(1)} + \hat{S}_{t+6}^{(1)})/24$$

So, after seasonal adjustment, the initial estimation has

$$(TC \cdot I)_t^{(1)} = \frac{Y_t}{S_t^{(1)}}$$

Step two: Calculation of tentative trend cycle element and final season factor

Using "Henderson" moving average formula, tentative trend cycle element can be calculated. That is:

$$TC_t^{(2)} = \sum_{j=-H}^{H} h_j^{(2H+1)} (TC \cdot I)_{t+j}^{(1)}$$

Among them, " $h_j^{(2H+1)}$ " is "Henderson" moving average coefficient.

Then $$(SI)_t^{(2)} = \frac{Y_t}{TC_t^{(2)}}$$

By "3 ×5 "moving average process, tentative season factor can be calculated. That is

$$\hat{S}_t^{(2)} = [(SI)_{t-36}^{(2)} + 2(SI)_{t-24}^{(2)} + 3(SI)_{t-12}^{(2)} + 3(SI)_t^{(2)} + 3(SI)_{t+12}^{(2)} + 2(SI)_{t+24}^{(2)} + (SI)_{t+36}^{(2)}]/15$$

So , final season factor can be drawn. That is

$$S_t^{(2)} = \hat{S}_t^{(2)} - (\hat{S}_{t-6}^{(2)} + 2\hat{S}_{t-5}^{(2)} + \cdots + 2\hat{S}_{t+5}^{(2)} + \hat{S}_{t+6}^{(2)})/24$$

At this point, second estimate results of seasonal adjustment shows as follows

$$(TC \cdot I)_t^{(2)} = \frac{Y_t}{S_t^{(2)}}$$

Step three: Calculation of final trend cycle element and season factor

Using "Henderson" moving average formula, final trend cycle element can be calculated. That is

(4) $$TC_t^{(3)} = \sum_{j=-H}^{H} h_j^{(2H+1)} (TC \cdot I)_{t+j}^{(2)}$$

So final season factor has

(5) $$I_t^{(3)} = \frac{(TC \cdot I)_t^{(2)}}{TC_t^{(3)}}$$

So far, the final decomposition multiplication model of series " Y_t " can be expressed as

(6) $$Y_t = TC_t^{(3)} \times S_t^{(2)} \times I_t^{(3)}$$

3.3 Improvement of modeling method on Census X12 seasonal adjustment

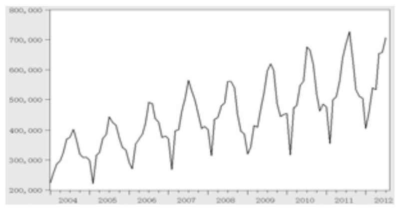
Census X12 seasonal adjustment method usually analysis and process time series data based on the X12-ARIMA model. Although in processing non-stationary time series the ARIMA model has obvious effect, but the effect less than SARIMA model in the treatment of seasonal time series. Therefore, as regional power load time series has obvious trend cyclical, seasonal periodic and non-stationary characteristics, the paper try to establish Census X12-SARIMA model, then analysis and forecast mid-long-term power load of this region. Its model form can be expressed as:

(7) $$\Phi_p(L)A_P(L^s)(1-L)^d(1-L^s)^d(y_t - \sum_{i=1}^{r}\beta_i x_{it}) = \Theta_q(L)B_Q(L^s)v_t$$

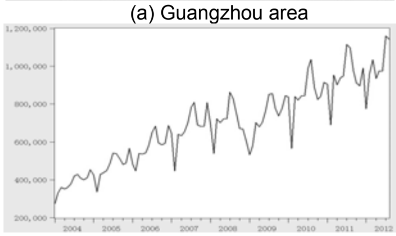
For quarterly power load data " $s = 4$ ",for monthly load data " $s = 12$ "[8], " x_{it} " is a exogenous regression factor, and " $i = 1, \cdots, r$ ". For order "p,d,q" and "P,D,Q" in Census X12-SARIMA, they will be determined according to the order identification of SARIMA model.

4 Empirical test and mid-long-term prediction of regional power load

To test the validity and ubiquity of this method, the paper selects the monthly power load statistics of the whole society power consumption from 2004.1 to 2012.8 in Guangzhou and Suzhou for modeling and analyses. Part I power load time series from 2004.1 to 2011.8 are used for modeling, order identification and forecasting. Part II power load time series from 2011.9 to 2012.8 used to test the prediction accuracy.



(a) Guangzhou area



(b) Suzhou area

Fig. 1 Time series graph of series " y_t "

4.1 Data smoothing and order identification of Census X12-SARIMA model

The power load time series graph (Fig.1) of Guangzhou and Suzhou shows that both are non-stationary series which has obvious upward trend, seasonal periodicity and incremental variance. First step, take "1"-order difference after the logarithm for part I series data, then take "ADF" test. The test results showed that, through "1"-order difference the tendency of time series have been eliminated, but "1"-order difference time series diagram shows seasonal periodicity still exist. Therefore, second step, another seasonal difference which order is "s=12" should be taken after "1"-order difference, then take "ADF" test again after seasonal difference. Test results are shown at Figure.2. The test results showed that, after such treatment, the trend and seasonal periodicity of two sets of series basically eliminated. At the same time the order of Census X12-SARIMA model of two areas can be determined, that is "d=1"，"D=1"，"P=1"，"Q=1"。

Meanwhile, from the Figure 3, the partial correlation of "$\Delta^1\Delta^{12}\ln y_t$"series of two areas all showed "p=2" or "p=3" probability, and the autocorrelation showed that "q=1" is more suitable. So the possible numerical of "(p,q)" is "(2,1)" or "(3,1)". Therefore, the possible model may be Census X12-SARIMA(2,1,1) × (1,1,1)$_{12}$ or Census X12-SARIMA(3,1,1)×(1,1,1)$_{12}$.

Take monthly power load statistics to test Census X12-SARIMA(2,1,1)×(1,1,1)$_{12}$ and Census X12-SARIMA(3,1,1)×(1,1,1)$_{12}$, through the comparison of the values of "$\overline{R}^2$"、"AIC" and "SC", Census X12-SARIMA(2,1,1)×(1,1,1)$_{12}$ can be selected as the final seasonal adjustment model of Guangzhou and Suzhou two area. Its corresponding model equations can be expressed as:

$$(1-\phi_1 L-\phi_2 L^2)(1-\alpha_1 L^{12})(1-L)(1-L^{12})(y_t-\sum_{i=1}^{r}\beta_i x_{it}) = (1+\theta_1 L)(1+\beta_1 L^{12})\nu_t \tag{8}$$

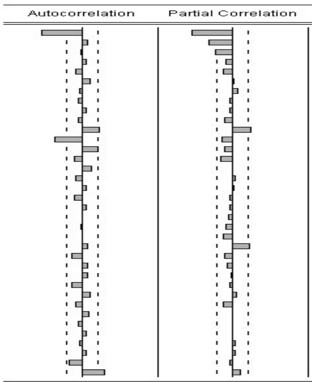
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.906654	0.0000
Test critical values:	1% level	-3.507394	
	5% level	-2.895109	
	10% level	-2.584738	

(a) Guangzhou area

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-8.800336	0.0000
Test critical values:	1% level	-3.506484	
	5% level	-2.894716	
	10% level	-2.584529	

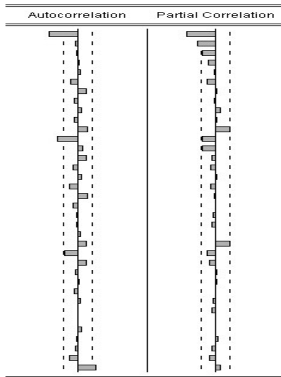
(b) Suzhou area

Fig. 2 ADF test table of "$\Delta\ln y_t$" 1st season difference



	AC	PAC	Q-Stat	Prob
1	-0.551	-0.551	28.198	0.000
2	0.074	-0.329	28.709	0.000
3	-0.019	-0.236	28.745	0.000
4	0.059	-0.085	29.081	0.000
5	-0.087	-0.124	29.810	0.000
6	0.111	0.020	31.036	0.000
7	-0.042	0.068	31.209	0.000
8	-0.062	-0.044	31.603	0.000
9	0.051	-0.034	31.869	0.000
10	-0.051	-0.112	32.136	0.000
11	0.229	0.262	37.652	0.000
12	-0.370	-0.140	52.165	0.000
13	0.213	-0.110	57.052	0.000
14	-0.102	-0.170	58.179	0.000
15	0.128	-0.003	59.973	0.000
16	-0.087	0.038	60.821	0.000
17	0.065	0.020	61.303	0.000
18	-0.109	-0.030	62.670	0.000
19	0.058	-0.035	63.066	0.000
20	-0.005	-0.065	63.070	0.000
21	-0.010	-0.093	63.081	0.000
22	-0.001	-0.133	63.081	0.000
23	0.081	0.234	63.896	0.000
24	-0.149	-0.108	66.674	0.000
25	0.079	-0.069	67.461	0.000
26	0.071	-0.018	68.107	0.000
27	-0.140	-0.036	70.691	0.000
28	0.101	0.055	72.058	0.000
29	-0.088	-0.118	73.115	0.000
30	0.096	0.008	74.397	0.000
31	-0.061	0.000	74.915	0.000
32	0.060	0.002	75.435	0.000
33	-0.041	0.037	75.679	0.000
34	0.047	0.025	76.012	0.000
35	-0.178	-0.031	80.780	0.000
36	0.297	0.100	94.284	0.000

(a) Guangzhou area



	AC	PAC	Q-Stat	Prob
1	-0.441	-0.441	18.275	0.000
2	-0.025	-0.273	18.336	0.000
3	-0.009	-0.192	18.344	0.000
4	0.019	-0.113	18.379	0.001
5	0.036	-0.016	18.509	0.002
6	-0.109	-0.127	19.687	0.003
7	0.125	0.024	21.249	0.003
8	-0.062	-0.014	21.644	0.006
9	0.064	0.069	22.065	0.009
10	-0.057	0.014	22.409	0.013
11	0.153	0.216	24.887	0.009
12	-0.306	-0.200	34.951	0.000
13	0.072	-0.194	35.511	0.001
14	0.115	-0.058	36.972	0.001
15	-0.068	-0.072	37.486	0.001
16	0.065	0.018	37.955	0.002
17	-0.126	-0.073	39.774	0.001
18	0.133	-0.014	41.833	0.001
19	-0.073	0.002	42.463	0.002
20	-0.013	-0.025	42.483	0.002
21	-0.020	-0.049	42.530	0.004
22	0.039	-0.008	42.715	0.005
23	0.115	0.218	44.365	0.005
24	-0.204	-0.139	49.602	0.002
25	0.122	-0.087	51.505	0.001
26	-0.025	0.015	51.590	0.002
27	0.020	0.008	51.643	0.003
28	-0.051	-0.003	51.986	0.004
29	0.029	-0.031	52.099	0.005
30	-0.006	-0.061	52.103	0.007
31	-0.006	-0.001	52.108	0.010
32	0.044	0.004	52.383	0.013
33	-0.015	0.025	52.414	0.017
34	-0.037	-0.049	52.621	0.022
35	-0.127	-0.097	55.047	0.017
36	0.276	0.072	66.795	0.001

(b) Suzhou area

Fig. 3 Auto correlation-partial auto correlation analysis of series

Taking part I statistical data and using "Eviews6.0", parameters of above equation can be estimated. The estimating command of "Eviews" for above equation is

DLOG(Yt,1,12) C AR(1) AR(2) SAR(12) MA(1) SMA(12)

The estimated results of Guangzhou area is:

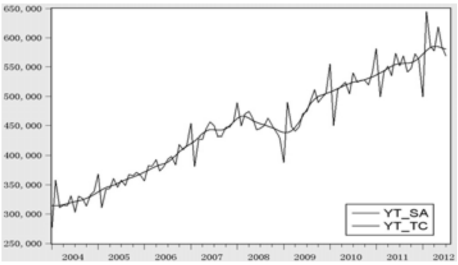
$$(1+0.5680L+0.3183L^2)(1+0.9137L^{12})(-0.0005+\Delta\Delta_{12}\ln y_t)=(1\text{-}0.3324L)(1+0.6380L^{12})\nu_t$$

(0.04) (0.19) (0.10) (0.01) (0.08) (0.25)

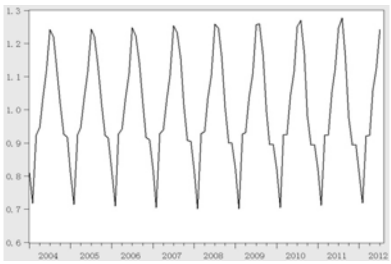
$R^2=0.41$， $s.e.=0.09$， $Q_{32}=18.27$， $\chi^2_{0.05(32-13-1)}=28.869$ ll model parameters is significant, and there has $Q_{32}=18.27\le\chi^2_{0.05(32-13-1)}=28.869$. So, the test of parameter estimation passed. Similarly, parameters of Suzhou area can be estimated. All model parameters is significant, and the test of parameter estimation passed.

3.2 Factor decomposition and load forecasting of Census X12-SARIMA model

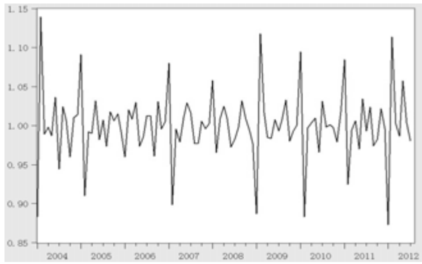
Do seasonal adjustment to the monthly power load time series of Guangzhou and Suzhou using Census X12-SARIMA(2,1,1)×(1,1,1)$_{12}$ model, to confirm the trend cycle element "TC_t"、seasonal element "S_t" and irregular element "I_t". Each element change trend is shown in Figure.4 and Figure.5.



(a) adjusted series"SA" and trend cycle element"TC_t"

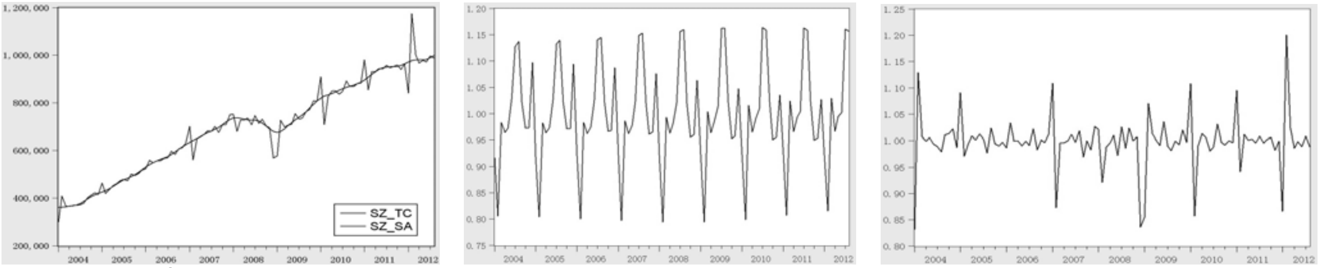


(b) seasonal element"S_t"



(c) irregular element "I_t"

Fig.4 Seasonal adjustment component for power series "y_t" of Guangzhou



(a) adjusted series"SA" and trend cycle element" TC_t " (b) seasonal element" S_t " (c) irregular element " I_t "

Fig.5 Seasonal adjustment component for power series" y_t "of Suzhou

From the curve of seasonal adjusted component graph, the power load of two area has obvious seasonal periodic, and every year in July and August the whole society power consumption reach the highest, but it reduce to the lowest in January and February, This is mainly due to the "summer high temperature" and "Spring Festival effect". Also, adjusted series and trend cycle component showed slow ascendant trend generally, and it associated with the stable development of the regional economy. In addition, from the irregular component curve, the influence of uncertain factors is more obvious, especially the " financial crisis " in 2008. At time of Census X12-SARIMA(2,1,1)×(1,1,1)$_{12}$seasonal adjustment, using Eviews6.0, the power load of next 12 months would be predicted, as shown in Table 1. For the prediction accuracy, usually adopts the "MAPE"[9]($MAPE = \frac{100}{k}\sum_{t=T+1}^{T+k}\left|\frac{\hat{y}_t - y_t}{y_t}\right|$, $t = T+1,\cdots T+k$)which smaller, the prediction accuracy is higher.

Tab 1. Forecasting results

year.month	Guangzhou area （million kwh）			Suzhou area （million kwh）		
	actual value	forecasting value	error ratio（%）	actual value	forecasting value	error ratio（%）
2011.09	6307.66	6588.46	4.452	9753.50	9909.56	1.600
2011.10	5330.51	5601.19	5.078	9105.95	9213.08	1.176
2011.11	5111.76	5190.05	-0.425	8964.86	9600.23	7.083
2011.12	5037.06	5191.96	3.075	9901.27	9937.35	0.364
2012.01	4052.89	4553.78	(12.359)	7756.73	8802.83	(13.486)
2012.02	4626.95	4001.72	(-13.512)	9575.02	8153.40	(-14.847)
2012.03	5387.75	5372.96	-0.275	10333.97	10287.95	-0.445
2012.04	5331.35	5440.79	2.053	9350.88	9797.65	4.777
2012.05	6538.90	6019.30	-7.946	9746.19	10260.53	5.277
2012.06	6576.74	6503.97	-1.106	9737.38	10242.05	5.182
2012.07	7069.11	7259.63	2.695	11144.40	11687.63	4.874
2012.08	7346.34	7429.43	1.131	10973.30	11481.32	4.629
MAPE	4.509%			5.312%		

Table 1 shows that, based on the Census X12-SARIMA seasonal adjustment model, the average prediction accuracy of two areas power load from 2011.9 to 2012.8 can reach more than 94%. At the same time, because of "Spring Festival effect ", the prediction error rate of two area power load in January and February is slightly higher. If removed the high error rate in January and February caused by " Spring Festival effect", the "MAPE" of Guangzhou and Suzhou are 2.834% and 3.541% respectively, and the average prediction accuracy of two regions power load are close to 97%. Furthermore, table 1 shows that, the region's load forecasting accuracy is not decreased obviously with the increase of predictive step.

4 Conclusions

Based on SARIMA model, using the improved Census X12-SARIMA seasonal adjustment method, the power load time series of Guangzhou and Suzhou is modeled and predicted in this paper. The empirical results indicated that, in processing the non-stationary power load time series which has obvious trend cyclical, seasonal periodic etc, the Census X12-SARIMA seasonal adjustment model showed good performance, and able to meet the requirements of regional power system mid-long-term load forecasting. It is a preferred tool of the analysis and prediction of mid-long-term regional electricity load.

REFRENCES

[1] NIU Dong-xiao,CAO Shu-hua, ZHAO Lei. Power Load Forecast Technology and Application[M]. Beijing: China Electric Power Press, 1998,10.

[2] Ghysels E, Lee H S, Noh J.Testing for unit roots in seasonal time series[J].Journal of Econometrics, 1994,(3):415~424

[3] Hylleberg S. Modeling seasonal variation, in Hargreaves, C P. (ed.) Nonstationary time series analysis and co-integration [M]·Oxford: Oxford University Press,1994.54~66

[4] THE PEOPLE'S BANK OF CHINA Statistics and Analysis Department: Time series X-12-ARIMA seasonal adjustment---theory and method, [M], China's financial press, 2006.

[5] Box G, Jenkins G, Reinsel C. Time Series Analysis: Forecasting and Control[M]. 3rd Edition. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[6] GAO Tiemei. The econometric analysis method and modeling -Eviews applications and examples[M]. Beijing:Tsinghua University press, 2006: 85~96.

[7] Catherine C. Hood. Comparison of time series characteristics for seasonal adjustments from SEATS and X-12-ARIMA[J]. ASA proceedings, October, 2002.

[8] U.S. Census Bureau. X-12ARIMA reference manual version 0·2·10,1~48[R].2002

[9] Maraval , A. and D.A. Pierce A prototypical seasonal adjustment model. Journal of Time Series Analysis, 8(1987)177-93

***Authors**: A.Prof.* zhanjunqiao, North China Institute of Science and Technology,*101601,*Beijing·Dongyanjiao,China.E-mail: zhanjunqiao@126.com. *A.Prof. Fu-ling LI,* North China Institute of Science and Technology,*101601,*Beijing·Dongyanjiao,China*.*
E-mail: lifuling0@ncist.edu.cn.