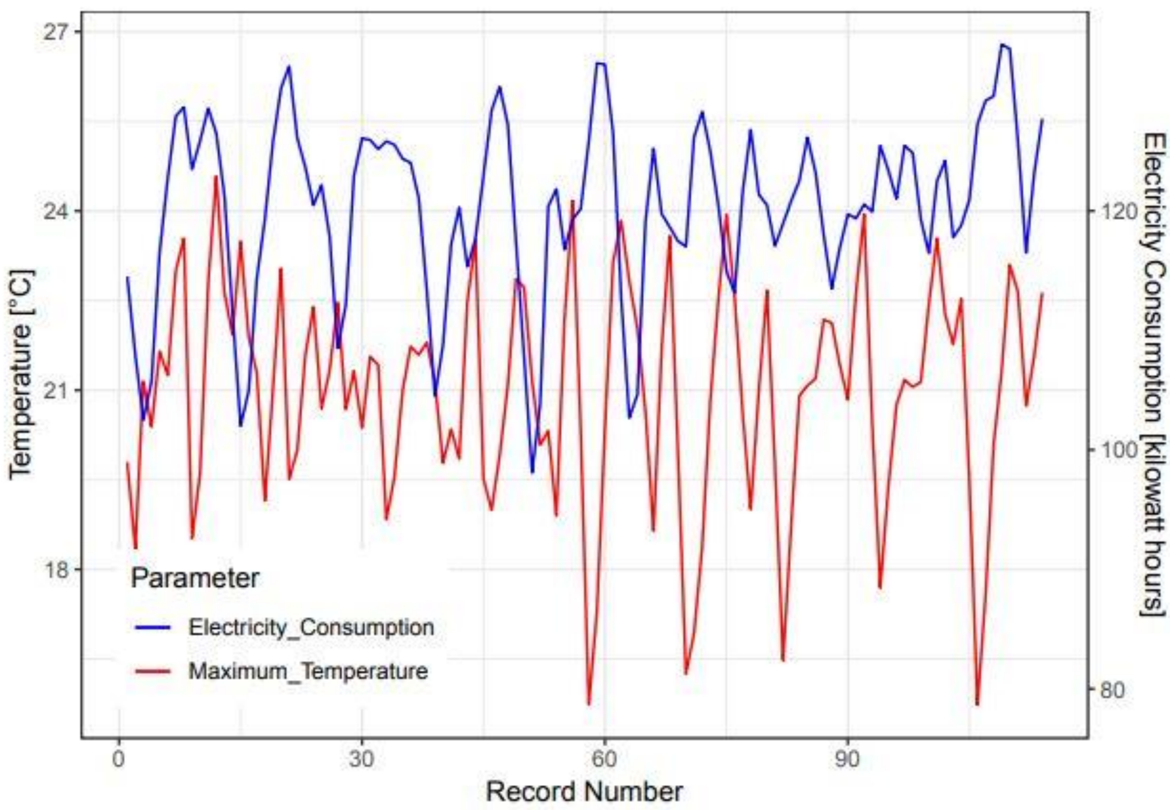
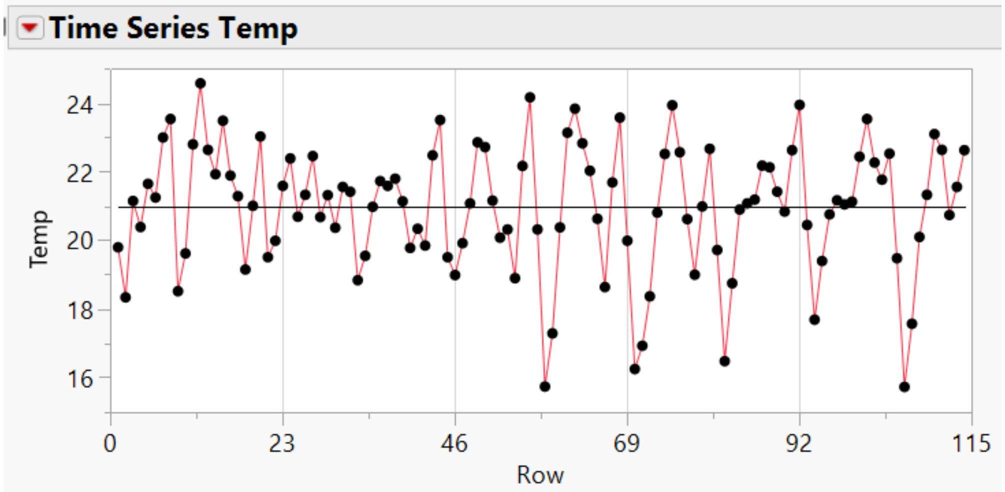
Problem Description A small city XYZ has been taking record of its monthly electricity consumption (in kilowatt hours) for about 10 years from 1997. Meanwhile, the monthly maximum temperature (in degree C) has also been recorded. The time series plots shown below indicate the varying of electricity consumption as well as temperature from Jan. 1997 to Dec. 2006. The data are included in the csv file (train.csv). A small part of them is shown as below.

```
p
```



Q.1) Use the Holt-Winter forecasting method to build a temperature forecasting model. Choose the best exponential smoothing parameter.

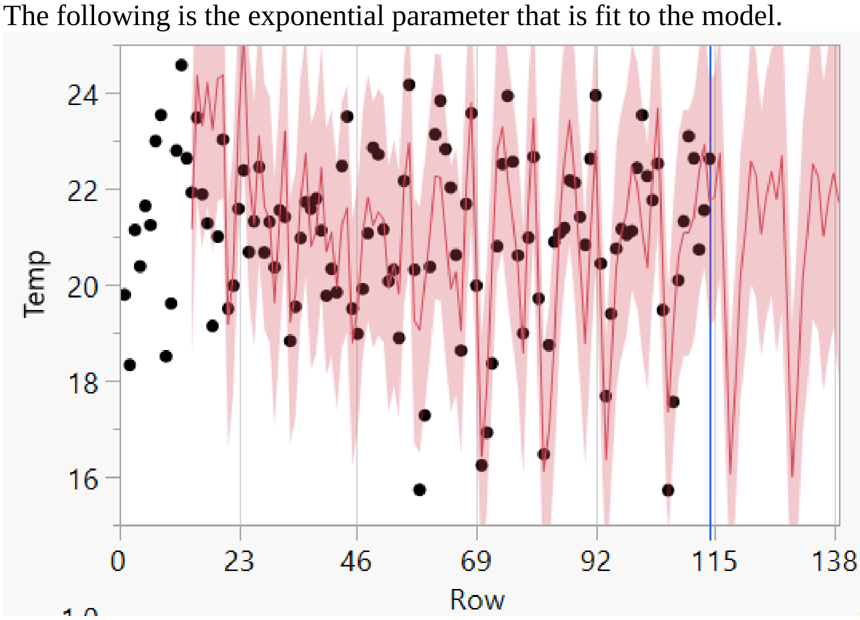


- Model Chosen : Holt-Winters Additive Model
- Reason: Seasonal pattern is approximately constant in time and it is independent of the average.

Holt-Winters Additive Model

yt= Lt+St+Et

The following is the exponential parameter that is fit to the model.

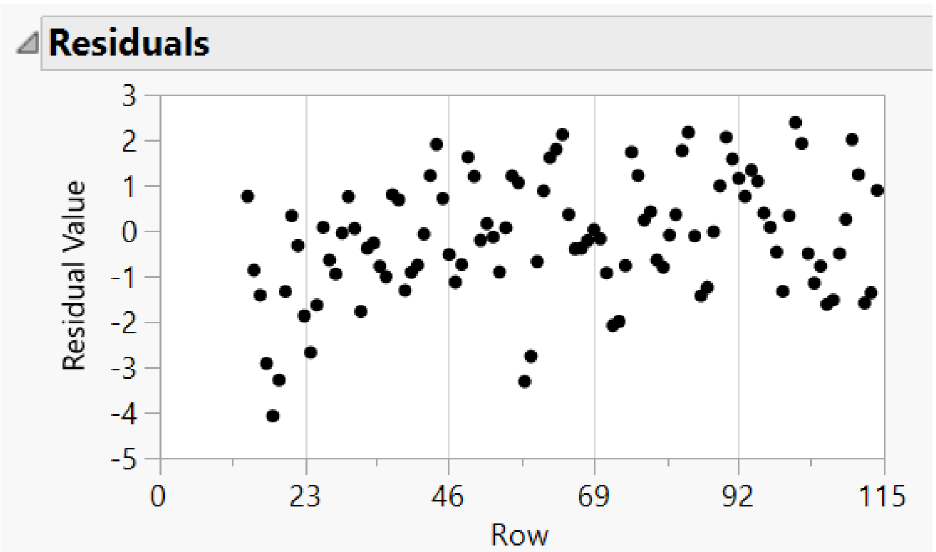


Parameter Estimates Obtained:

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Level Smoothing Weight	1.74478e-8	1.8354e-7	0.10	0.9245
Trend Smoothing Weight	0.00009885	0.0012006	0.08	0.9346
Seasonal Smoothing Weight	0.76083326	0.1390828	5.47	<.0001*

- Seasonal smoothing weight is only the significant estimate here.
- Residual plot supports the statement stated above.

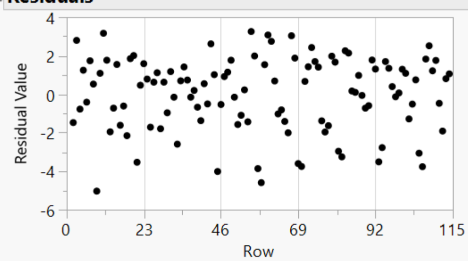


- No unusual pattern observed here. Hence, we can say that the model is adequate.
- We do not need the Level Smoothing and trend smoothing weight. The coefficient for seasonal smoothing is 0.7608 and is the best exponential smoothing parameter.

Q.2) Develop a seasonal ARIMA model for temperature. Please clearly describe each step of your model building process. Compare this model with the model you found for Question 1), in terms of some performance statistics including MSE, MAD, and MAPE.

- Fitting ARIMA model to the seasonal data: Take the first difference of the data.
  I)d=1
  II) D=1
  III) d=1 and D=1
- Plots And Adequacy Checks
  1. d=1
  - Residuals are normal
  - ACF and PACF do not show seasonal difference.

Residuals

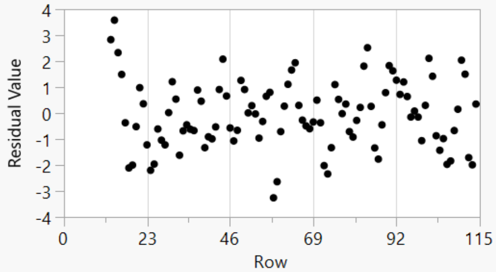


Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.1027	1.2247	0.2684	1	0.1027
2	-0.4906	29.4051	<.0001*	2	-0.5065
3	-0.2351	35.9327	<.0001*	3	-0.1456
4	0.0392	36.1155	<.0001*	4	-0.2282
5	0.0611	36.5652	<.0001*	5	-0.1749
6	-0.0206	36.6169	<.0001*	6	-0.2133
7	0.1460	39.2309	<.0001*	7	0.0812
8	0.0696	39.8301	<.0001*	8	-0.0856
9	-0.1544	42.8096	<.0001*	9	-0.0790
10	-0.4326	66.4223	<.0001*	10	-0.5829
11	0.0083	66.4310	<.0001*	11	-0.2262
12	0.7001	129.502	<.0001*	12	0.3115
13	0.1271	131.602	<.0001*	13	-0.2123
14	-0.3156	144.677	<.0001*	14	0.1075
15	-0.1912	149.523	<.0001*	15	-0.0641
16	0.0171	149.562	<.0001*	16	-0.0313
17	0.0491	149.888	<.0001*	17	0.1251
18	0.0175	149.931	<.0001*	18	0.0820
19	0.1098	151.596	<.0001*	19	-0.0177
20	0.0226	151.668	<.0001*	20	-0.0089
21	-0.1307	154.080	<.0001*	21	-0.1328
22	-0.3380	170.393	<.0001*	22	-0.0385
23	0.0169	170.434	<.0001*	23	-0.0007
24	0.5190	209.758	<.0001*	24	-0.0294
25	0.0931	211.039	<.0001*	25	-0.1233

2) D=1

- Residuals do not show any pattern.
- ACF and PACF within limits for the difference term.
- Potential candidate for the ARIMA model.

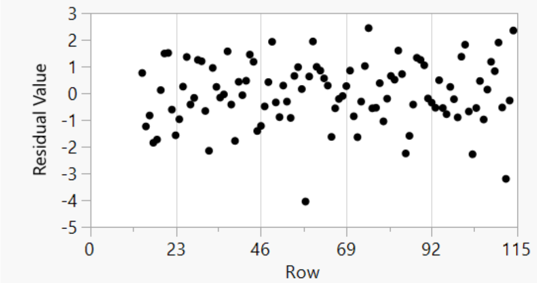
Residuals



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.5696		34.0795	<.0001*	1	0.5696	
2	-0.0279		34.1620	<.0001*	2	-0.5216	
3	-0.2638		41.6212	<.0001*	3	0.1168	
4	-0.2588		48.8710	<.0001*	4	-0.2020	
5	-0.1389		50.9823	<.0001*	5	0.0714	
6	0.0655		51.4570	<.0001*	6	0.1005	
7	0.2107		56.4132	<.0001*	7	0.0417	
8	0.1462		58.8260	<.0001*	8	-0.0750	
9	-0.0405		59.0131	<.0001*	9	-0.0495	
10	-0.2025		63.7406	<.0001*	10	-0.1238	
11	-0.2261		69.7001	<.0001*	11	0.0034	
12	-0.1028		70.9453	<.0001*	12	0.0065	
13	0.0635		71.4266	<.0001*	13	0.0240	
14	0.0738		72.0838	<.0001*	14	-0.1730	
15	-0.0193		72.1294	<.0001*	15	0.0140	
16	-0.0747		72.8186	<.0001*	16	-0.0378	
17	-0.0742		73.5054	<.0001*	17	0.0226	
18	-0.0105		73.5194	<.0001*	18	0.0663	
19	0.0686		74.1205	<.0001*	19	-0.0177	
20	0.0616		74.6110	<.0001*	20	-0.0946	
21	-0.0302		74.7307	<.0001*	21	-0.0451	
22	-0.1029		76.1351	<.0001*	22	-0.0368	
23	-0.1654		79.8104	<.0001*	23	-0.1430	
24	-0.1931		84.8797	<.0001*	24	-0.0809	
25	-0.1284		87.1514	<.0001*	25	-0.0914	

3) d=1 and D=1

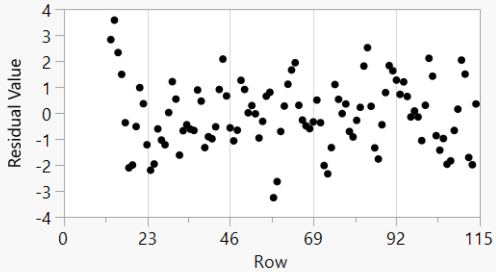
- ACF and PACF within limits.
- But the initial lags have may insignificant values.



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.1965		4.0182	0.0450*	1	0.1965	
2	-0.4216		22.6934	<.0001*	2	-0.4787	
3	-0.2893		31.5763	<.0001*	3	-0.0981	
4	-0.1045		32.7475	<.0001*	4	-0.2787	
5	-0.0627		33.1731	<.0001*	5	-0.2431	
6	0.0735		33.7652	<.0001*	6	-0.1175	
7	0.2311		39.6780	<.0001*	7	0.0112	
8	0.1197		41.2818	<.0001*	8	-0.0343	
9	-0.0209		41.3313	<.0001*	9	0.0802	
10	-0.1369		43.4730	<.0001*	10	-0.0625	
11	-0.1625		46.5240	<.0001*	11	-0.0495	
12	-0.0641		47.0048	<.0001*	12	-0.0623	
13	0.1630		50.1461	<.0001*	13	0.1095	
14	0.1419		52.5529	<.0001*	14	-0.0395	
15	-0.0389		52.7360	<.0001*	15	-0.0246	
16	-0.0972		53.8933	<.0001*	16	-0.0688	
17	-0.1032		55.2121	<.0001*	17	-0.1252	
18	-0.0048		55.2150	<.0001*	18	0.0003	
19	0.1485		58.0128	<.0001*	19	0.1015	
20	0.0838		58.9151	<.0001*	20	-0.0423	
21	-0.0268		59.0083	<.0001*	21	0.0534	
22	-0.0104		59.0226	<.0001*	22	0.0573	
23	-0.0428		59.2663	<.0001*	23	0.0107	
24	-0.1463		62.1590	<.0001*	24	-0.0617	
25	-0.1665		65.9519	<.0001*	25	-0.1924	

- Choose D=1
- Why? Because the other models have insignificant values or fail to explain seasonality.

Residuals



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.5696		34.0795	<.0001*	1	0.5696	
2	-0.0279		34.1620	<.0001*	2	-0.5216	
3	-0.2638		41.6212	<.0001*	3	0.1168	
4	-0.2588		48.8710	<.0001*	4	-0.2020	
5	-0.1389		50.9823	<.0001*	5	0.0714	
6	0.0655		51.4570	<.0001*	6	0.1005	
7	0.2107		56.4132	<.0001*	7	0.0417	
8	0.1462		58.8260	<.0001*	8	-0.0750	
9	-0.0405		59.0131	<.0001*	9	-0.0495	
10	-0.2025		63.7406	<.0001*	10	-0.1238	
11	-0.2261		69.7001	<.0001*	11	0.0034	
12	-0.1028		70.9453	<.0001*	12	0.0065	
13	0.0635		71.4266	<.0001*	13	0.0240	
14	0.0738		72.0838	<.0001*	14	-0.1730	
15	-0.0193		72.1294	<.0001*	15	0.0140	
16	-0.0747		72.8186	<.0001*	16	-0.0378	
17	-0.0742		73.5054	<.0001*	17	0.0226	
18	-0.0105		73.5194	<.0001*	18	0.0663	
19	0.0686		74.1205	<.0001*	19	-0.0177	
20	0.0616		74.6110	<.0001*	20	-0.0946	
21	-0.0302		74.7307	<.0001*	21	-0.0451	
22	-0.1029		76.1351	<.0001*	22	-0.0368	
23	-0.1654		79.8104	<.0001*	23	-0.1430	
24	-0.1931		84.8797	<.0001*	24	-0.0809	
25	-0.1284		87.1514	<.0001*	25	-0.0914	

- There are 2 significant lags in PACF
- Hence, we can try AR1 MA2 along with seasonal difference.

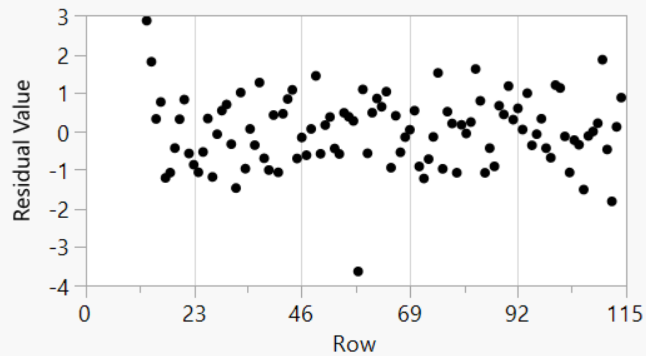
Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
AR1,1	1	1	-0.036385	0.2159751	-0.17	0.8666
MA1,1	1	1	-1.077503	0.1938059	-5.56	<.0001*
MA1,2	1	2	-0.377771	0.1766128	-2.14	0.0349*
Intercept	1	0	-0.030892	0.1993360	-0.15	0.8772

Constant Estimate	Mu
-0.0320162	-0.0308922

Forecast

Residuals



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.0186		0.0362	0.8490	1	0.0186	
2	-0.0742		0.6210	0.7331	2	-0.0746	
3	-0.1233		2.2493	0.5223	3	-0.1211	
4	-0.1312		4.1115	0.3911	4	-0.1357	
5	-0.0446		4.3295	0.5030	5	-0.0643	
6	-0.0152		4.3551	0.6287	6	-0.0542	
7	0.1993		8.7900	0.2681	7	0.1636	
8	0.0268		8.8712	0.3533	8	-0.0066	
9	-0.0169		8.9038	0.4462	9	-0.0076	
10	-0.1327		10.9342	0.3627	10	-0.1077	
11	-0.0908		11.8958	0.3715	11	-0.0548	
12	-0.1110		13.3471	0.3443	12	-0.1270	
13	0.1127		14.8623	0.3160	13	0.0932	
14	0.0463		15.1202	0.3700	14	-0.0513	
15	-0.0393		15.3082	0.4295	15	-0.0811	
16	-0.0124		15.3271	0.5008	16	-0.0398	
17	-0.0437		15.5650	0.5549	17	-0.0085	
18	-0.0405		15.7723	0.6084	18	-0.0431	
19	0.0613		16.2527	0.6404	19	0.0967	
20	0.0565		16.6662	0.6745	20	-0.0158	
21	-0.0585		17.1138	0.7042	21	-0.0962	
22	0.0065		17.1193	0.7567	22	-0.0016	
23	-0.1116		18.7915	0.7133	23	-0.1099	
24	0.0222		18.8584	0.7595	24	0.0175	
25	-0.2282		26.0355	0.4057	25	-0.2697	

- As we can see that, there are no significant lags we cannot proceed with this model.
- We try fitting ARIMA model with the help of R software’s optimal ARIMA.
- Obtained model was $(2,0,1)x(0,1,1)_{12}$

Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
AR1,1	1	1	0.6383751	0.1631358	3.91	0.0002*
AR1,2	1	2	-0.4142309	0.1365723	-3.03	0.0031*
MA1,1	1	1	-0.3819288	0.1769813	-2.16	0.0334*
MA2,12	2	12	0.1795443	0.1043152	1.72	0.0884

- Parameter estimates show that the MA2,12 term is not significant.
- Thus, this model is discarded.
- Now trying to fit a seasonal ARIMA of order $(2,0,1)x(0,1,0)_{12}$

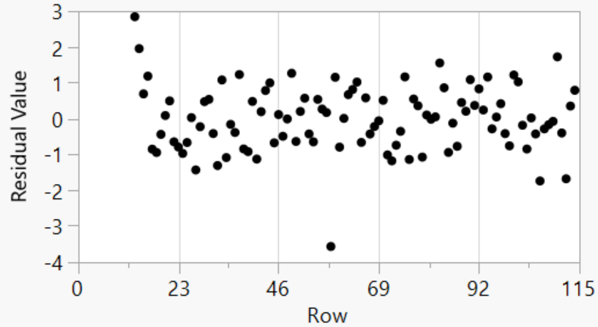
Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
AR1,1	1	1	0.6246971	0.1625715	3.84	0.0002*
AR1,2	1	2	-0.3922484	0.1379396	-2.84	0.0054*
MA1,1	1	1	-0.4022916	0.1713703	-2.35	0.0209*

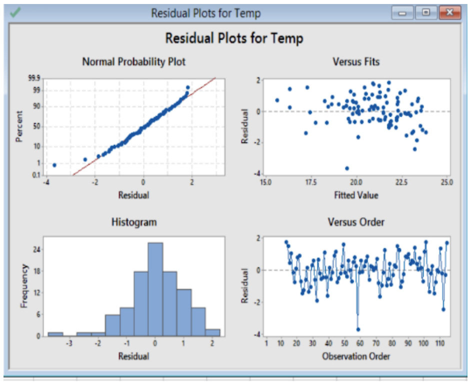
- All parameters are significant.

- ACF and PACF indicates that the model is adequate.

Residuals



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.0200		0.0418	0.8379	1	0.0200	
2	-0.0217		0.0918	0.9552	2	-0.0221	
3	0.0132		0.1105	0.9905	3	0.0141	
4	-0.0496		0.3772	0.9843	4	-0.0507	
5	-0.0755		1.0002	0.9625	5	-0.0730	
6	-0.0370		1.1518	0.9792	6	-0.0369	
7	0.1496		3.6499	0.8191	7	0.1504	
8	-0.0166		3.6811	0.8847	8	-0.0249	
9	-0.0354		3.8236	0.9226	9	-0.0364	
10	-0.0999		4.9733	0.8930	10	-0.1170	
11	-0.0664		5.4875	0.9053	11	-0.0547	
12	-0.1438		7.9247	0.7910	12	-0.1312	
13	0.0779		8.6487	0.7989	13	0.0961	
14	0.0122		8.6665	0.8518	14	-0.0354	
15	-0.0481		8.9488	0.8802	15	-0.0588	
16	0.0064		8.9538	0.9153	16	-0.0233	
17	-0.0329		9.0893	0.9374	17	-0.0210	
18	-0.0445		9.3391	0.9514	18	-0.0356	
19	0.0431		9.5768	0.9626	19	0.0833	
20	0.0392		9.7754	0.9721	20	-0.0208	
21	-0.0895		10.8248	0.9661	21	-0.1214	
22	-0.0072		10.8318	0.9771	22	-0.0325	
23	-0.0882		11.8773	0.9723	23	-0.1023	
24	0.0275		11.9803	0.9801	24	0.0369	
25	-0.2498		20.5800	0.7158	25	-0.2631	



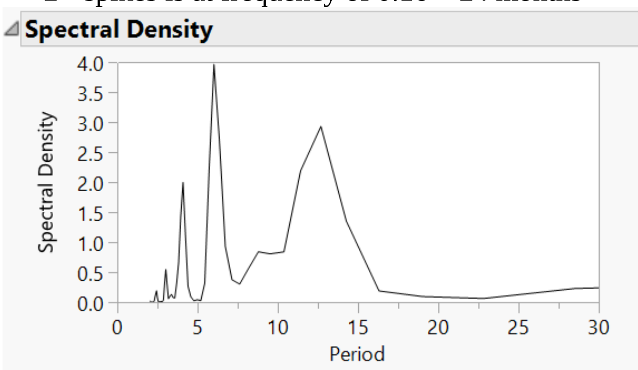
- Residual plots show normal probability plot which looks linear, residual vs. fits is scattered, histogram shows normal curve and residual vs. observation order shows observations around mean almost equidistant.
- Comparison of Winters vs. ARIMA model

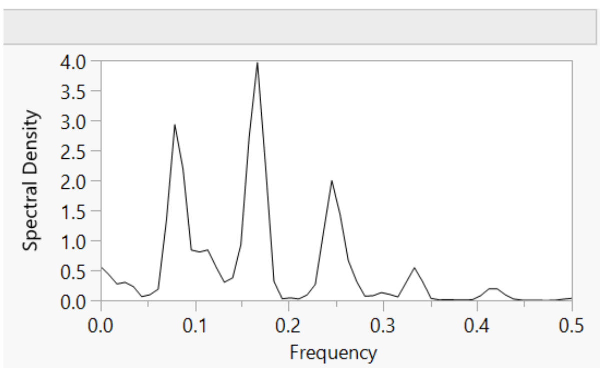
	WINTERS		ARIMA
MAD	1.056		0.70
MAPE	5.09		3.43
MSE	181.40		0.81

- As the MAD, MAPE and MSE of ARIMA is lesser than winters method, ARIMA model is the better of the two models.

Q.3) Perform a spectral analysis of the temperature process to identify the significant frequency and make explanation.

- The important frequencies are listed as follows:
- 1st spikes obtained is in the period of 0.08 ~ 12 months
- 2nd spikes is at frequency of 0.16 ~ 24 months





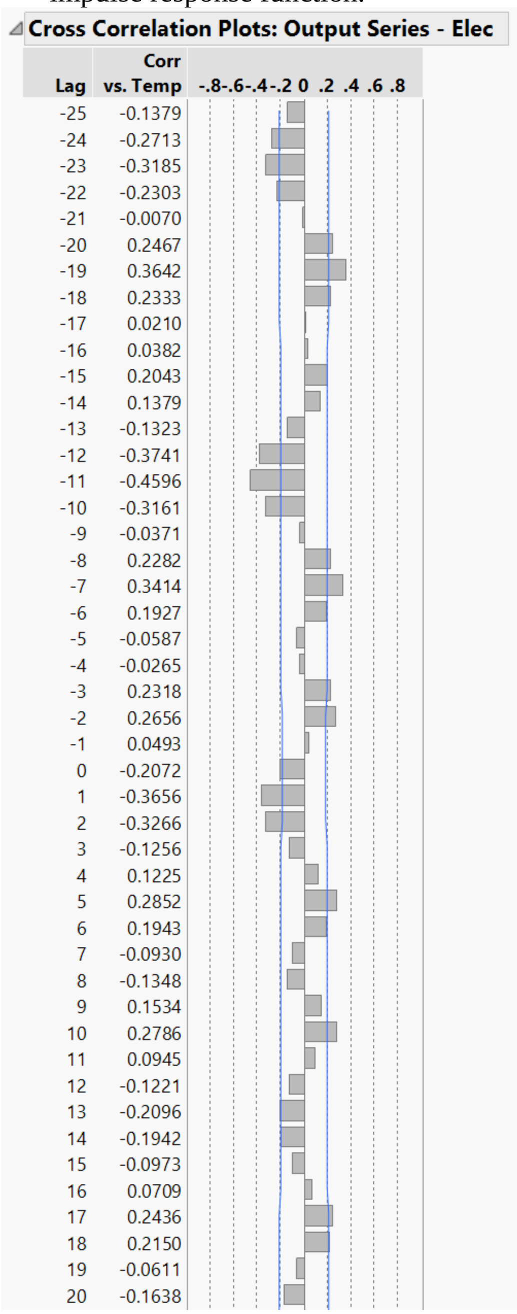
White Noise test

Fisher's Kappa	11.752839
Prob > Kappa	0.0001323
Bartlett's Kolmogorov-Smirnov	0.3899686

- Thus, we can say that Periodicity in the data with period = 12 months.

Q.4) Develop a transfer function model to illustrate the change of electricity consumption with its relation to the exogenous variable – temperature. You may choose any temperature model developed for Question 1), 2) or 3). Please clearly describe each step of your model building process. Remember to show your model adequacy checks.

- Check cross correlation of Electricity and temperature.:
  It indicates that there is autocorrelation between electricity and temperature.
- Perform Pre-whitening in order to remove the auto correlation; this helps in finding the possible impulse response function.



Cross Correlation Plots: Output Series - Elec

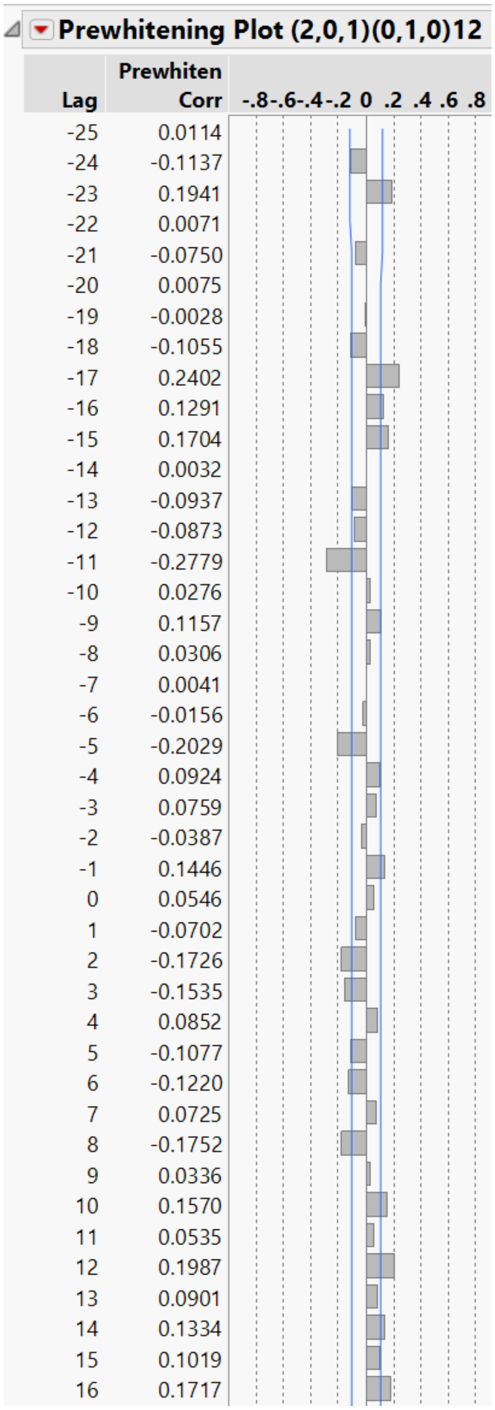
Lag	Corr vs. Temp
-25	-0.1379
-24	-0.2713
-23	-0.3185
-22	-0.2303
-21	-0.0070
-20	0.2467
-19	0.3642
-18	0.2333
-17	0.0210
-16	0.0382
-15	0.2043
-14	0.1379
-13	-0.1323
-12	-0.3741
-11	-0.4596
-10	-0.3161
-9	-0.0371
-8	0.2282
-7	0.3414
-6	0.1927
-5	-0.0587
-4	-0.0265
-3	0.2318
-2	0.2656
-1	0.0493
0	-0.2072
1	-0.3656
2	-0.3266
3	-0.1256
4	0.1225
5	0.2852
6	0.1943
7	-0.0930
8	-0.1348
9	0.1534
10	0.2786
11	0.0945
12	-0.1221
13	-0.2096
14	-0.1942
15	-0.0973
16	0.0709
17	0.2436
18	0.2150
19	-0.0611
20	-0.1638

- Selecting ARIMA model of question 2 of order $(2,0,1)x(0,1,0)_{12}$ for the temperature data to find out the possible impulse response function.

The following is the ACF and PACF of temp using the above ARIMA model.

Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.0200		0.0418	0.8379	1	0.0200	
2	-0.0217		0.0918	0.9552	2	-0.0221	
3	0.0132		0.1105	0.9905	3	0.0141	
4	-0.0496		0.3772	0.9843	4	-0.0507	
5	-0.0755		1.0002	0.9625	5	-0.0730	
6	-0.0370		1.1518	0.9792	6	-0.0369	
7	0.1496		3.6499	0.8191	7	0.1504	
8	-0.0166		3.6811	0.8847	8	-0.0249	
9	-0.0354		3.8236	0.9226	9	-0.0364	
10	-0.0999		4.9733	0.8930	10	-0.1170	
11	-0.0664		5.4875	0.9053	11	-0.0547	
12	-0.1438		7.9247	0.7910	12	-0.1312	
13	0.0779		8.6487	0.7989	13	0.0961	
14	0.0122		8.6665	0.8518	14	-0.0354	
15	-0.0481		8.9488	0.8802	15	-0.0588	
16	0.0064		8.9538	0.9153	16	-0.0233	
17	-0.0329		9.0893	0.9374	17	-0.0210	
18	-0.0445		9.3391	0.9514	18	-0.0356	
19	0.0431		9.5768	0.9626	19	0.0833	
20	0.0392		9.7754	0.9721	20	-0.0208	
21	-0.0895		10.8248	0.9661	21	-0.1214	
22	-0.0072		10.8318	0.9771	22	-0.0325	
23	-0.0882		11.8773	0.9723	23	-0.1023	
24	0.0275		11.9803	0.9801	24	0.0369	
25	-0.2498		20.5800	0.7158	25	-0.2631	

- As there are no significant lags, we use the above model to pre-white the temp variable.
- Following shows the prewhitening plot for the chosen ARIMA model.



Prewhitening Plot (2,0,1)(0,1,0)12

Lag	Prewhiten Corr
-25	0.0114
-24	-0.1137
-23	0.1941
-22	0.0071
-21	-0.0750
-20	0.0075
-19	-0.0028
-18	-0.1055
-17	0.2402
-16	0.1291
-15	0.1704
-14	0.0032
-13	-0.0937
-12	-0.0873
-11	-0.2779
-10	0.0276
-9	0.1157
-8	0.0306
-7	0.0041
-6	-0.0156
-5	-0.2029
-4	0.0924
-3	0.0759
-2	-0.0387
-1	0.1446
0	0.0546
1	-0.0702
2	-0.1726
3	-0.1535
4	0.0852
5	-0.1077
6	-0.1220
7	0.0725
8	-0.1752
9	0.0336
10	0.1570
11	0.0535
12	0.1987
13	0.0901
14	0.1334
15	0.1019
16	0.1717

From the above plot, following are the possible impulse response functions:
$b = 2, r = 1, s = 1$

Comparison of impulse response function:

1) b = 2, r = 1, s = 1

Model Summary

DF	107
Sum of Squared Errors	4843.74737
Variance Estimate	45.2686664
Standard Deviation	6.72819934
Akaike's 'A' Information Criterion	742.130779
Schwarz's Bayesian Criterion	752.9689
RSquare	0.22439723
RSquare Adj	0.20324443
MAPE	4.07791606
MAE	4.79979714
-2LogLikelihood	734.13078

Parameter Estimates

Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Temp	Num0,0	0	0	-1.3865	0.328937	-4.22	<.0001*
Temp	Num1,1	1	1	-1.4306	0.318899	-4.49	<.0001*
Temp	Den1,1	1	1	0.7532	0.120491	6.25	<.0001*
	Intercept	0	0	117.0371	6.397970	18.29	<.0001*

$$\text{Elec}_t = 117.0371 + \left(\frac{\left(-1.3865 + 1.4306 \bullet B \right)}{\left(1 - 0.7532 \bullet B \right)} \right) \bullet \text{Temp}_{t-2} + e_t$$

- Clearly, the first model b = 2, r = 1, s = 1 is the best because of the best parameters obtained.
- We will need to check adequacy for this model or else we will have to consider second model.
- Fitting the noise term for b = 2, r = 1, s = 1.
- ACF and PACF of response function:

Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.6097		43.5050	<.0001*	1	0.6097	
2	0.1198		45.2004	<.0001*	2	-0.4010	
3	-0.0445		45.4368	<.0001*	3	0.1851	
4	-0.0720		46.0602	<.0001*	4	-0.1503	
5	-0.2389		52.9832	<.0001*	5	-0.2932	
6	-0.3961		72.1987	<.0001*	6	-0.1065	
7	-0.2567		80.3445	<.0001*	7	0.1493	
8	-0.1084		81.8105	<.0001*	8	-0.2476	
9	-0.1096		83.3222	<.0001*	9	-0.0242	
10	-0.0188		83.3672	<.0001*	10	0.2062	
11	0.3593		99.9421	<.0001*	11	0.4147	
12	0.6208		149.902	<.0001*	12	0.0889	
13	0.3737		168.186	<.0001*	13	-0.2451	
14	0.0353		168.350	<.0001*	14	0.0270	
15	-0.0410		168.574	<.0001*	15	0.0328	
16	0.0007		168.575	<.0001*	16	0.0849	
17	-0.1425		171.345	<.0001*	17	-0.0415	
18	-0.2958		183.401	<.0001*	18	-0.0147	
19	-0.2030		189.138	<.0001*	19	0.0049	
20	-0.0851		190.158	<.0001*	20	-0.0046	
21	-0.0980		191.524	<.0001*	21	0.0285	
22	-0.0565		191.983	<.0001*	22	-0.0068	
23	0.2173		198.844	<.0001*	23	-0.0045	
24	0.4176		224.465	<.0001*	24	0.0348	
25	0.2651		234.908	<.0001*	25	0.0356	

- Trying Seasonal ARIMA (2,1,1)x(1,1,0)12 here (p=2 q=1 d=1, P=1 Q=0 D=1),
- All the parameters that are obtained are significant.
- Ignoring the insignificant intercept term.

Model Summary

DF	90
Sum of Squared Errors	780.494935
Variance Estimate	8.67211342
Standard Deviation	2.94484523
Akaike's 'A' Information Criterion	503.120047
Schwarz's Bayesian Criterion	523.799787
RSquare	0.43730574
RSquare Adj	0.39495241
MAPE	1.61009529
MAE	1.93621525
-2LogLikelihood	487.120047

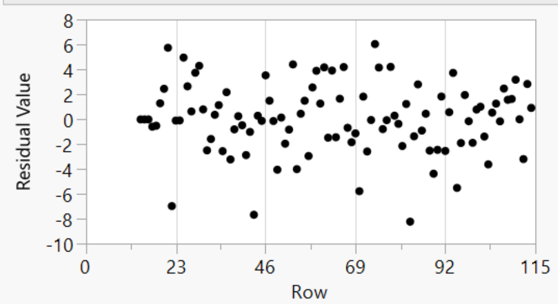
Parameter Estimates

Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Temp	Num0,0	0	0	-0.135726	0.0134464	-10.09	<.0001*
Temp	Num1,1	1	1	-0.136886	0.0135261	-10.12	<.0001*
Temp	Den1,1	1	1	0.952139	0.0995505	9.56	<.0001*
Elec	AR1,1	1	1	1.002047	0.0848539	11.81	<.0001*
Elec	AR1,2	1	2	-0.526135	0.0855363	-6.15	<.0001*
Elec	AR2,12	2	12	-0.244629	0.1078992	-2.27	0.0258*
Elec	MA1,1	1	1	1.000000	0.0314689	31.78	<.0001*
	Intercept	0	0	-0.424804	0.6036267	-0.70	0.4834

$$(1-B)\bullet\left(1-B^{12}\right)\bullet \mathrm{Elec}_t = -0.4248 + \left(\frac{\left(-0.1357 + 0.1369 \bullet B\right)}{\left(1 - 0.9521 \bullet B\right)}\right)\bullet \mathrm{Temp}_{t-2} + \left(\frac{(1-B)}{\left(\left(\left(1 - 1.002 \bullet B\right) + 0.5261 \bullet B^{2}\right)\bullet\left(1 + 0.2446 \bullet B^{12}\right)\right)}\right)\bullet e_t$$

- Checking Model Adequacy:
- No specific pattern observed in the residual plot.
- ACF and PACF plots are within the limits. Hence, the model is adequate.

Residuals



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.0026		0.0007	0.9786	1	0.0026	
2	-0.0202		0.0436	0.9784	2	-0.0202	
3	0.0674		0.5258	0.9132	3	0.0675	
4	-0.0668		1.0040	0.9092	4	-0.0680	
5	-0.0377		1.1578	0.9489	5	-0.0344	
6	0.0837		1.9249	0.9265	6	0.0777	
7	-0.0044		1.9271	0.9638	7	0.0020	
8	0.0779		2.6066	0.9566	8	0.0824	
9	-0.0866		3.4544	0.9435	9	-0.1057	
10	-0.0959		4.5058	0.9217	10	-0.0839	
11	-0.0279		4.5959	0.9491	11	-0.0355	
12	-0.0349		4.7381	0.9662	12	-0.0218	
13	0.1725		8.2557	0.8266	13	0.1875	
14	-0.0755		8.9366	0.8351	14	-0.1155	
15	-0.0623		9.4068	0.8553	15	-0.0490	
16	0.0411		9.6137	0.8860	16	0.0161	
17	-0.1214		11.4382	0.8329	17	-0.0808	
18	-0.1968		16.2922	0.5722	18	-0.1769	
19	0.2273		22.8461	0.2442	19	0.1918	
20	-0.0279		22.9458	0.2915	20	-0.0256	
21	-0.0528		23.3083	0.3278	21	-0.0762	
22	-0.1145		25.0358	0.2954	22	-0.1523	
23	-0.1055		26.5212	0.2769	23	-0.0600	
24	-0.2556		35.3460	0.0634	24	-0.2733	
25	-0.0397		35.5612	0.0785	25	-0.0494	

We fit another noise term for the same b = 2, r = 1 and s = 1. That is $(1,0,0)(0,1,0)_{12}$

Model Summary

DF	94
Sum of Squared Errors	1117.98221
Variance Estimate	11.8934267
Standard Deviation	3.44868478
Akaike's 'A' Information Criterion	531.493429
Schwarz's Bayesian Criterion	544.469028
RSquare	0.49334146
RSquare Adj	0.47244832
MAPE	2.90178971
MAE	3.43170048
-2LogLikelihood	521.493429

Parameter Estimates

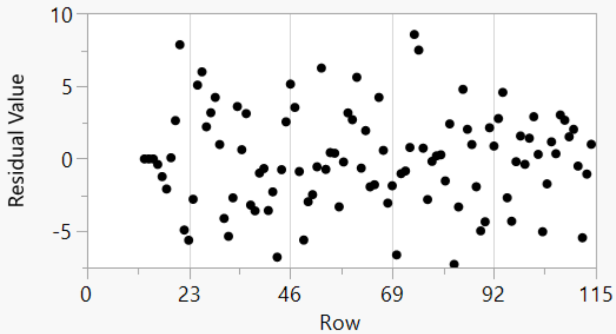
Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Temp	Num0,0	0	0	-0.214952	0.001254	-171.5	<.0001*
Temp	Num1,1	1	1	-0.214073	0.001247	-171.6	<.0001*
Temp	Den1,1	1	1	1.066609	0.091855	11.61	<.0001*
Elec	AR1,1	1	1	0.651167	0.074921	8.69	<.0001*
	Intercept	0	0	-0.444996	1.422598	-0.31	0.7551

$$\left(1-B^{12}\right)\bullet Elec_t = -0.445 + \left(\frac{\left(-0.215+0.2141\bullet B\right)}{\left(1-1.0666\bullet B\right)}\right)\bullet Temp_{t-2} + \left(\frac{1}{\left(1-0.6512\bullet B\right)}\right)\bullet e_t$$

- The estimated parameters are significant as can be seen from above.

- But, the ACF and PACF show that initial lags are not within limits.

Residuals



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.3313		11.5248	0.0007*	1	0.3313	
2	-0.1552		14.0796	0.0009*	2	-0.2976	
3	-0.3158		24.7657	<.0001*	3	-0.1855	
4	-0.3379		37.1242	<.0001*	4	-0.2450	
5	-0.0990		38.1962	<.0001*	5	-0.0093	
6	0.1190		39.7615	<.0001*	6	-0.0074	
7	0.1368		41.8510	<.0001*	7	-0.0511	
8	0.1510		44.4258	<.0001*	8	0.0761	
9	-0.0516		44.7299	<.0001*	9	-0.1445	
10	-0.1398		46.9823	<.0001*	10	-0.0087	
11	-0.1037		48.2369	<.0001*	11	-0.0516	
12	-0.0758		48.9149	<.0001*	12	-0.0702	
13	0.1459		51.4525	<.0001*	13	0.1500	
14	0.0429		51.6742	<.0001*	14	-0.2113	
15	-0.0344		51.8182	<.0001*	15	0.0338	
16	-0.0277		51.9127	<.0001*	16	-0.0579	
17	-0.1533		54.8473	<.0001*	17	-0.1522	
18	-0.0801		55.6579	<.0001*	18	0.0025	
19	0.2406		63.0593	<.0001*	19	0.2186	
20	0.1404		65.6087	<.0001*	20	-0.0992	
21	0.0093		65.6200	<.0001*	21	-0.0644	
22	-0.1444		68.3868	<.0001*	22	-0.0742	
23	-0.2288		75.4161	<.0001*	23	-0.0830	
24	-0.1794		79.7953	<.0001*	24	-0.1849	
25	0.0328		79.9436	<.0001*	25	0.0769	

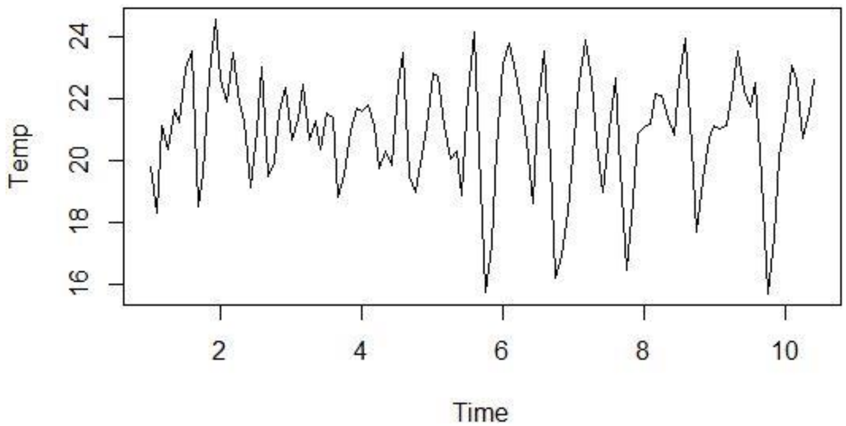
So, comparing b = 2, r = 1 and s = 1 (1,0,0)(0,1,0)$_{12}$ and b = 2, r = 1 and s = 1 (2,1,1)(1,1,0)$_{12}$. We find the later to be a better model because it has a lower AIC, BIC, MAPE and MAE.

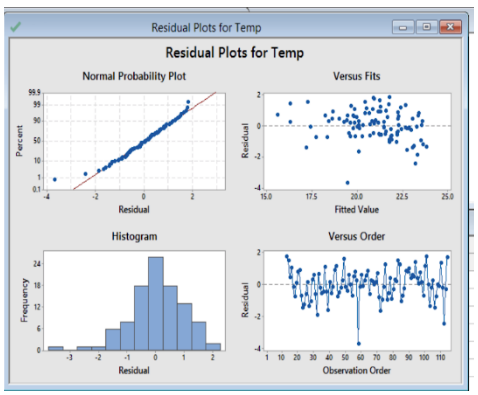
Q.5) Forecast the future streamflow of both monthly maximum temperature and electricity consumption for the next 6 months after Jun. 2006.

i) Using R to fit the ARIMA model for the temperature time series we get, Temp $(2,0,1)x(0,1,1)_{12}$

```r
library(forecast)
electricity <- read.csv("C:/Users/User/Desktop/520 project/elec.csv",header=TRUE)
Temperature <- read.csv("C:/Users/User/Desktop/520 project/Temp.csv",header=TRUE)
x1<- ts(Temperature,frequency = 12)
plot.ts(x1)

model_fit<-auto.arima(x1, seasonal =TRUE, ic = 'aic')
print(model_fit)
forecast(model_fit,6)
```

```
> plot.ts(x1)
>
> model_fit<-auto.arima(x1, seasonal =TRUE, ic = 'aic')
> print(model_fit)
Series: x1
ARIMA(2,0,1)(0,1,1)[12]

Coefficients:
         ar1      ar2     ma1     sma1
      0.6384  -0.4142  0.3819  -0.1795
s.e.  0.1631   0.1366  0.1770   0.1043

sigma^2 estimated as 0.7684:  log likelihood=-130.1
AIC=270.19   AICc=270.82   BIC=283.32
> forecast(model_fit,6)
       Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Jul 10       23.25970 22.13627 24.38312 21.54157 24.97783
Aug 10       23.35191 21.74694 24.95689 20.89732 25.80651
Sep 10       19.44388 17.81695 21.07080 16.95571 21.93204
Oct 10       15.67176 14.01654 17.32699 13.14032 18.20321
Nov 10       17.70992 16.02685 19.39299 15.13589 20.28395
Dec 10       20.26294 18.57848 21.94740 17.68678 22.83910
>
```



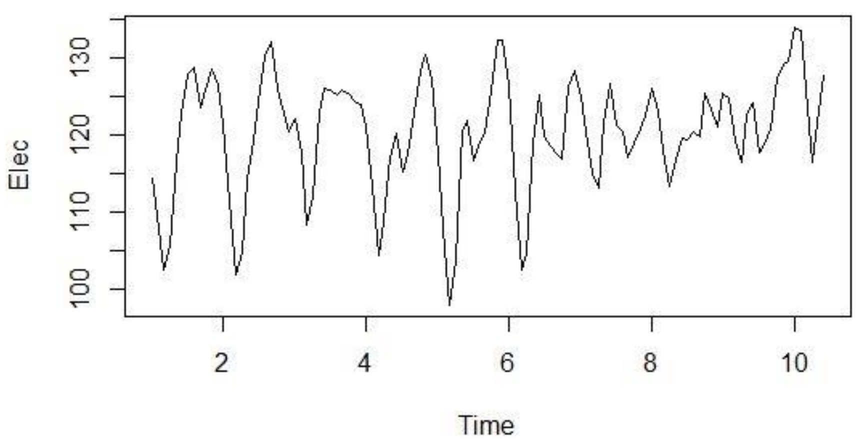


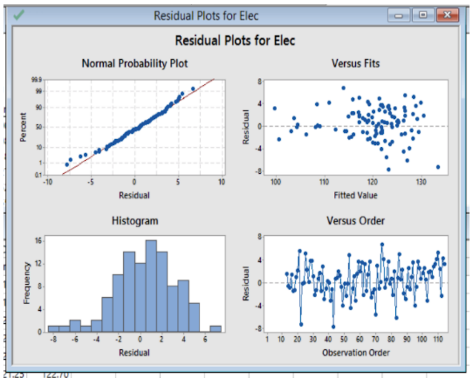
All the plots in the residual plots look pretty normal, hence, we can say that the model is adequate.

ii) Using R to fit the ARIMA model for the temperature time series we get, Elec $(2,0,0)x(1,1,1)_{12}$

```r
library(forecast)
electricity <- read.csv("C:/Users/User/Desktop/520 project/elec.csv",header=TRUE)
Temperature <- read.csv("C:/Users/User/Desktop/520 project/Temp.csv",header=TRUE)
x1<- ts(electricity,frequency = 12)
plot.ts(x1)

model_fit<-auto.arima(x1, seasonal =TRUE, ic = 'aic')
print(model_fit)
forecast(model_fit,6)
```

```
> plot.ts(x1)
>
> model_fit<-auto.arima(x1, seasonal =TRUE, ic = 'aic')
> print(model_fit)
Series: x1
ARIMA(2,0,0)(1,1,1)[12] with drift

Coefficients:
         ar1      ar2    sar1     sma1   drift
      1.0542  -0.4997  0.4408  -0.7504  0.0550
s.e.  0.0870   0.0854  0.2461   0.2271  0.0318

sigma^2 estimated as 8.281:  log likelihood=-252.04
AIC=516.08   AICc=516.97   BIC=531.83
> forecast(model_fit,6)
       Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Jul 10       122.5677 118.8781 126.2574 116.9249 128.2106
Aug 10       122.3426 116.9813 127.7038 114.1432 130.5419
Sep 10       122.4392 116.6223 128.2561 113.5431 131.3353
Oct 10       126.9475 121.1144 132.7806 118.0265 135.8685
Nov 10       128.7815 122.9103 134.6527 119.8023 137.7607
Dec 10       129.0084 123.0657 134.9511 119.9198 138.0970
```





In this model as well, all the residual plots look normal hence, the chosen model is pretty adequate.

Forecasts from Temperature time series:

Temp	**Forecast**	**Lo 80**	**Hi 80**	**Lo 95**	**Hi 95**
Jul 10	23.2597	22.13627	24.38312	21.54157	24.97783
Aug 10	23.35191	21.74694	24.95688	20.89732	25.8065
Sep 10	19.44388	17.81695	21.0708	16.95571	21.93204
Oct 10	15.67176	14.01654	17.32698	13.14032	18.20321
Nov 10	17.70992	16.02685	19.39299	15.13589	20.28396
Dec 10	20.26294	18.57848	21.9474	17.68678	22.8391

Forecasts from Electricity time series:

Elec	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jul 10	122.5677	118.8781	126.2574	116.9249	128.2106
Aug 10	122.3426	116.9813	127.7038	114.1432	130.5419
Sep 10	122.4392	116.6223	128.2561	113.5431	131.3353
Oct 10	126.9475	121.1144	132.7806	118.0265	135.8685
Nov 10	128.7815	122.9103	134.6527	119.8023	137.7607
Dec 10	129.0084	123.0657	134.9511	119.9198	138.097