DNSC 6219 TIME SERIES FORECATING

TIME SERIES PROJECT
LAND TEMPERATURE SERIES

SECTION 11
Chen Gao, Renlu Zhang, Xiang Li, Yuwei Yao, Haoning Liu

Part 1. Introduction and Overview

Our project topic is concerned about global temperature and climate trends. We found the monthly datasets from Kaggle "Climate Change: Earth Surface Temperature Data" and National Center for Environmental Information "Global Time Series- Climate at a Glance".

1.1 Attributes Description

- dt: Date starts in 1750 for average land temperature and 1850 for max and min land temperatures and global ocean and land temperatures.
- Land Average Temperature: global average land temperature in Celsius.
- Land Max Temperature: global average maximum land temperature in Celsius.
- Land Min Temperature: global average minimum land temperature in Celsius.
- Land And Ocean Average Temperature: global average land and ocean temperature in Celsius.
- Global Ocean Temperature Anomalies: global and hemispheric anomalies of ocean temperature with respect to the 20^{th} century average.
- Global Land Temperature Anomalies: global and hemispheric anomalies of land temperature with respect to the 20^{th} century average.
- Global Land and Ocean Temperature Anomalies: global and hemispheric anomalies of land and ocean temperature with respect to the 20^{th} century average.

1.2 Main series – Monthly Land Average Temperature

1.2.1 Series Plot

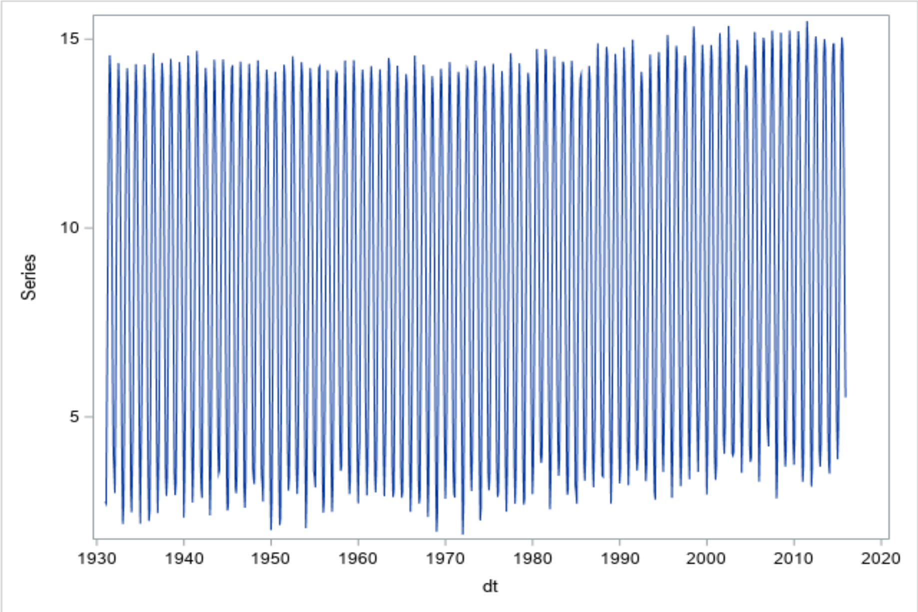

Figure 1.2.1 Land Average Temperature Plot

From the Land Average Temperature series plot (Figure 1.2.1), we can see the series has a slightly increasing trend over time.

1.2.2 Box Plot

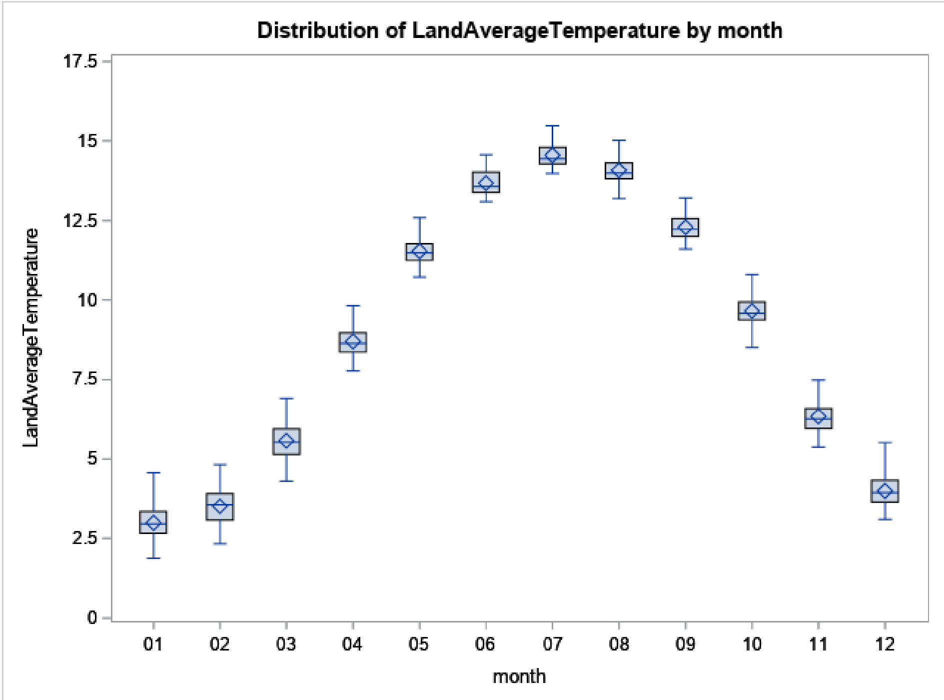

Figure 1.2.2 Land Average Temperature Box Plot

Based on the seasonal box plots (Figure 1.2.2), we see that the series has higher temperatures from May to September and comparatively low temperatures from November to March in the next year. This suggests the existence of seasonality in this series.

1.2.3 ACF Plot

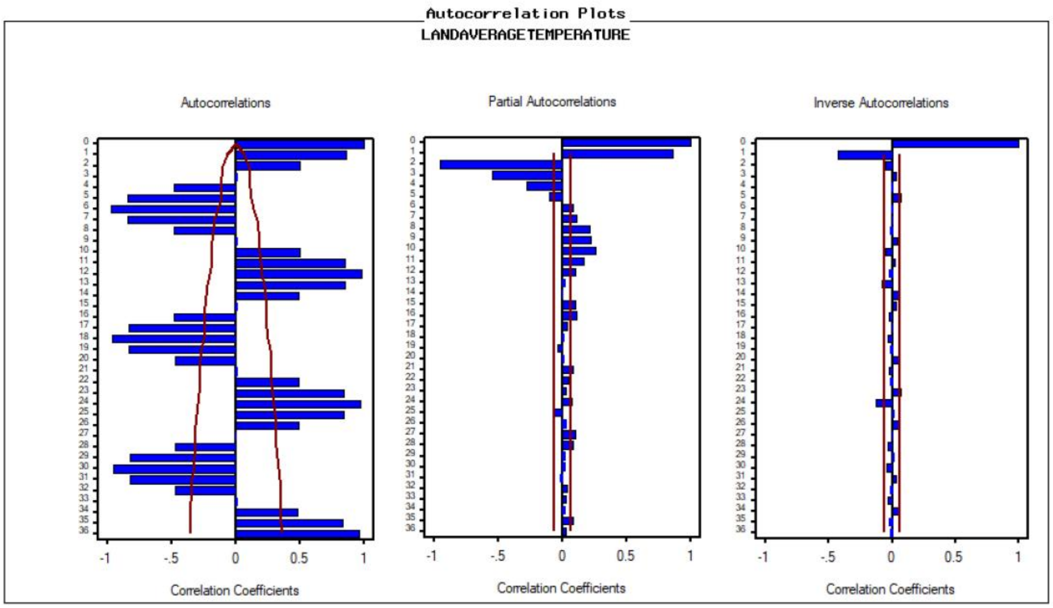

Figure 1.2.3 Land Average Temperature ACF Plot

The sample ACF plot (Figure 1.2.3) shows that the series is stationary in non-seasonal component since the ACF is decaying sinusoidally and non-stationary in seasonal component since the series is decaying slowly at lag 12, 24, 36.

1.3 Sample Period selection

- Series duration: January 1931 – December 2015
- Fit sample: January 1931 – December 2006
- Hold-out sample: January 2007 – December 2015

Part 2. Univariate Time-series Models

Reviewing the ACF of the original series (Figure 1.2.3), it is shown that the autocorrelation of seasonal lags (e.g., lag 12, lag 24, lag 36) are beyond two standard error bounds. Therefore, there is seasonality in this series.

2.1 Deterministic Time Series Models

2.1.1 Model with Seasonal Dummies and Linear Trend

Parameter Estimates
LANDAVERAGETEMPERATURE
Seasonal Dummies + Linear Trend

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	3.54600	0.0451	78.6418	<.0001
Seasonal Dummy 1	-0.99737	0.0569	-17.5327	<.0001
Seasonal Dummy 2	-0.46229	0.0569	-8.1266	<.0001
Seasonal Dummy 3	1.55369	0.0569	27.3126	<.0001
Seasonal Dummy 4	4.68603	0.0569	82.3770	<.0001
Seasonal Dummy 5	7.52633	0.0569	132.3080	<.0001
Seasonal Dummy 6	9.66768	0.0569	169.9522	<.0001
Seasonal Dummy 7	10.55862	0.0569	185.6151	<.0001
Seasonal Dummy 8	10.07554	0.0569	177.1232	<.0001
Seasonal Dummy 9	8.28102	0.0569	145.5768	<.0001
Seasonal Dummy 10	5.64421	0.0569	99.2228	<.0001
Seasonal Dummy 11	2.31492	0.0569	40.6955	<.0001
Linear Trend	0.0008328	0.000044	18.8818	<.0001
Model Variance (sigma squared)	0.12296	.	.	.

Figure 2.1.1 Seasonal Dummies and Linear Trend Parameter Estimate

After modeling the series with seasonal dummies and linear trend, we can find that the P value of all the coefficients is smaller than 0.05 (Figure 2.1.1). It can be concluded that the land temperature for January to November are statistically different than that of the reference month December and the series has an upward trend. Therefore, the Land Average Temperature series has a seasonal trend.

2.1.2 Model with Seasonal Dummies and Differenced Error Model

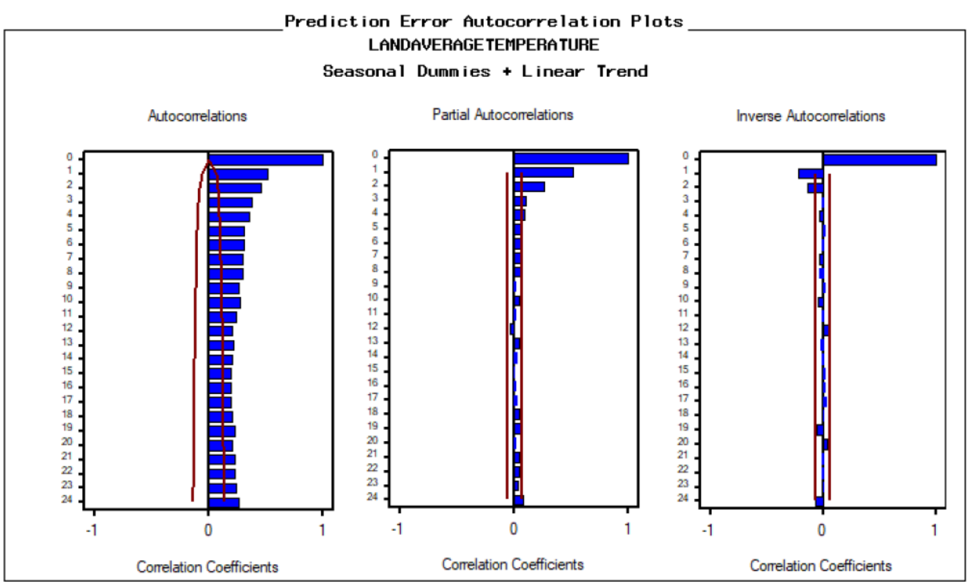

Figure 2.1.2.1 Seasonal Dummies and Linear Trend Residuals ACF Plot

From residual ACF plot of the seasonal dummies and linear trend (Figure 2.1.2.1), we find that AC decays slowly and the series is non-stationary. Thus, we take difference of the series to make residual stationary first. The differencing of the residual is achieved by taking difference to the entire series and the linear trend is dropped.

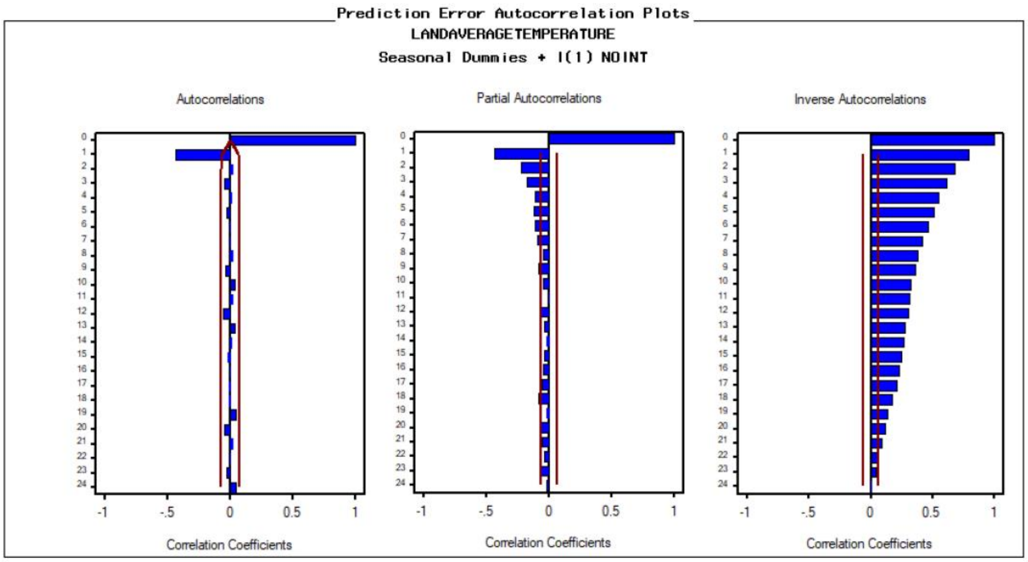

Figure 2.1.2.2 Differenced Seasonal Dummies and Linear Trend Residuals ACF Plot

After taking the difference of the original series and adding the first order difference, the ACF of the residual chops off after lag 1 and PACF decays quickly (Figure 2.1.2.2). Therefore, we firstly fit the MA (1) to the residuals.

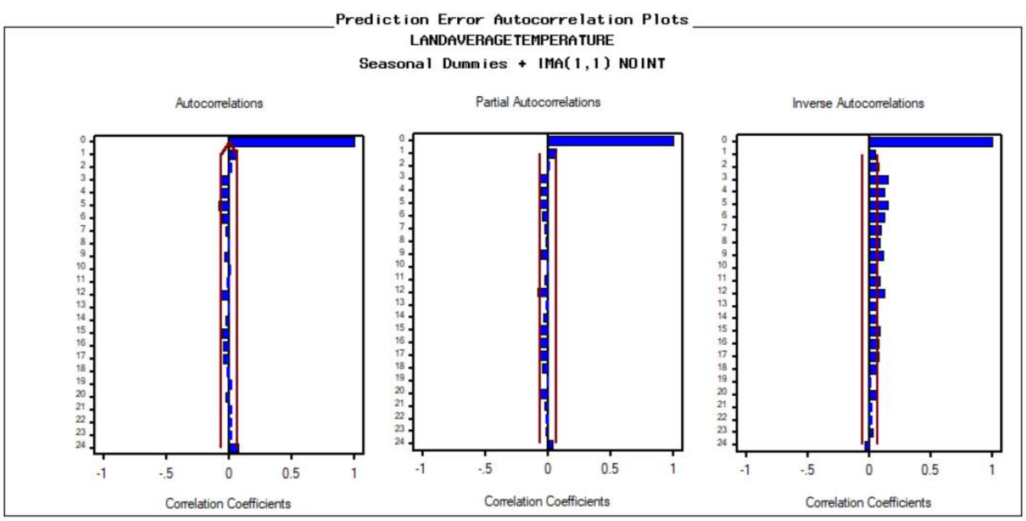

Figure 2.1.2.2 Differenced Seasonal Dummies Error Model Residuals ACF Plot

From Figure 2.1.2.2, the ACs of the residuals are within 2 standard error bound for most lags after modeling with MA (1). Thus, the residuals are practically white noise.

Parameter Estimates
LANDAVERAGETEMPERATURE
Seasonal Dummies + IMA(1,1) NOINT

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Moving Average, Lag 1	0.70803	0.0236	29.9779	<.0001
Seasonal Dummy 1	-0.99162	0.0420	-23.6058	<.0001
Seasonal Dummy 2	0.53591	0.0419	12.7857	<.0001
Seasonal Dummy 3	2.01682	0.0419	48.1174	<.0001
Seasonal Dummy 4	3.13317	0.0419	74.7515	<.0001
Seasonal Dummy 5	2.84113	0.0419	67.7840	<.0001
Seasonal Dummy 6	2.14218	0.0419	51.1085	<.0001
Seasonal Dummy 7	0.89178	0.0419	21.2761	<.0001
Seasonal Dummy 8	-0.48225	0.0419	-11.5056	<.0001
Seasonal Dummy 9	-1.79368	0.0419	-42.7939	<.0001
Seasonal Dummy 10	-2.63599	0.0419	-62.8897	<.0001
Seasonal Dummy 11	-3.32845	0.0419	-79.4105	<.0001
Seasonal Dummy 12	-2.31409	0.0419	-55.2099	<.0001
Model Variance (sigma squared)	0.08901	.	.	.

Fit Range: JAN1931 to DEC2006

Figure 2.1.2.3 Differenced Seasonal Dummies Error Model Parameter Estimates

Furthermore, the P value of parameters are all smaller than 0.05 (Figure 2.1.2.3). it can be concluded that all the parameters are statistically significant than 0. Thus, the model is appropriate for the Land Average Temperature Series.

2.1.3 Cyclical Model

In order to build the cyclical model, we firstly remove observation after 912 periods (108 hold-out periods) and use SAS command to estimate amplitude for each harmonic. Below Table 2.1.3.1, shows Top 15 periodogram values. We find that observations 2, 77 ,153, 22 and 305 have high P_01 value.

Obs	FREQ	PERIOD	P_01	Harmonics
77	0.5236	12	15630.2	76
2	0.00689	912	22.22	1
153	1.0472	6	11.86	152
22	0.14468	43.429	3.01	21
305	2.0944	3	2.19	304
229	1.5708	4	1.96	228
10	0.06201	101.333	1.95	9
457	3.14159	2	1.77	456
39	0.2618	24	1.56	38
5	0.02756	228	1.54	4
31	0.20668	30.4	1.51	30
78	0.53049	11.844	1.43	77
52	0.35136	17.882	1.29	51
9	0.05512	114	1.17	8
28	0.18602	33.778	1.16	27

Table 2.1.3.1 Harmonics Amplitude Table

The four harmonics with the highest amplitudes are 1, 76, 152 and 304. Then, we use these regressors to build the cyclical model. Since observation 22 is not significant when we build the model, we drop this variable.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	8.34680	0.02367	352.65	<.0001
TIME	1	0.00106	0.00004208	25.16	<.0001
COS1	1	0.21169	0.01420	14.91	<.0001
SIN1	1	0.10929	0.01784	6.13	<.0001
COS76	1	-4.83356	0.01392	-347.28	<.0001
SIN76	1	-3.30689	0.01392	-237.57	<.0001
COS152	1	-0.07669	0.01392	-5.51	<.0001
SIN152	1	-0.15562	0.01392	-11.18	<.0001
COS304	1	-0.02643	0.01392	-1.90	0.0578
SIN304	1	0.06311	0.01392	4.53	<.0001

Figure 2.1.3.2 Cyclical Model Parameter Estimate

From the parameter estimate (Figure 2.1.3.2), it can be seen that the P value of all the parameters are smaller than 0.05 except the cosine 304. Since the P value for this period is slightly higher than 0.05, we decide to keep this pair.

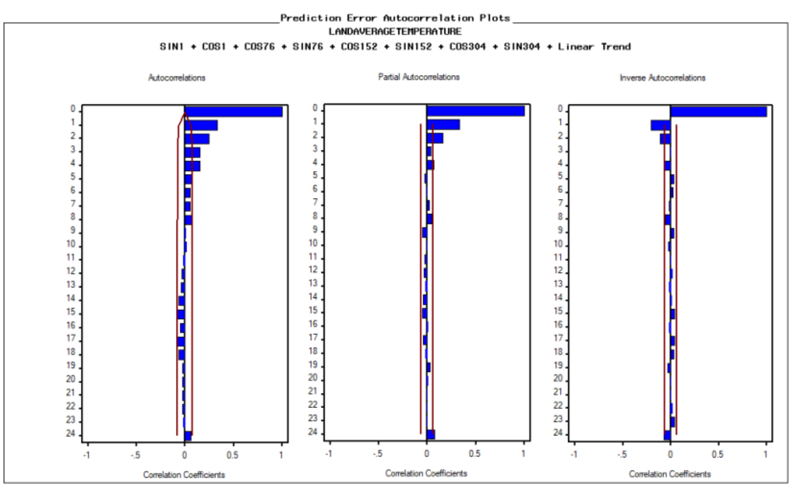

Figure 2.1.3.3 Cyclical Model Residuals ACF Plot

From the cyclical model, we can find that ACF of residuals decays quickly (Figure 2.1.3.3), this residuals series is stationary. Therefore, we can use error model to model residuals. Furthermore, PAC chops off after lag2 and we choose AR (2) for error model first.

2.1.4 Cyclical Model + Error Model

Option 1: AR (2) Error Model

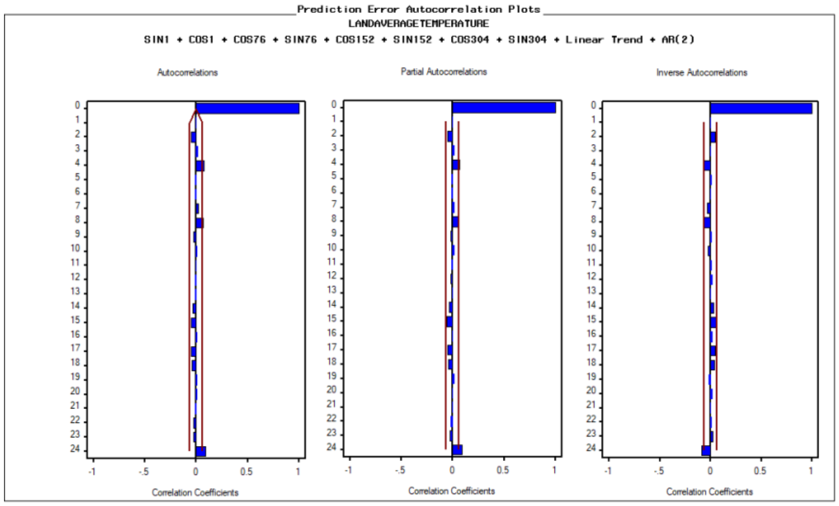

Figure 2.1.4.1 Cyclical Model + AR (2) Error Model Residuals ACF Plot

We firstly check the ACF of the residuals. The AC for residuals of error model is statiscally significantly less than 2 standard error bounds in pratical sense (Figure 2.1.4.1). Therefore, the series has been reduced to white noise.

Parameter Estimates
LANDAVERAGETEMPERATURE
SIN1 + COS1 + COS76 + SIN76 + COS152 + SIN152 + COS304 + SIN304 + Linear Trend + AR(2)

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	8.34204	0.0524	159.1090	<.0001
Autoregressive, Lag 1	0.27126	0.0328	8.2762	<.0001
Autoregressive, Lag 2	0.18697	0.0328	5.7040	<.0001
SIN1	0.11294	0.0403	2.8003	0.0062
COS1	0.21283	0.0253	8.3995	<.0001
COS76	-4.83056	0.0187	-258.0480	<.0001
SIN76	-3.30696	0.0187	-176.3929	<.0001
COS152	-0.07339	0.0133	-5.5280	<.0001
SIN152	-0.14319	0.0133	-10.7804	<.0001
COS304	-0.02888	0.0112	-2.5825	0.0113
SIN304	0.06321	0.0112	5.6530	<.0001
Linear Trend	0.00107	0.000108	9.9152	<.0001
Model Variance (sigma squared)	0.08641	.	.	.

Figure 2.1.4.2 Cyclical Model + AR(2)Error Model Parameter Estimate

From the parameter estimate table (Figure 2.1.4.2), we find that the p-values associated with all parameters are all less than 0.05(even for cosine of Harmonics 304). We can conclude that all the parameters are statistically different than 0.

In order to achieve a better fit, we try to use other error to find whether we could get a better fit. Then, we choose an ARMA (1,1) model for the error model.

Option2: ARMA (1,1) Error model

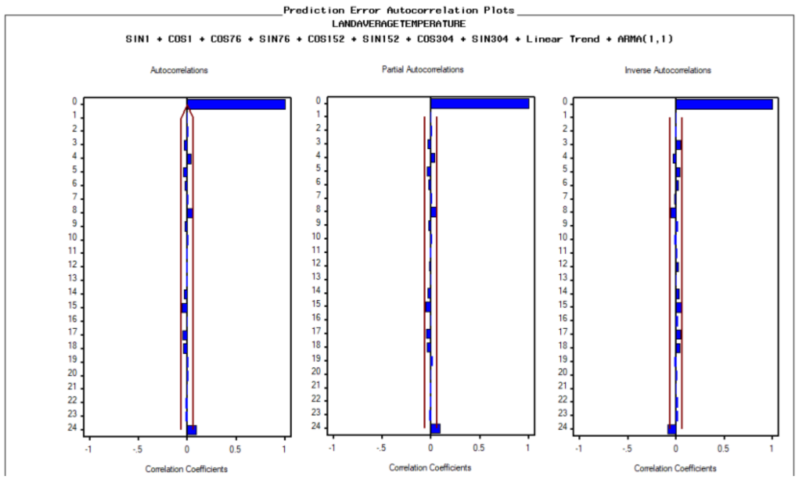

Figure 2.1.4.3 Cyclical Model + AR(1,1)Error Model Residuals ACF Plot

We find that ACs residuals of the error model are all within than two standard error bounds as Figure 2.1.4.2. Therefore, the series has been reduced to white noise.

Parameter Estimates
LANDAVERAGETEMPERATURE
SIN1 + COS1 + COS76 + SIN76 + COS152 + SIN152 + COS304 + SIN304 + Linear Trend + ARMA(1,1)

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	8.34051	0.0589	141.7090	<.0001
Moving Average, Lag 1	0.47937	0.0691	6.9412	<.0001
Autoregressive, Lag 1	0.75072	0.0520	14.4471	<.0001
SIN1	0.11395	0.0454	2.5117	0.0137
COS1	0.21289	0.0286	7.4499	<.0001
COS76	-4.83064	0.0169	-285.4226	<.0001
SIN76	-3.30687	0.0170	-195.0901	<.0001
COS152	-0.07333	0.0132	-5.5509	<.0001
SIN152	-0.14317	0.0132	-10.8335	<.0001
COS304	-0.02879	0.0118	-2.4360	0.0167
SIN304	0.06333	0.0118	5.3599	<.0001
Linear Trend	0.00107	0.000121	8.8673	<.0001
Model Variance (sigma squared)	0.08620	.	.	.

Fit Range: JAN1931 to DEC2006

Figure 2.1.4.4 Cyclical Model + AR(1,1)Error Model Parameter Estimate

The p-values associated with coefficients are all less than 0.05, which are statistically significant(Figure 2.1.4.4). We can conclude that all the coefficients are significantly different than 0.

Compared with AR (2) error model, ARMA (1,1) model get a better Root mean squared error. Therefore, we get a best Cyclical Model with ARMA (1,1) error model.

2.2 Seasonal ARIMA Model

Figure 2.1.4.4 shows that ACF of the main series (Land Average Temperature) decays slowly at seasonal lags showing that the series is nonstationary at both seasonal lags. Therefore, the series is taken the first difference at the seasonal lags.

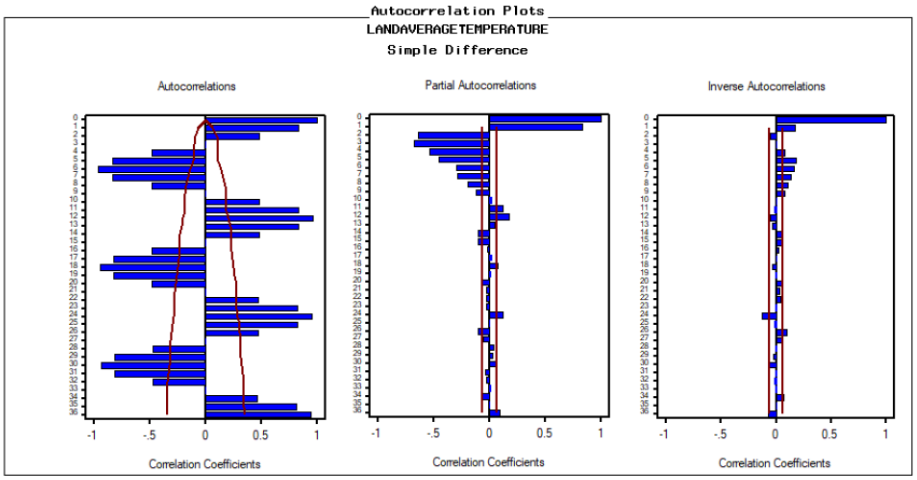

Figure 2.2.1 Land Average Temperature ACF Plot

After taking the seasonal difference to the Land Average Temperature series, it is shown (Figure 2.2.2) that the ACF at non-seasonal lags are decaying slowly as well. The series is also not stationary at non-seasonal lags. Therefore, we take the difference to the non-seasonal lags.

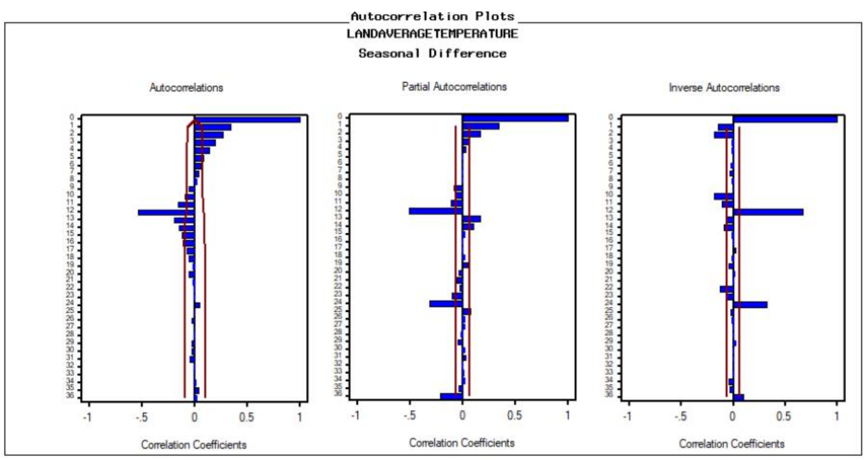

Figure 2.2.2 Land Average Temperature (seasonal differenced) ACF Plot

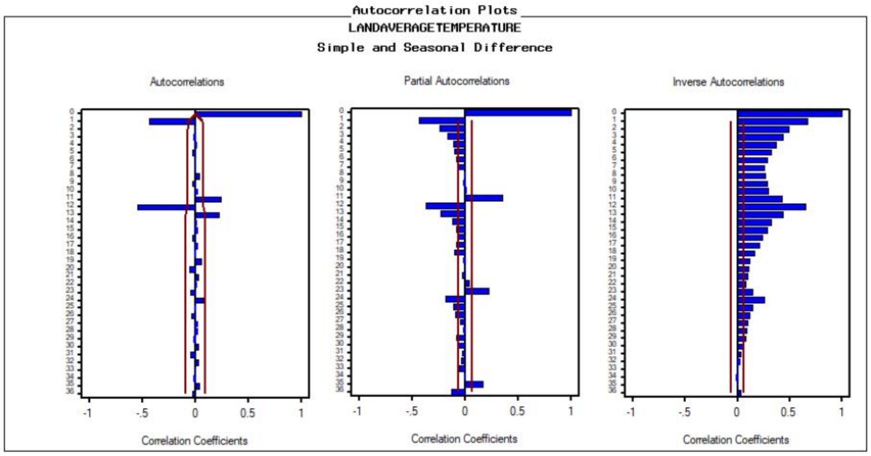

Figure 2.2.3 Land Average Temperature (seasonal and non-seasonal differenced) ACF Plot

After both differencing, the ACF of the series decays quickly at the seasonal and non-seasonal lags and the series is stationary (Figure 2.2.2). For the non-seasonal lags, the ACF chops off after lag 1 and the PACF decays quickly. We first fit the MA (1) to the differenced the non-seasonal lags. For the seasonal lags, the ACF chops off at the first lag (lag 12). We also fit the MA (1) to the differenced seasonal lags.

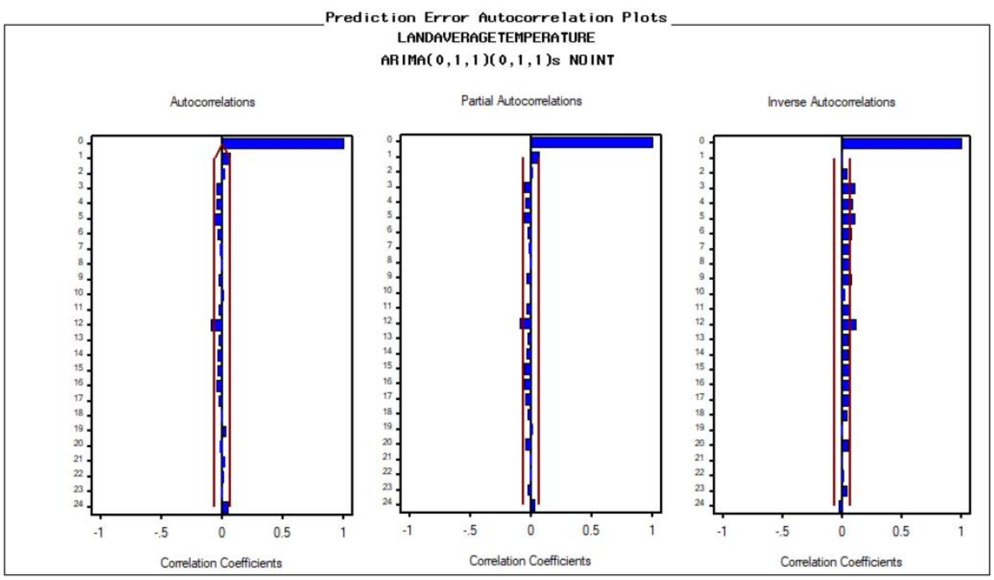

Figure 2.2.3 Seasonal ARIMA (0,1,1) (0,1,1) Residual ACF Plot

The ACF of the residual is inside the two standard error bounds of the most lags, but slightly outside two standard bounds at lag 1 and lag 12 as Figure 2.2.3. It can be defined that the residuals are white noise practically.

Parameter Estimates
LANDAVERAGETEMPERATURE
ARIMA(0,1,1)(0,1,1)s NOINT

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Moving Average, Lag 1	0.71482	0.0231	30.9076	<.0001
Seasonal Moving Average, Lag 12	0.95101	0.0142	66.7895	<.0001
Model Variance (sigma squared)	0.09042	.	.	.

Fit Range: JAN1931 to DEC2006

Figure 2.2.4 Seasonal ARIMA (0,1,1) (0,1,1) Parameter Estimate

Furthermore, the P value of all parameters are smaller than 0.05, which means all the parameters are statistically significant (Figure 2.2.4). And the Seasonal ARIMA (0,1,1) (0,1,1) model is appropriate for the series.

2.3 Model Comparison

2.3.1 Fit Sample (January 1931 to December 2006)

	Seasonal Dummies and Trend Model	**Seasonal Dummies + Differencing error model (ARIMA(0,1,1))**	**Cyclical Trend Model**	**Cyclical Trend Model + ARMA (1,1) Error Model**	**Seasonal ARIMA (1,0,1)(0,1,1) Model**
Mean Square Error	0.12121	0.08788	0.09946	0.08508	0.09426
Root Mean Square Error	0.34815	0.29644	0.31537	0.29168	0.30703
Mean Absolute Percent Error	4.78034	4.19875	4.37522	4.08762	4.29984

Mean Absolute Error	0.27687	0.23072	0.24466	0.22495	0.23869
R-Square	0.993	0.995	0.994	0.995	0.995
Model Variance (Sigma squared)	0.12296	0.08901	0.10056	0.0862	0.09042

Table 2.3.1 Model Comparison with Fit Period

As the table 2.3.1 shows, our five models have similar R-square and model variance. But since the cyclical trend model with ARMA (1,1) error model has relatively lower errors, it is the best model estimated by the fit period.

2.1.1 Hold-out Sample (January 2007 to December 2015)

	Seasonal Dummies and Trend Model	**Seasonal Dummies + Differencing error model (ARIMA(0,1,1))**	**Cyclical Trend Model**	**Cyclical Trend Model + ARMA (1,1) Error Model**	**Seasonal ARIMA (0,1,1)(0,1,1) Model**
Mean Square Error	0.20478	0.09153	0.0849	0.08216	0.09293
Root Mean Square Error	0.45252	0.30254	0.29137	0.28663	0.30485
Mean Absolute Percent Error	5.17597	3.64980	3.43306	3.40512	3.70530
Mean Absolute Error	0.38781	0.22786	0.21233	0.21379	0.22712
R-Square	0.988	0.995	0.995	0.995	0.995
Model Variance	0.12296	0.08901	0.10056	0.0862	0.09042

(Sigma squared)					

Table 2.3.2 Model Comparison with Hold-out Period

Furthermore, we use hold-out samples to evaluate models as table 2.3.2. The re We compare all five models and it turns out that that best model is the 4th model with the lowest error and constant level of R-square and model variance. Thus, we choose cyclical trend model with ARMA (1,1) error model as the best model here.

Part 3. Multivariate Time Series Models

3.1 Variable Selection

In this part, the dependent variable is Global Land Average Temperature, this series is both seasonal and non-seasonal nonstationary, so we take the first and seasonal difference to make it stationary. Excluding main variable, there are six variables in the data set. Since we consider the land max temperature, land min temperature, land and ocean average temperature are similar to our original variable, we exclude those three variables first. For the remaining three variables, we choose global land temperature anomalies and global ocean temperature anomalies as two independent variables to develop the transfer function model separately for the reason that we think these two anomalies may have some impact on the land average temperature.

3.2 Transfer Function Model Estimate

3.2.1 Global Land Temperature Anomalies

We firstly take a look at the Global Land Temperature Anomalies series. From the ACF plot (Figure 3.2.1.1), we can find this series is nonstationary, therefore, we firstly take the first difference to make the series stationary.

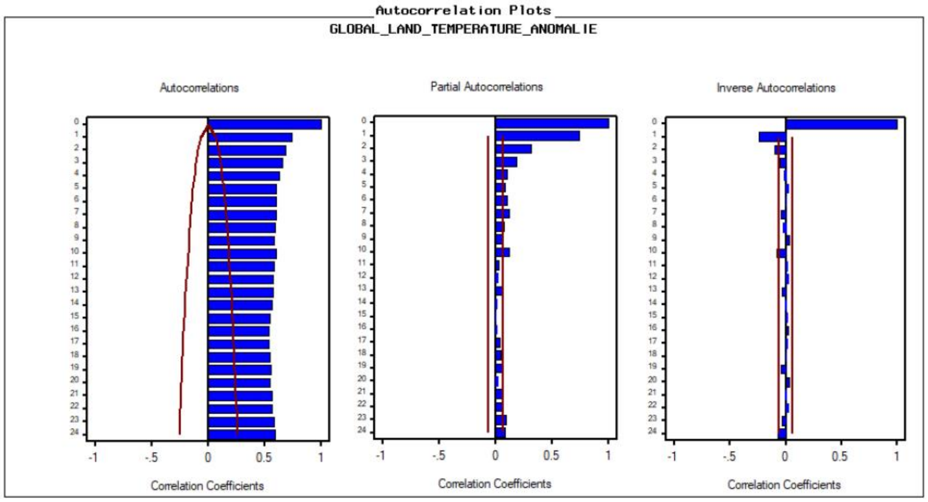

Figure 3.2.1.1 Global Land Temperature Anomalies ACF Plot

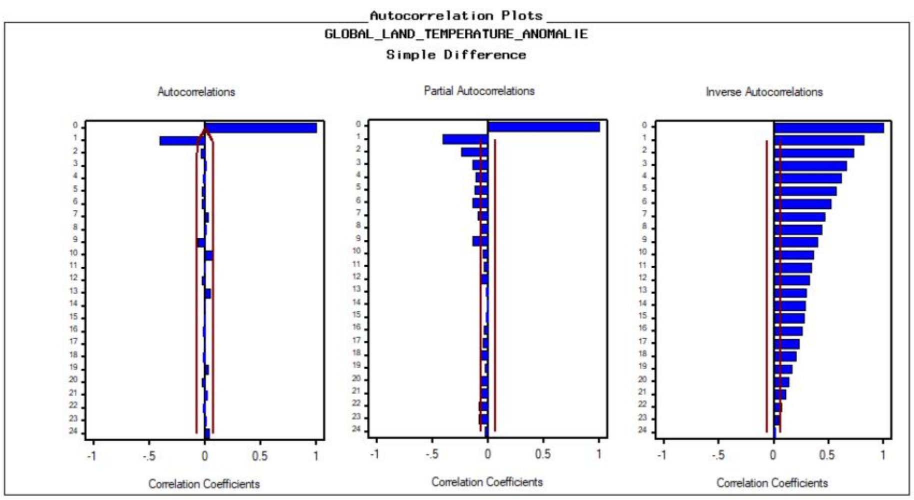

Figure 3.2.1.2 Global Land Temperature Anomalies ACF Plot After 1st Differencing

Then we check the ACF of the differenced series, the ACF (Figure 3.2.1.1) at some lags are still outside two standard error bounds, which indicate the series is not white noise series. After researching the ACF, PACF and IACF of the differenced series, we decide to use ARIMA (1,1,1) to prewhiten the series.

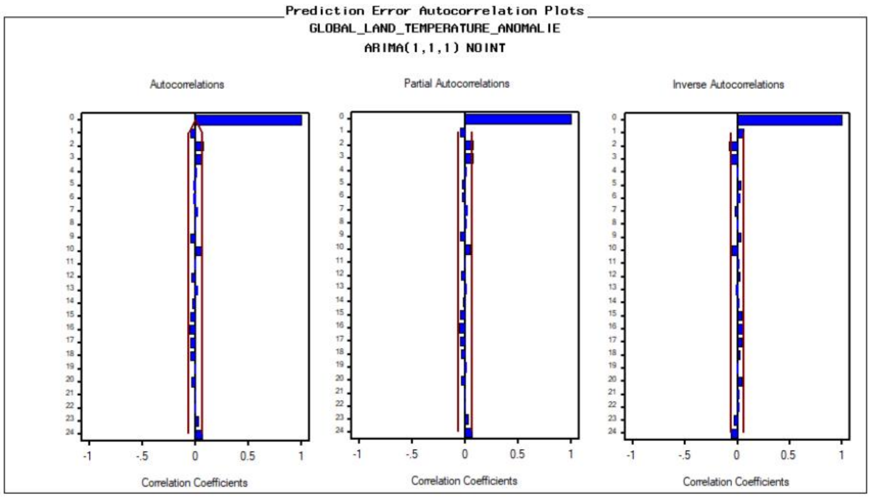

Figure 3.2.1.3 ACF of ARIMA (1,1,1) Prewhitening

The AC of residuals (Figure 3.2.1.3) are mostly within two standard error bounds and the series becomes white noise series practically thus we can use it to get the transfer function. After prewhitening, we try to find the CCF between Land Average Temperature and Global Land Temperature Anomalies. Since the dependent variable Land Average Temperature is nonstationary at both seasonal and non-seasonal lags, we need to take the seasonal and non-seasonal difference to make the series stationary using SAS command. Since we also need to difference the independent variable, we create a new variable DIFTEMP of the seasonal differenced Land Average Temperature series and use it to the following models.

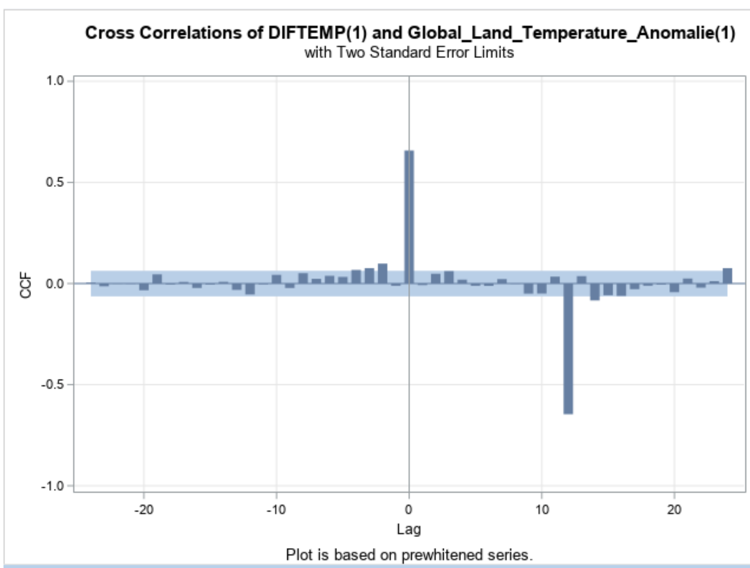

Figure 3.2.1.4 CCF Plot between Land Average Temperature and Global Land Temperature Anomalies

From the CCF plot (Figure 3.2.1.4), the first response is at lag 0, so b=0. The sample CCF chopped of after lag 12 (which is the first seasonal lag), so r=0 and s=12. We can conclude that global land temperature has a seasonal influential effect on the land average temperature.

Crosscorrelation Check of Residuals with Input Global_Land_Temperature_Anomalie

To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	3.56	4	0.4681	-0.007	-0.017	0.030	-0.046	0.012	-0.004
11	16.33	10	0.0906	-0.039	0.009	0.012	0.027	-0.088	0.049
17	20.63	16	0.1932	-0.001	0.013	-0.040	0.039	-0.031	-0.002
23	29.08	22	0.1427	0.027	0.029	-0.051	0.008	0.048	-0.044
29	34.43	28	0.1872	0.050	-0.040	0.010	0.003	0.007	0.033
35	39.01	34	0.2546	-0.041	0.016	-0.005	0.010	-0.034	0.036
41	39.96	40	0.4720	-0.022	0.009	0.002	-0.012	-0.003	0.015
47	42.70	46	0.6115	-0.007	-0.004	0.032	-0.023	0.031	-0.013

Figure 3.2.1.5 CCF between Global Land Temperature Anomalies and Residuals

Furthermore, we check the cross correlation between residuals and independent variable Land Average Temperature Anomalies(Figure 3.2.1.5). The P value is larger than 0.05 from lag 1 to lag 47. Thus, we can reject the null hypothesis that the Land Average Temperature Anomalies have a correlation with residuals.

Then we make the TF model in the forecasting system.

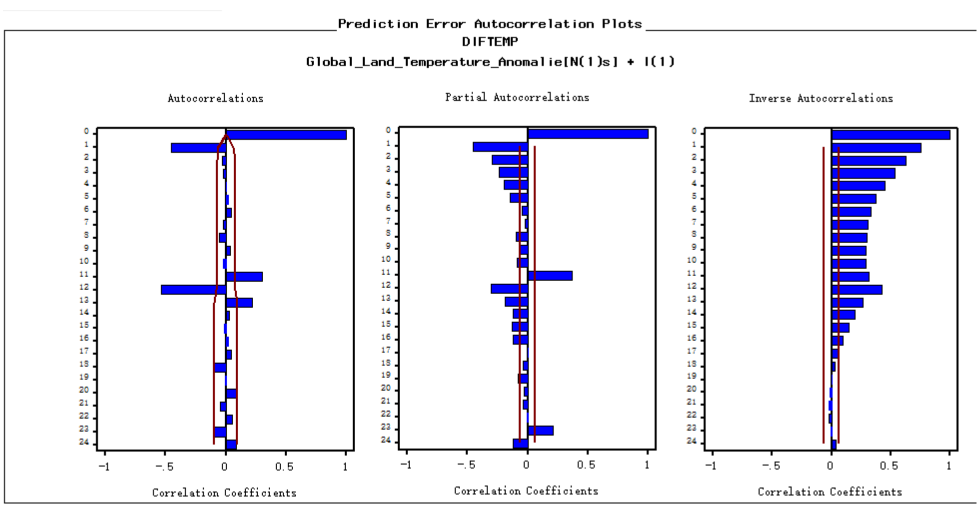

Figure 3.2.1.6 Global Land Temperature Anomalies (with seasonal differencing) Residuals ACF Plot

From the ACF plot (Figure 3.2.1.6), we can see the residuals of the model are not white noise series, so we need to identify a model to reduce them to white noise.

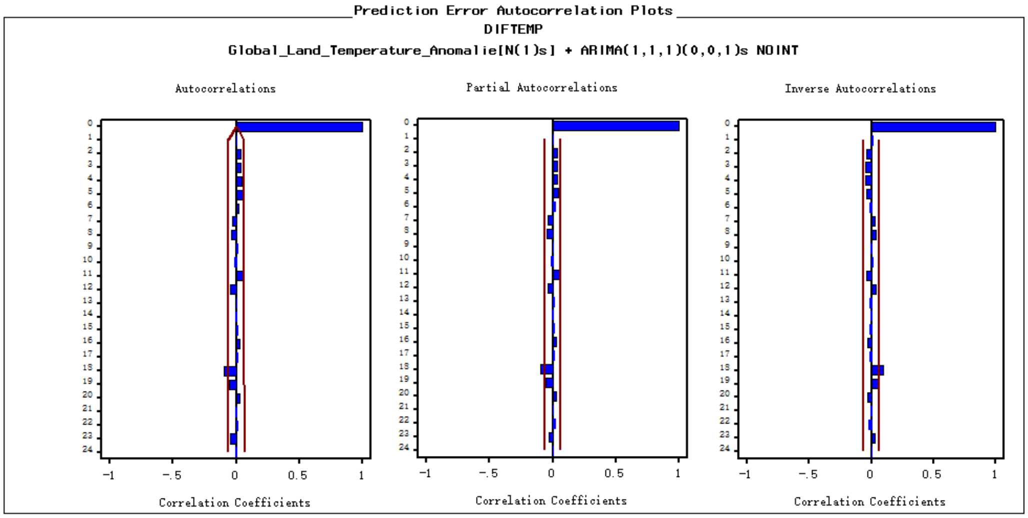

Figure 3.2.1.7 ACF Plot of noise-TF model

We add a first autoregressive order, a first and seasonal moving average order to the model. The ACF plot of residuals, are mostly within 2 standard error bounds as Figure 3.2.1.7. And we can conclude the residuals become white noise.

Parameter Estimates
DIFTEMP
Global_Land_Temperature_Anomalie[N(1)s] + ARIMA(1,1,1)(0,0,1)s NOINT

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Moving Average, Lag 1	0.97150	0.0094	103.6305	<.0001
Seasonal Moving Average, Lag 12	0.97896	0.0205	47.7832	<.0001
Autoregressive, Lag 1	0.17755	0.0340	5.2295	<.0001
GLOBAL_LAND_TEMPERATURE_ANOMALIE[N(1)s]	0.92572	0.0119	77.8096	<.0001
GLOBAL_LAND_TEMPERATURE_ANOMALIE[N(1)s]	0.92702	0.0122	76.1195	<.0001
Model Variance (sigma squared)	0.01197	.	.	.

Figure 3.2.1.8 Parameters of noise-TF model

Furthermore, the P value of parameters are all less than 0.05(Figure 3.2.1.8). Therefore, we can conclude that all the parameters are statistically difference than 0. The TF model and TF-noise are appropriate for the series after the seasonal differencing.

3.2.2 Global Ocean Temperature Anomalies

Then we look at the variable Global Ocean Temperature Anomalies. From the ACF of the series (Figure 3.2.2.1) This series is not stationary, so we first difference it.

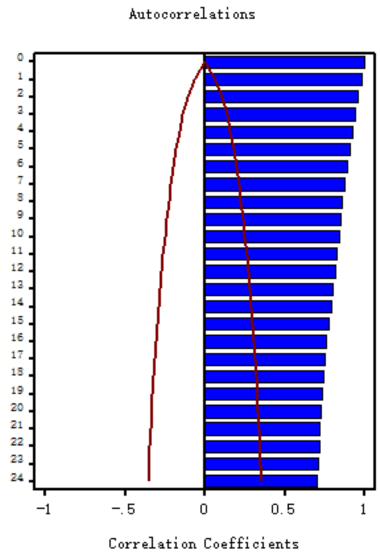

Figure 3.2.2.1 Global Ocean Temperature Anomalies ACF Plot

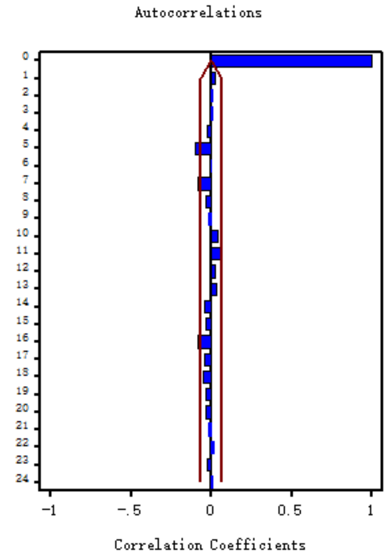

Figure 3.2.2.2 Global Ocean Temperature Anomalies ACF Plot After 1st Differencing

After first difference, we found the ACF of the differenced series is a white noise series in a practical sense (Figure 3.2.2.1). Therefore, we can use the differenced series to check the CCF between the Land Average Temperature and Global Ocean Temperature Anomalies.

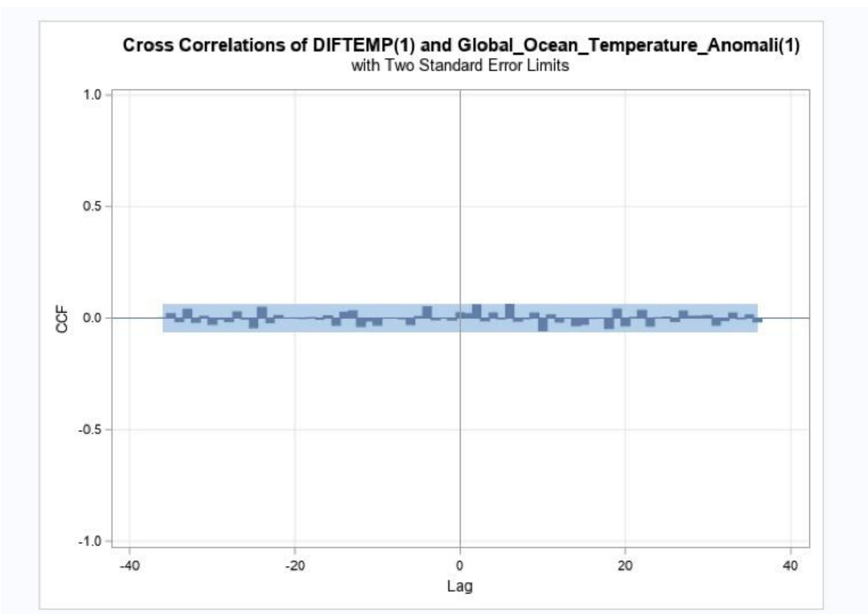

Figure 3.2.2.3 CCF between Land Average Temperature and Global Ocean Temperature Anomalies

From the CCF plot (Figure 3.2.2.3), the sample CCF is not significant both at negative and positive lags. It shows that there is no relationship between Land Average Temperature and

Global Ocean Temperature Anomalies. Therefore, we cannot define a TF model between these two variables.

3.3 Further Analysis of Contemporary Regression

Since the independent variable Global Ocean Temperature Anomalies cannot be used to model the transfer function model. We conduct the further research to see if they can do a contemporary regression between those variables.

We firstly regress Global Ocean Temperature Anomalies on Land Average Temperature. The ACs of residuals decay quickly as below Figure 3.2.3.1, therefore, the residuals are stationary. The ACs of residuals are outside two standard error bounds. We added an Error model of Seasonal ARMA (1,0,1) (0,1,1), and the residuals are white noise.

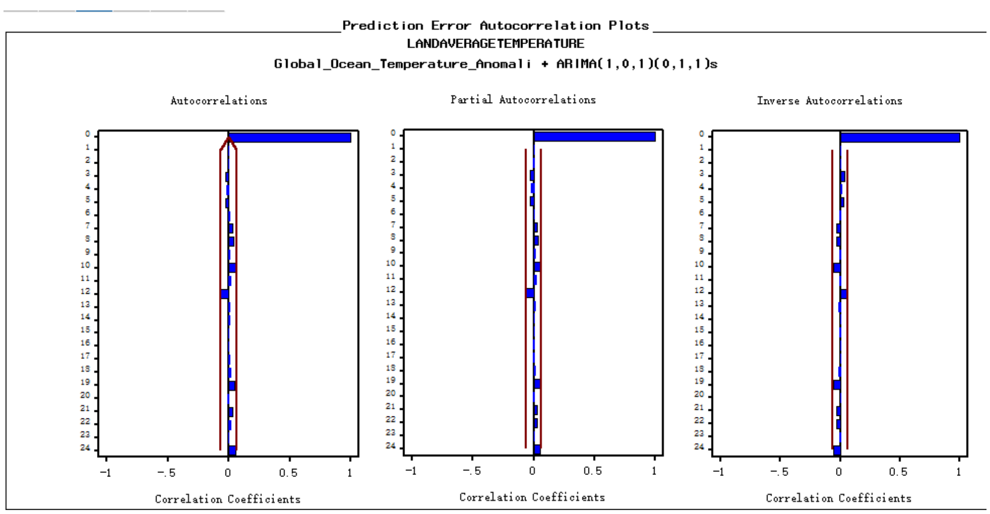

Figure 3.2.3.1 Land Average Temperature and Global Ocean Temperature Anomalies Contemporary Regression Residuals ACF Plot

Parameter Estimates
LANDAVERAGETEMPERATURE
Global_Ocean_Temperature_Anomali + ARIMA(1,0,1)(0,1,1)s

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	0.00407	0.0020	1.9879	0.0495
Moving Average, Lag 1	0.54351	0.0600	9.0595	<.0001
Seasonal Moving Average, Lag 12	0.94139	0.0158	59.4400	<.0001
Autoregressive, Lag 1	0.80468	0.0429	18.7775	<.0001
Global_Ocean_Temperature_Anomali	0.89757	0.1394	6.4387	<.0001
Model Variance (sigma squared)	0.08398	.	.	.

Figure 3.2.3.2 Land Average Temperature and Global Ocean Temperature Anomalies Contemporary Regression Parameter Estimate

From the parameter estimate (Figure 3.2.3.2), we found that the P value of all the parameters is smaller than 0.05, which means they are statistically different than 0.

For Global Land Temperature Anomalies and Global Ocean Temperature Anomalies together, after regress them on the main variable, we add an error model and seasonal trend of ARIMA (1,0,1)(0,1,1)s, and the residuals become white noise (Figure 3.2.3.3).

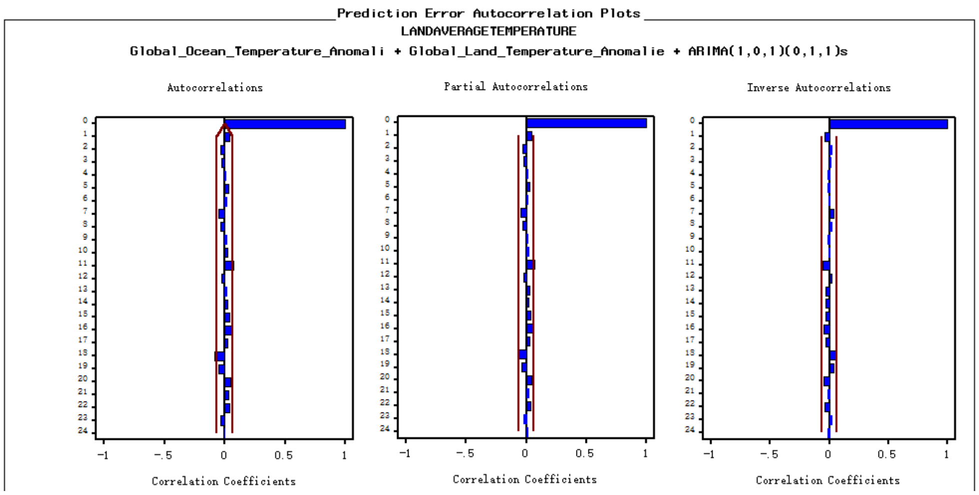

Figure 3.2.3.3 Land Average Temperature and Global Ocean Temperature Anomalies & Global Land Temperature Anomalies Contemporary Regression Residuals ACF Plot

Parameter Estimates

LANDAVERAGETEMPERATURE

Global_Ocean_Temperature_Anomali + Global_Land_Temperature_Anomalie + ARIMA(1,0,1)(0,1,1)s

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	-0.00176	0.000456	-3.8499	0.0002
Moving Average, Lag 1	0.65460	0.0772	8.4778	<.0001
Seasonal Moving Average, Lag 12	0.97310	0.0160	60.6650	<.0001
Autoregressive, Lag 1	0.80796	0.0608	13.2870	<.0001
Global_Ocean_Temperature_Anomali	0.12750	0.0445	2.8627	0.0051
Global_Land_Temperature_Anomalie	0.91181	0.0120	75.7934	<.0001
Model Variance (sigma squared)	0.01170	.	.	.

Figure 3.2.3.4 Land Average Temperature and Global Ocean Temperature Anomalies & Global Land Temperature Anomalies Contemporary Regression Parameter Estimates

From the parameter estimation, we found that the P value of all the parameters are smaller than 0.05 (Figure 3.2.3.4), which means they are statistically different than 0. Thus, we can use the contemporary regression with error model to estimate the Land Average Temperature.

Part 4. Conclusion

- The land average temperature shows a seasonal pattern and every year the average temperature increases slowly.
- Cyclical model with ARMA (1,1) error model is the best to model global land average temperature in terms of the error magnitude based on hold-out sample.

- When we use the cyclical model with ARMA (1,1) error model, we found there exists some hidden period, for example the Observation 77, in the land average temperature series. However, it may be difficult to explain.
- After building multivariate time series models, it can be concluded that the non-seasonal differenced Global Land Temperature Anomalies influences seasonal and non-seasonal differenced series of Land Average Temperature.
- We further compare the model we performed best in the univariate part with the TF model. The result is as table 4.1.

	Cyclical Trend Model + ARMA (1,1) Error Model	**Transfer Function Model(Land Average Temperature & Global Land Temperature Anomalies)**
Mean Square Error	0.08508	0.01372
Root Mean Square Error	0.29168	0.11715
Mean Absolute Percent Error	4.08762	77.18531
Mean Absolute Error	0.22495	0.9912
R-Square	0.995	0.926
Model Variance (Sigma squared)	0.0862	0.01197

Table 4.1 Model Comparison

The two models is superior than each other based on the different error measurement. Therefore, the Cyclical Trend Model with ARMA (1,1) Error Model and transfer function model are the two best models to describe Land Average Temperature series.

Data Source

https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data

https://www.ncdc.noaa.gov/cag/global/time-series/globe/land/2/1/1880-2019