A Time Series Analysis Study on Sea Surface Temperature (SSTs) Prediction

Xuechang Liu

May 2016

Abstract

A 34-year monthly sea surface temperature (SST) time series dataset is used in order to find out a best model that fits it and can make reasonable prediction. Multiple AR, MA, ARMA, ARIMA, ARIMAX, seasonal ARIMA and seaonal ARIMAX models are formulated and tested with different terms and moments. Another time series data southern oscillation index (SOI) is utilized as the extermal regressor in the ARIMAX and SARIMAX models. Based on the AIC and BIC values calculated for each model and the diagnosis analyses, the best fitting model is determined as SARIMA$(0,1,26)(0,1,1)_{12}$ model with a AIC value of 19.18. The model is then fitted and compared with real data and the fittng curve shows a good agreement. Though the model does not precisely predict the abnormal increase in temperature caused by the extreme El Niño event occurred in last year, the real data curve is still within the confidence interval.

1 Introduction

The atmosphere, hydrosphere and cryosphere form a fully coupled climate system. Sea surface temperature (SST) is the water temperature close to the ocean's surface. It is also a key climate and weather measurement obtained by satellite microwave radiometers, infrared (IR) radiometers, in situ moored and drifting buoys, and ships of opportunity. SST affects the behavior of the Earth's atmosphere above, so their

initialization into atmospheric models is important. While sea surface temperature is important for tropical cyclogenesis, it is also important in determining the formation of sea fog and sea breezes. This very direct measurement of the temperature change over ocean, especially the long time record of the variation of tens of years' data, can give us more information on the trends and change thus a better prediction for future temperature. The atmosphere and ocean system exhibits a number of large-scale phenomena that has large impact on both sea currents and climate change, such as the El Niño – Southern Oscillation (ENSO), Atlantic Multidecadal Oscillation (AMO), Arctic Oscillation (AO), etc. A climate pattern is any recurring characteristic of the climate, and may come in the form of a regular, a quasi-periodic (like El Niño) or an irregular event. A mode of variability is a climate pattern with identifiable characteristics, specific regional effects, and often oscillatory behavior. While these modes of variability are not exactly periodic, they are oscillatory in character, and their state is monitored using so-called climate indices that represent the general climatic state of a region affected by a given climate pattern. Each of these scalar indices is a combination of several climate variables [*Viron et al.*, 2013]. These patterns can be seen as the proxy of the climate change of the regions that are affected by certain variables.

Numerous authors have studied such teleconnections and the, possibly lagged, correlations between up to four indices [*Ambaum et al.*, 2001; *Mooley and Munot*, 1993; *Tsonis et al.*, 2007; *Wyatt et al.*, 2012; *Wang et al.*, 2012]. Other studies have considered the dynamics of coupled nonlinear oscillators as a possible source for such correlations [*Ghil and Mo*, 1991; *Kimoto and Ghil*, 1993; *Feliks et al.*, 2010] or studied the network of time series of a single climate field, such as the SSTs, at the nodes of a regular grid [*Tsonis and Swanson*, 2008; *Donges et al.*, 2009]. Here, I use a set of for time intervals that range between 1981 and 2015, to study the predictability of SST data and the possibility to use its relation between ENSO indices - Southern Oscillation Indices (SOIs) to forecast it. In this study, I hypothesize that (1) it is possible to find out a proper time series model that uses past SST record to fit and predict its behavior in the future; (2) it is also possible to use SOIs as a regressor to forecast SSTs.

To confine the study to a reasonable field and specify the outcoming results, the SSTs used in this project only come from most ENSO-affected area. El Niño, the warming phase of ENSO events, is a warming of the central to eastern tropical Pacific that occurs every two to seven years, on average. The term also commonly refers to the atmospheric rearrangements that occur with the oceanic warming. During an El Niño event, sea surface temperatures across a watery expanse often as large as the United States can warm by 1–3°F or more for a period of from a few months to a year or two. To qualify as an El Niño event, according to the definition used by the National Oceanic and Atmospheric Administration (NOAA), SSTs must remain at or above 0.5°C (about 1°F) for at least three months across the region labeled Niño 3.4 (defined by NOAA in Fig. 1), which is the region chosen for this study. The location of this region is [5°S - 5°N, 170°W - 120°W]

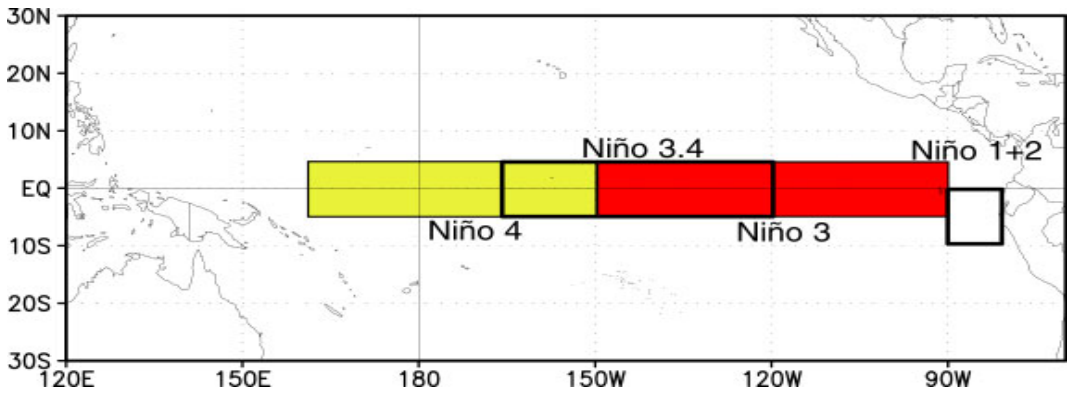

Figure 1: NOAA defined El Niño identification region. Niño 3.4 is shaded in the box with red and yellow.

In this report[1], I will first describe my datasets and how they are obtained and the hypothesis I wish to test in Section 2. In Section 3 I investigate the connections between the two time series data and calculated important statistical moments including their autocorrelation, cross-correlation, periodograms and compare some fitting models of the data. Results and discussion will be presented in Section 4.

2 Data and Hypothesis

The two time series datasets used in this project are assumed to be independently collected in a random fashion over time and each variable is indexed according to the order it is obtained in time. Each datapoint is indexed as once per month during this period of 34 years and this gives a total of 406 time indices. These time indices are given as natural numbers from 1 (December 1981) to 406 (September 2015). Thirty-four year monthly SST data is obtained from NOAA ESRL. The monthly data is packed in Network Common Data Form (NetCDF), derived by a linear interpolation of the weekly optimum interpolation (OI) version 2 fields to daily fields then averaging the daily values over a month. The OI SST analysis is produced weekly on a one-degree grid. The analysis uses multi-platform measurements from in situ (ships, buoys, etc.) and satellite SST's plus SST's simulated by sea-ice cover. Before the analysis is computed, the satellite data is adjusted for biases using Reynolds and Marsico methods. The dates used in the netCDF files are the dates of the beginning of the 7-day average period. For the more recent period, 1990-present, the weeks are defined to be centered on Wednesday. During the period 1981-1989, the weeks are centered on Sunday. The data is gridded by each latitude and longitude, but only the NOAA defined Niño 3.4 area data is averaged as one point per month for this study. Monthly climate indices data (SOI, in this case) are downloaded from NOAA Climate Prediction Center (CPC). Monthly SOI is updated on 10th of each month, calculatedd with the sea level pressure (SLP) values from Darwin and Tahiti station. The calculation is give as:

$$STD_{Tahiti} = \sqrt{\frac{\Sigma_1}{N}}$$

where,

[1]Based on suggestion and comments given by Pegah Fakhari, this project has gone through major revisions - the hypothesis that made eralier this semester has been changed.

Σ_1 = the sum of all TA^2.
STD - Standard Deviation
TA - Tahiti anomaly = actual(SLP) - mean(SLP)).
N - number of months.
So,

$$Standardized \quad Tahiti \quad = \quad \frac{TA}{STD_{Tahiti}}$$

Same way, for Darwin station:

$$Standardized \quad Darwin \quad = \quad \frac{DA}{STD_{Darwin}}$$

To calculate the monthly standard deviation:

$$Monthly \quad Standard \quad Deviation \quad (MSD) \quad = \quad \sqrt{\frac{\Sigma_3}{N}}$$

where,
Σ_3 is the sum of ((Standardized Tahiti - Standardized Darwin)^2)
The SOI equation looks as follows:

$$SOI = \frac{Standardized \quad Tahiti - Standardized \quad Darwin}{MSD}$$

The original SST data and SOI data are plotted with R program in Fig. 2 and Fig. 3

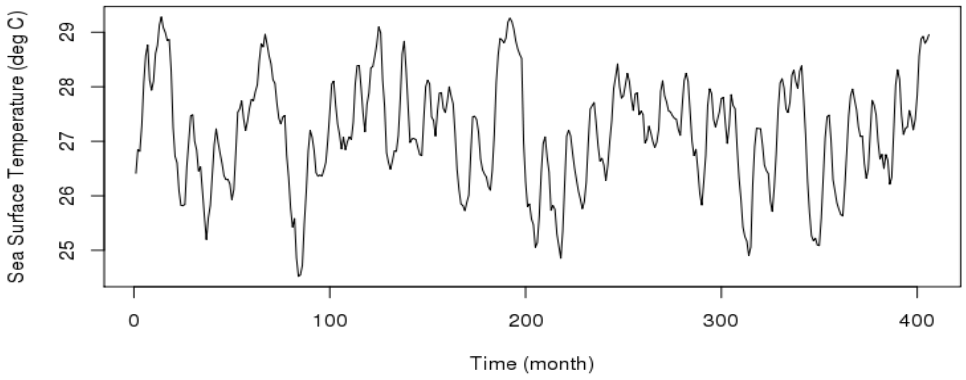

Figure 2: 34-year (December 1981 - September 2015) original monthly Niño 3.4 SST. Indexed from 1 to 406.

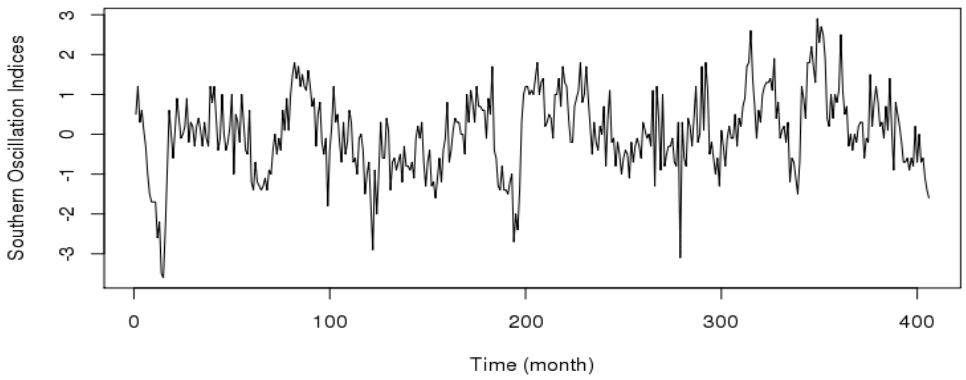

Figure 3: Original SOI datapoints, indexed same way as SSTs.

A negative and lagged relation between these two time series plots can be roughly recognized from the plots. Based on this, this project form following hypotheses: (1) there is a model that can simulate the sea surface temperature data by its past records and make reasonable prediction for the future; and (2) the standardized pressure difference indices (SOI) can act as a contributor in forecasting SSTs.

3 Methodology

Before the datasets can be fitted into any models, the stationarity needs to be ensured. A stationary time series has no trend, a constant variance over time, and constant "wiggliness" over time, which means that its autocorrelations are constant over time. A periodogram of original SSTs is shown in Fig. 4. A pronounced peak is found at the 36th frequency, 0.833. This frenquency converts to a period of 12, which is the correct cycle for monthly data. To test the stationarity of the data series, the auto-correlation function (ACF) of original SST time series is plotted in the left panel of Fig. 5. We can see the ACF of original data series decays slowly and renders an oscilatory pattern over time, which indicates a non-stationarity. The larger values at every 12 lags confirms its cycle as well. To stationarize the data, SSTs are taken first order differencing. The ACF and partial autocorrelation function (PACF) of differenced SSTs are plotted in the middle and right panel of Fig. 5. The ACF of differenced data gives a relatively sharp cut-off at the first and second lag, while the third lag becomes negative, and also oscillates afterwards.

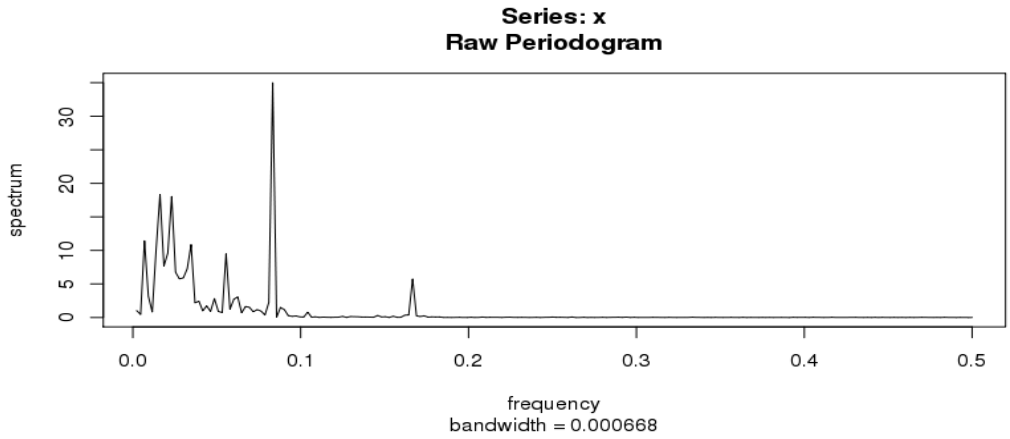

Figure 4: Periodogram of original SST time series. The peak frequency is 0.833, giving a period of 12.

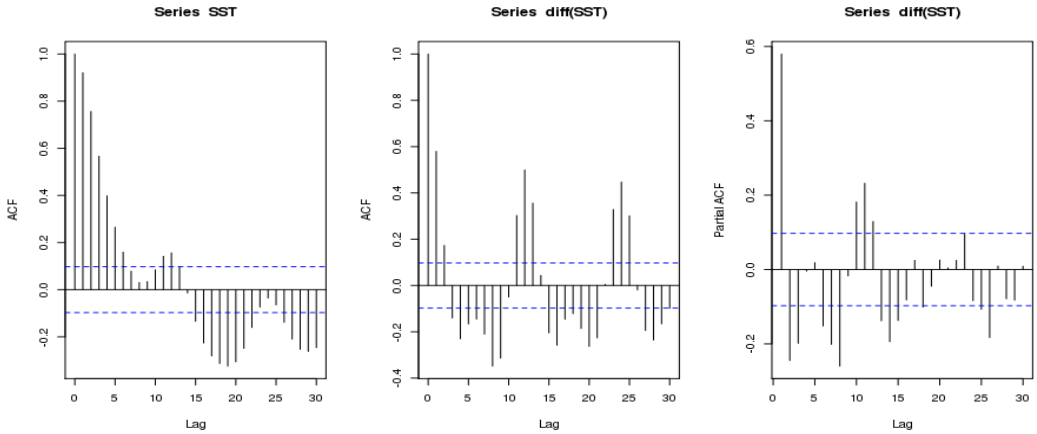

Figure 5: From left to right: Autocorrelation function (ACF) of SSTs, ACF of differenced SSTs, and partial autocorrelation function (PACF) of differenced SSTs.

3.1 AR, MA, and ARMA models

In this section, the differenced SST data is fitted into different autoregressive (AR) model, moving average (MA) model, and autoregressive moving average (ARMA) models. Model fitting AIC and BIC values are tabled in Table. 1. For convenience, all the AIC and BIC values for models that are used in this project are all listed and compared in this table. A linear model is fitted as a reference. The AIC and BIC for the linear model are calculated, both are greater than 1150.

AR(1) and MA(1) are the first two ARMA models fitted, and they give AIC values of 216.3 and 233.22, respectively. The BIC values are even larger. To test for the best model, the AR and MA terms and tried from 1 to 36. ACF, periodogram and Ljung-Box test are examined for the residuals of each fitting trial. All AR and MA models have shown significant seasonal pattern in their ACFs and sharp Ljung-Box p-value drop after a few lags. All AR models have smaller AIC and BIC values than corresponding MA models. A significant lag, however, is found at lag 36, that stands out in almost every experiment. From the model comparison table,

Table 1: Model Comparison

	AR(1)	MA(1)	AR(18)
AIC	216.3	233.22	62.49
BIC	228.31	245.24	142.57
	MA(18)	AR(26)	MA(26)
AIC	97.52	45.8	70.95
BIC	177.6	157.91	183.06
	AR(36)	MA(36)	ARMA(26,2)
AIC	55.14	76.38	47.63
BIC	207.29	228.53	163.74
	ARIMA(26,1,0)	ARIMA(0,1,26)	ARIMA(26,2,0)
AIC	43.89	68.95	82.76
BIC	152	177.05	190.79
	Linear	ARIMA(30,1,0)	SARIMA $(26,1,0)$, $(0,1,1)_{12}$
AIC	1154.56	46.73	30.86
BIC	1166.58	170.85	142.12
	SARIMA $(0,1,26)$, $(0,1,1)_{12}$	ARIMAX(0,1,26)	SARIMAX $(0,1,26)$, $(0,1,1)_{12}$
AIC	19.18	68.57	26.67
BIC	130.45	180.61	141.83

AIC andd BIC Values for models that are fitted in this project.

we can see that AIC values decrease largely before lag 26, and relent a little after this lag, while BIC values drop to minimum around lag 20 and start to increase. In order to find out the best lag to use in the model fitting, a balance needs to be reach between the AIC, BIC, computing efficiency and residual ACFs. Fig. 6 shows the ACF of the residuals after fitting AR(18) and AR(36). These two ACFs give a representative feature of all AR and MA model fitting between lag 1 and 36. Both show insignificant ACFs after lag 1 but with an exception at lag 26. This can ensure us that lag 26 is the best term to be used in our AR/MA models. But is AR(26) or MA(26) our best model? A diagnosis of these two are shown in Fig. 7. Plotted are the ACF of the residuals of first 40 lags, ACF of the residuals of the first 400 lags, and the Ljung-Box test p-values for 400 lags. Upper panels show the results of AR(26) and lower the MA(26). AR model does a slightly better job than MA since both the ACFs and Ljung-Box test results are better over the 400-lag run, but still renders an uncertainty after 60 lags. ARMA models with both AR and MA terms are examined too, but they give worse results than just AR or MA model. An ARMA(26,2) result is shown in the table, which indicates a low efficiency in calculating and unfavorable fitting.

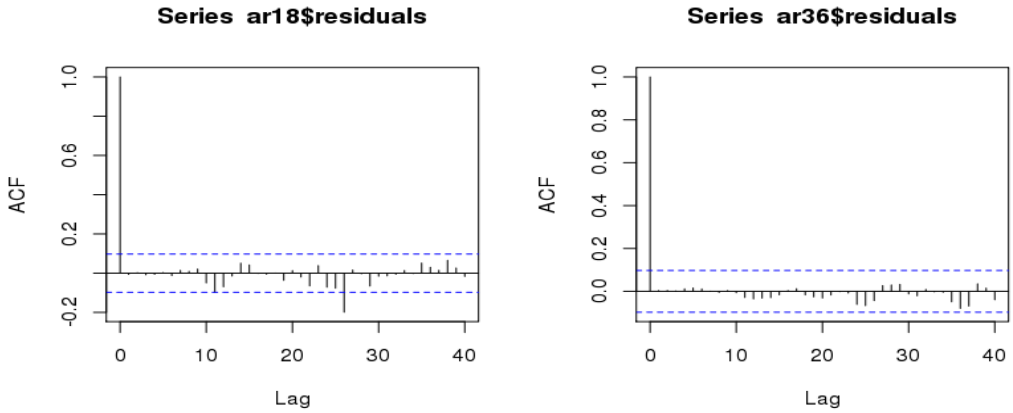

Figure 6: ACF of AR(18) and AR(36). Both are significant at lag 26.

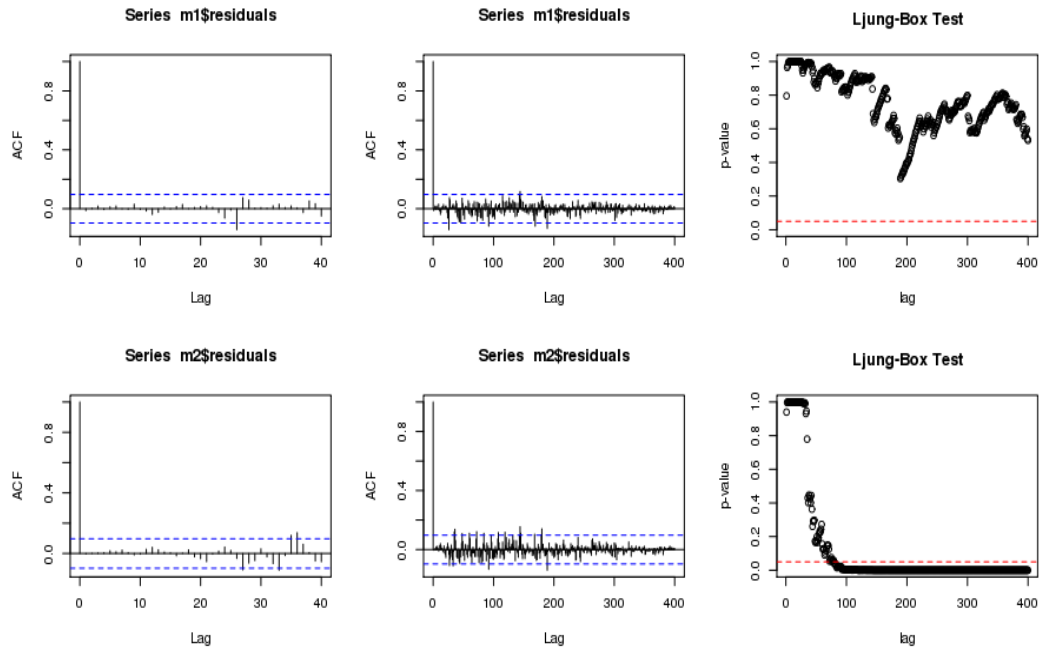

Figure 7: Diagnosis of AR(26) (upper panels) and MA(26) (lower panels) model. From left to right: ACF of first 40 lags, ACF of first 400 lags, and Ljung-Box test for first 400 lags.

3.2 ARIMA and Seasonal ARIMA models

An ARIMA model, or Arima function in R with differencing term, is also examined. Since we have already had the knowledge of best lag to use from previous trials, a lag of 26 is applied. ARIMA model with differencing order of 1 and AR term of 26 and MA term with 26 are run, and the AIC and BIC are given in Table. 1. Both of them have smaller AICs than AR or MA only model by about 12 in value. With the differencing feature provided in the function, a second order differencing is checked up with ARIMA(26,2,0). Compared the AIC value of ARIMA(26,1,0) of 43.80, ARIMA(26,2,0) has an AIC of 82.76. This means first order differencing is sufficient for the data and better than higher order differencing. To assure the best lag is chosen, ARIMA models with AR and MA term at 30 are examined too, but both ARIMA(30,1,0) and ARIMA(0,1,30) show greater AIC and BIC values than at lag 26, confirming a better choice at lag 26. Since ARIMA(26,1,0) gives the best AIC and BIC values so far, a diagnosis is run for it, displayed in Fig. 8. The Ljung-Box tests for the first 40 lags look ok, with very high p-values. The ACF of 400 lags shows a a few significant lags but not too serious; most are acceptable. But Ljung-Box test p-values for 400 lags are problematic as they drop below 40% after 90 lags and can hardly fluctuate above 70%. As for the periodogram, the peak frequency gives a cycle of 9.6, which does not make sense for out monthly data. At this point, a seasonal ARIMA model is introduced for fitting.

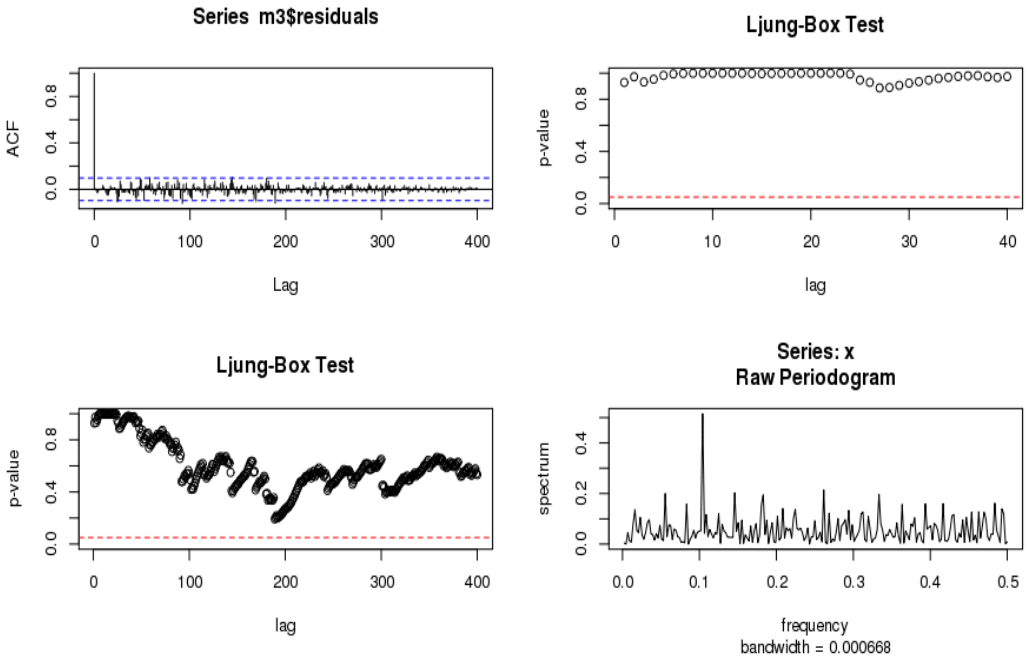

Figure 8: Diagnosis of ARIMA(26,1,0). Ljung-Box test for first 40 lags, ACF and Ljung-Box test p-values for 400 lags, and periodogram.

Seasonal ARIMA models, or SARIMAs, are tested with several popular seasonal terms such as SARIMA(p,0,0)$(0,1,1)_{12}$ and SARIMA(0,1,q)$(0,1,1)_{12}$, where p and q are corresponding AR and MA term. Recalling that for all previous experiments, AR(26) term always works better than MA(26), to my surprise, SARIMA model shows a contrary result. With AR term set to 26 and MA term 0, an AIC of 30.86 is calculated, while AR is set to 0 and MA set to 26 a smaller AIC of 19.18 is reached. For each calculation, seasonal ARMA terms are (0,1,1) and cycle is set to 12 as suggested by the periodogram of original dataset. The diagnosis plots of the better fitted one are shown in Fig. 9. We can see the ACF for both 40 and 400 lags is stable and insignificant after lag 0, and Ljung-Box test p-values for both are close to 1 except at very few lags. The residuals after model fitting look totally white noisy and the periodogram no longer shows any single major peak. All these diagnosis indicates a pretty good fitting by using SARIMA(p,0,0)$(0,1,1)_{12}$ model. A period of 6 for SARIMA model is tested too, but the results are not favorable as expected.

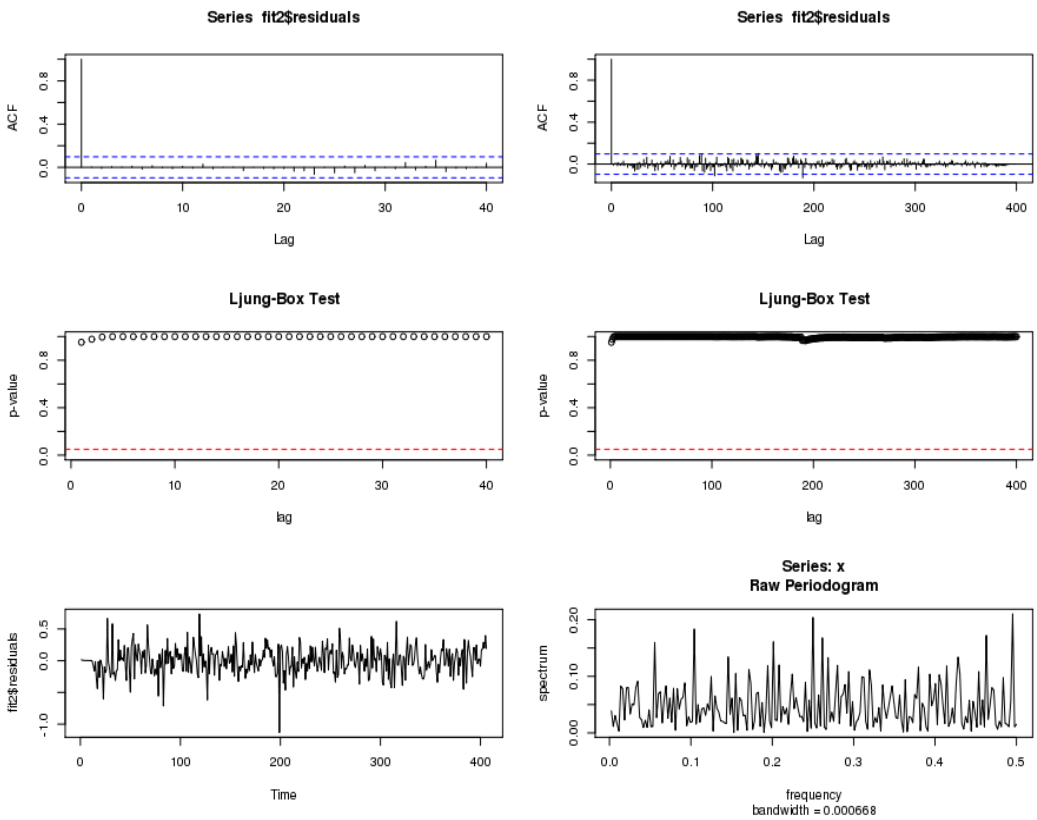

Figure 9: Diagnosis of Seasonal ARIMA(0,1,26),$(0,1,1)_{12}$. ACF and Ljung-Box test for first 40 lags, for 400 lags, and periodogram and residuals are plotted from left to right, top to bottom.

3.3 ARIMAX and Seasonal ARIMAX

It is also of my great interest to examine if the other dataset mentioned in the beginning in the report would be a good predictor for SSTs. The ACF of original and the ACF and PACF of differenced time series dataset SOI are plotted in Fig. 10. Slow decaying of ACF of original dataset is present with sharp decrease in ACF and PACF of differenced data. The ACF at lag 1 and the PACF at lag 0 and lag 1 of differenced data show strong negative value, which might suggest an overdifferenced feature and thus MA(1) term might be considered. The periodogram of original data is given in Fig. 11. No pronounced peak thus cycle/period of SOIs is indicated. This means SOI is not a seasonal data, but a some minor peaks shown can suggest its complicated irregular oscillatory behaviors - which is well-known as the characteristic for ENSO events.

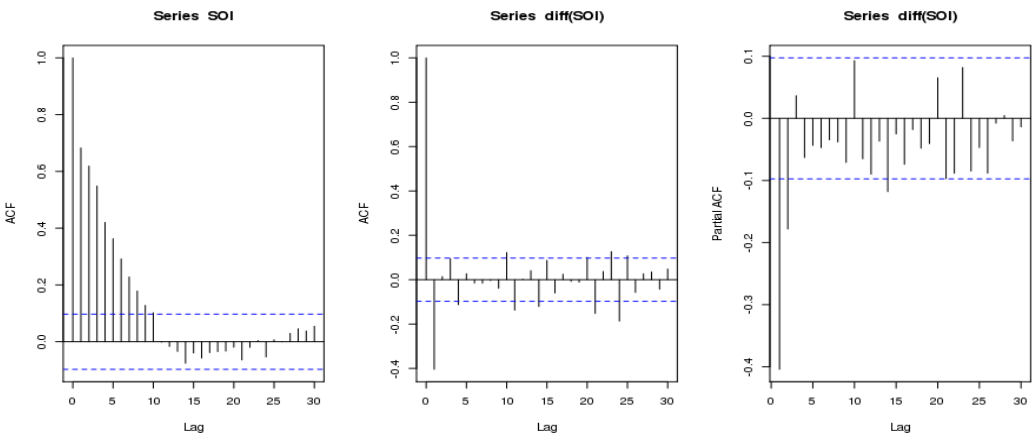

Figure 10: ACF original SOI, differenced SOI and PACF of differenced SOI.

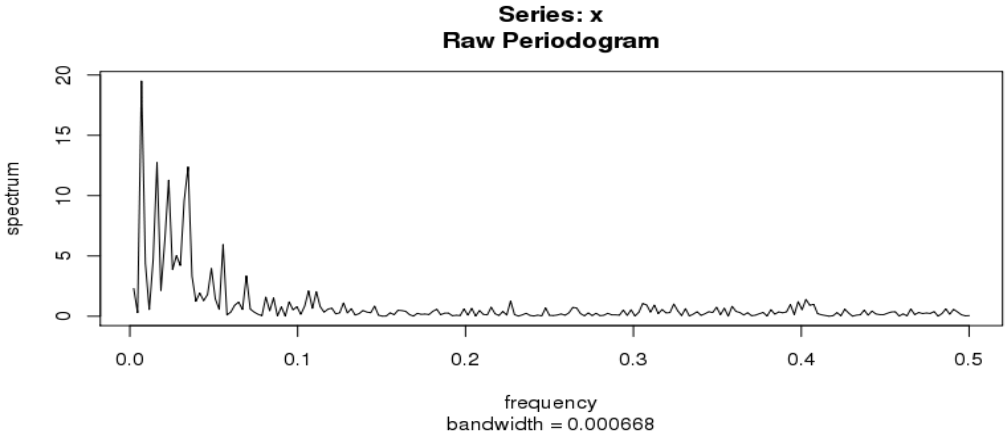

Figure 11: Periodogram of original SOI dataset. No significant cycle is inferred.

Whether or not using SOIs as another time series regressor in our fitting model cannot be told just by the statistical behavior of its own. The cross-correlation function between SSTs and SOIs is examined in Fig. 12 and not so surprisingly, strong negative correlation is present for -10 through +10 lags. The biggest CCF value is at negative lag 1, which suggests that SSTs precedes SOIs. Thus, SOIs are shifted once for use. A Granger-causalty test by using the "vars" package in R is done for 30 lags to check for the significance of shifted SOI data to SSTs. None of the lag test, however, shows that SOI would Granger cause SST. An inverse test is done as well, but the results also show SST does not Granger cause SOI. To fulfill the requirement of the project, an ARIMAX and SARIMAX model is still fitted to see if SOI is capable to act as a predictor for SST datasets. The AIC and BIC of ARIMAX(0,1,26) and SARIMAX(0,1,26)$(0,1,1)_{12}$ are given in the last row of Table. 1. With an AIC value of 68.75, which is only 0.38 smaller than ARIMA model, SOI is not likely to have a great capacity as the external regressor in ARIMAX. As for Seasonal ARIMAX model, the AIC value is even larger than that of SARIMA model by a numeric value of 7.49. The result does show an improvement after introducing SOI as an external regressor, but only as a slight one, rather a huge impact. The

diagnosis of fitting ARIMAX and SARIMAX are shown in Fig. 13 and Fig. 14. The peak in the periodogram of ARIMAX model residuals shows a pronounced cycle of 12, which means ARIMAX even fails to remove the seasonal feature of the data. Also, the Ljung-Box test p-values of ARIMAX residuals drop below significance level after 60 lags, also indicating a failed model fitting. The diagnosis of SARIMAX model is very similar to that of SARIMA's, but the Ljung-Box test at around 200 lags is lower.

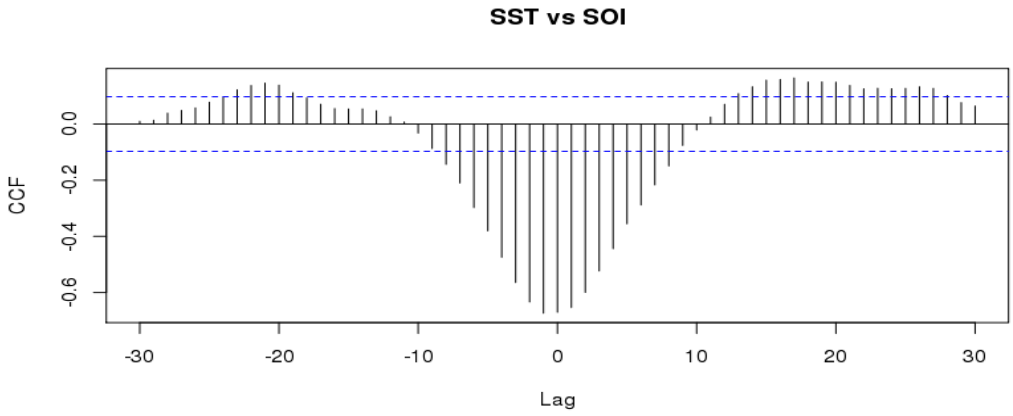

Figure 12: Cross-correlation Function (CCF) for SST and SOI data series at ±30 lags.

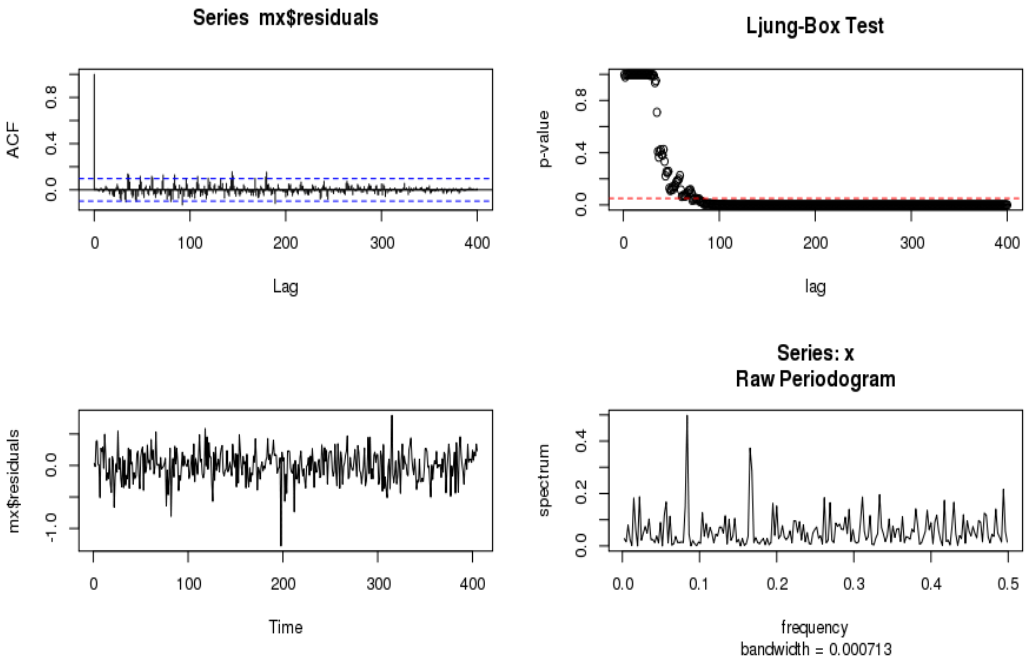

Figure 13: Diagnosis for fitting ARIMAX(0,1,26) model.

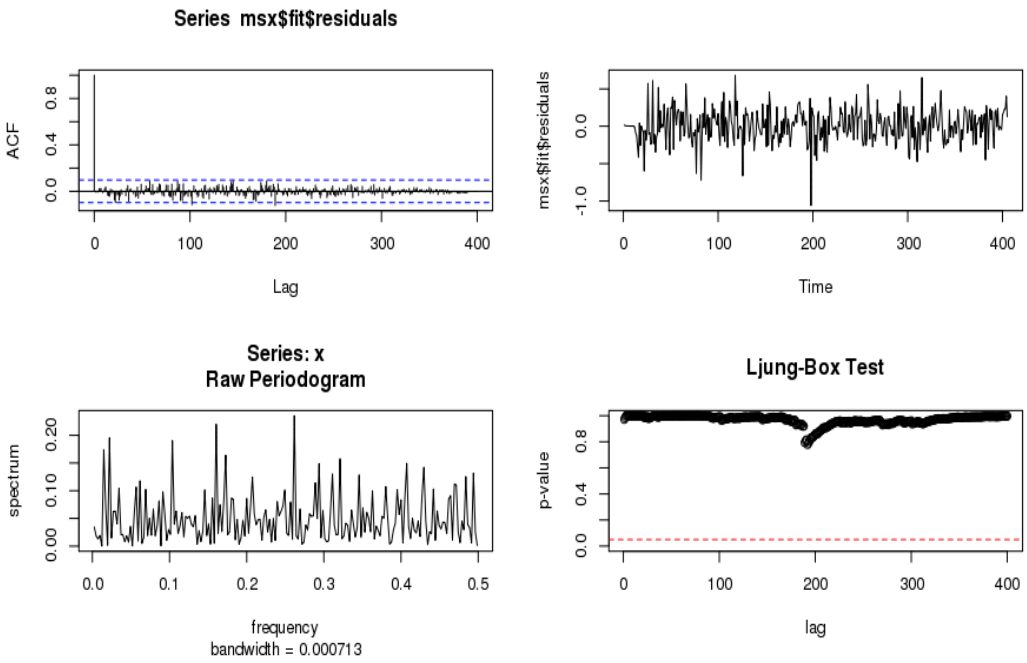

Figure 14: Diagnosis for fitting SARIMAX(0,1,26)$(0,1,1)_{12}$ model.

4 Results and Discussion

With the knowledge from the AICs, BICs and diagnosis of each model, we can conclude that SARIMA(0,1,26)$(0,1,1)_{12}$ is our best models. To have better and more straightforward illustration, the model fitting and the real data for the best model SARIMA and second best model SARIMAX(0,1,26)$(0,1,1)_{12}$ are plotted in Fig. 15: original SST data is plotted in black solid curve with SARIMA fitting in red dashed line and SARIMAX in blue dashed line. Both fitting curves are close to each other and fit the data very well, and both significantly underestimate the SST at the last few datapoints. But it looks to me that SARIMAX has more missing points. The underestimation only occurs at the end of the time series data, and this may imply the unpredictability of the data in the future.

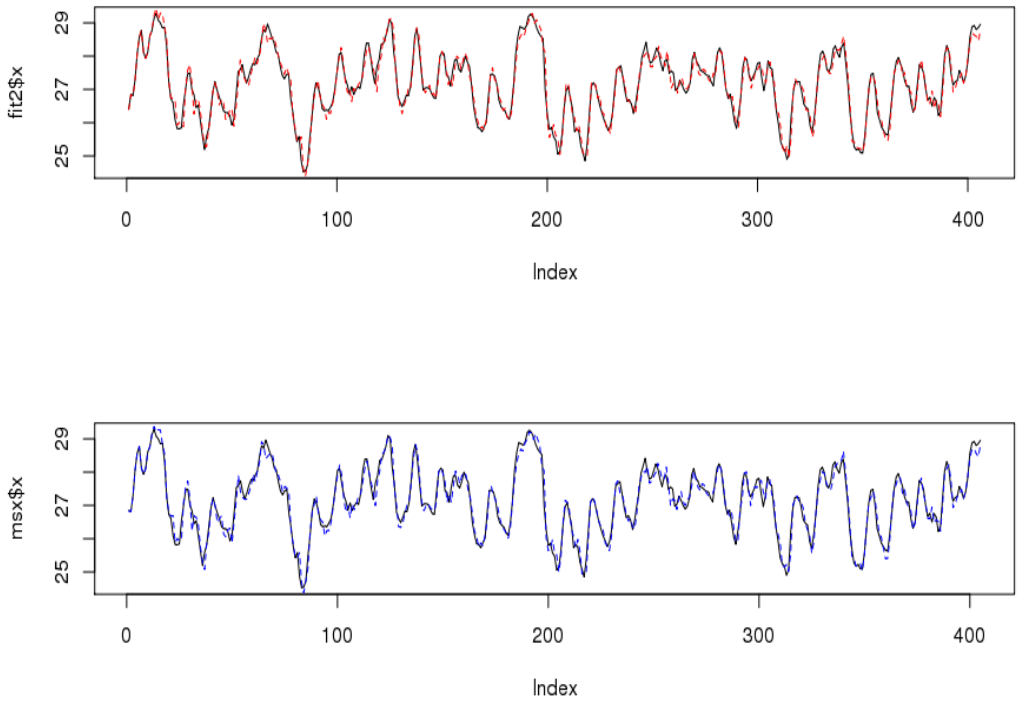

Figure 15: Model fitting of SARIMA (red, top panel) and SARIMAX (blue, bottom panel). Original data is plotted in balck.

Still, we are interested in the forecasting ability of our best model, SARIMA $(0,1,26)(0,1,1)_{12}$, so the the SARIMA model forecasted for 5 month of SST is compared with real data (with the last 5 month plotted in red) in Fig. 16 and Fig. 17. We can see that the SST keeps climbing and reaches a record maximum over the diagram period. The forecasted result is within the 95$ confidence interval. It does predict the first increasing and then decreasing characteristic, even though not precisely. I would not conclude that the model, with such high fitting accuracy, fails to predict the SST, but the on contrary, it confirms the extremely strong El Niño event of last year. That is why this irregular occurrence of ENSO presents a tough challenge for all the climatologists. As for the hypotheses I posed at the beginning of the report, I would say (1) SARIMA model used in this project fits SST data very well and does have the ability to make reasonable forecast, but in extreme occasions like El Niño, it might fail; and (2) SOI is not a significant contributor in SST forecasting and prediction, and even can reduce the precision in fitting by looking at the AICs, due to its own irregularity.

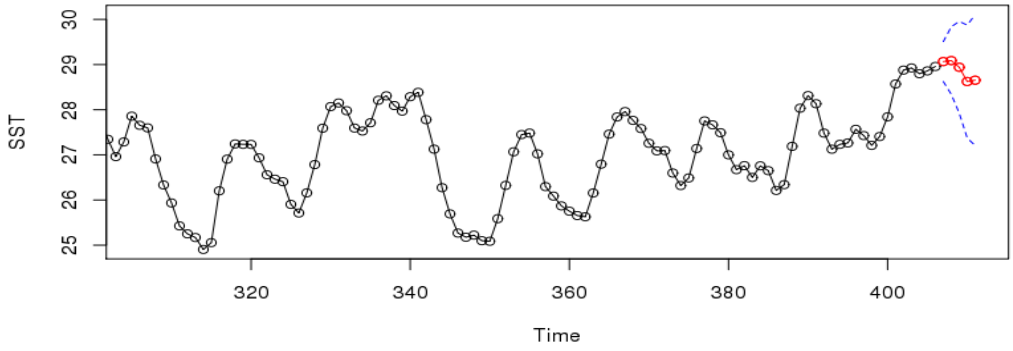

Figure 16: 5 points/months prediction of SARIMA model.

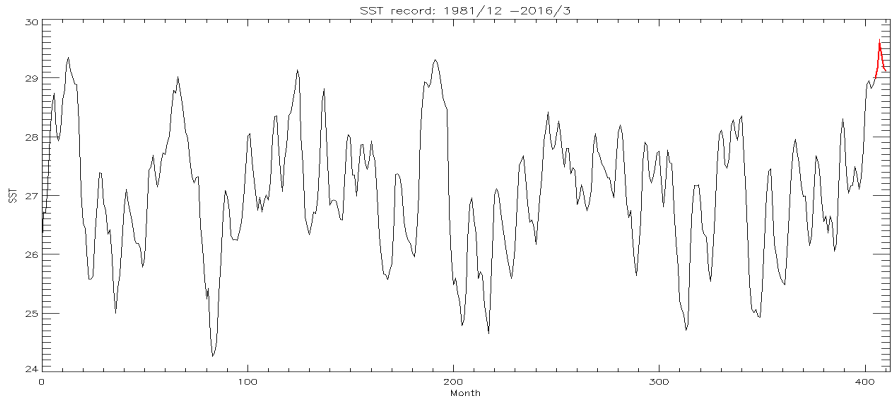

Figure 17: Real SST data, with 10/2015 - 3/2016 plotted in red. This figure is plotted with IDL software.

References

Ambaum, M. H., B. J. Hoskins, and D. B. Stephenson, Arctic oscillation or north atlantic oscillation?, *Journal of Climate*, *14*, 3495–3507, 2001.

Donges, J. F., Y. Zou, N. Marwan, and J. Kurths, The backbone of the climate network, *EPL (Europhysics Letters)*, *87*, 48,007, 2009.

Feliks, Y., M. Ghil, and A. W. Robertson, Oscillatory climate modes in the eastern mediterranean and their synchronization with the north atlantic oscillation, *Journal of Climate*, *23*, 4060–4079, 2010.

Ghil, M., and K. Mo, Intraseasonal oscillations in the global atmosphere. part i: Northern hemisphere and tropics, *Journal of the Atmospheric Sciences*, *48*, 752–779, 1991.

Kimoto, M., and M. Ghil, Multiple flow regimes in the northern hemisphere winter. part ii: Sectorial regimes and preferred transitions, *Journal of the atmospheric sciences*, *50*, 2645–2673, 1993.

Mooley, D., and A. Munot, Variation in the relationship of the indian summer monsoon with global factors, *Proceedings of the Indian Academy of Sciences-Earth and Planetary Sciences*, *102*, 89–104, 1993.

Tsonis, A. A., and K. L. Swanson, Topology and predictability of el nino and la nina networks, *Physical Review Letters*, *100*, 228,502, 2008.

Tsonis, A. A., K. Swanson, and S. Kravtsov, A new dynamical mechanism for major climate shifts, *Geophysical Research Letters*, *34*, 2007.

Viron, O., J. Dickey, and M. Ghil, Global modes of climate variability, *Geophysical Research Letters*, *40*, 1832–1837, 2013.

Wang, G., P. Yang, X. Zhou, K. L. Swanson, and A. A. Tsonis, Directional influences on global temperature prediction, *Geophysical Research Letters*, *39*, 2012.

Wyatt, M. G., S. Kravtsov, and A. A. Tsonis, Atlantic multidecadal oscillation and northern hemisphere's climate variability, *Climate dynamics*, *38*, 929–949, 2012.

Codes used in these project

Re-run with caution: all figures can be re-produced, but some lines need to be commented out or un-commented. Properties of models may need manually typing to show.

```
D <- read.csv("/N/u/xuecliu/Karst/Downloads/rpro/sst_nino34.csv", header=T)
SST=D$SST
SOI=D$SOI
summary(D)

plot(D$SST,xlab="Time (month)",ylab="Sea Surface Temperature (deg C)",type="l")
plot(D$SOI,xlab="Time (month)",ylab="Southern Oscillation Indices",type="l")

#Spectral analysis

install.packages("FitAR")
library(FitAR)

ts1=spectrum(SST,log="no")
1/ts1$freq[which.max(ts1$spec)] #the 36th freq=0.833, period=12

ts2=spectrum(SOI,log="no")
1/ts2$freq[which.max(ts2$spec)] #period = 144

par(mfrow=c(1,3))
acf(SST,lag.max = 30)
acf(diff(SST),lag.max = 30)
#acf((diff(SST)))
pacf(diff(SST),lag.max = 30)
par(mfrow=c(1,3))
acf(SOI,lag.max = 30)
acf(diff(SOI),lag.max = 30)
pacf(diff(SOI),lag.max = 30)

#AR & MA
m1=Arima(diff(SST), order=c(1,0,0))
m1=Arima(diff(SST), order=c(0,0,1))
m1=Arima(diff(SST), order=c(0,0,18))
ar18=Arima(diff(SST), order=c(18,0,0))
m1 = Arima(diff(SST), order=c(26,0,0) )
m2 = Arima(diff(SST), order=c(0,0,26) )
```

```
ar36 = Arima(diff(SST), order=c(36,0,0) )
m1 = Arima(diff(SST), order=c(0,0,36) )
m1 = Arima(diff(SST), order=c(26,0,2) )
dev.off()
par(mfrow=c(1,2))
  acf(ar18$residuals,lag.max = 40)
  acf(ar36$residuals,lag.max = 40)

dev.off()
par(mfrow=c(2,3))
  #plot(m1$residuals)
  acf(m1$residuals,lag.max = 40)
  acf(m1$residuals,lag.max = 400)
  LBQPlot(m1$residuals,lag.max=400,StartLag=1)
  acf(m2$residuals,lag.max = 40)
  acf(m2$residuals,lag.max = 400)
  LBQPlot(m2$residuals,lag.max=400,StartLag=1)
  P1=spectrum(m1$residuals,log="no")
  1/P1$freq[which.max(P1$spec)]

dev.off()
lmfit = lm(SST~c(1:406))
plot(lmfit)
plot(lmfit$residuals)
AIC(lmfit)
BIC(lmfit)

#ARIMA
m3 = Arima(SST, order=c(26,1,0) )
m3 = Arima(SST, order=c(0,1,26) )
m3 = Arima(SST, order=c(0,2,26) )
m3 = Arima(SST, order=c(26,2,0) )


par(mfrow=c(2,2))
  acf(m3$residuals,lag.max = 400)
  LBQPlot(m3$residuals,lag.max=40,StartLag=1)
  LBQPlot(m3$residuals,lag.max=400,StartLag=1)
  P3=spectrum(m3$residuals,log="no")
  1/P3$freq[which.max(P1$spec)]
  plot(m3$residuals)

  par(mfrow=c(1,1))
```

```
  pacf(m3$residuals)
  plot(m3$residuals)

Box.test(m3$residuals,lag=20)

par(mfrow=c(1,1))
fit <- Arima(SST, order=c(26,0,26) )
plot(fit$x,col="black",type="l") #original
lines(fitted(fit),col="red") #fitted

dev.off()
#SARIMA
library(astsa)
ms1 = sarima(SST, 0,1,26, 0,1,1, 6, details=FALSE)
ms1 = sarima(SST, 26,1,0, 0,1,1, 12, details=FALSE)
fit1 <-Arima(SST, order=c(26,1,0), seasonal=list(order=c(0,1,1),period=12))
par(mfrow=c(2,2))
  plot(ms1$fit$residuals,type="l")
  acf(ms1$fit$residuals,lag.max = 400)
  P2=spectrum(ms1$fit$residuals,log="no")
    1/P2$freq[which.max(P1$spec)] # =6
  LBQPlot(ms1$fit$residuals,lag.max=400,StartLag=1)

ms2 = sarima(SST, 0,1,26, 0,1,1, 12, details=FALSE)
par(mfrow=c(2,2))
  plot(ms2$fit$residuals,type="l")
  acf(ms2$fit$residuals,lag.max = 400)
  P2=spectrum(ms2$fit$residuals,log="no")
    1/P2$freq[which.max(P2$spec)]
  LBQPlot(ms2$fit$residuals,lag.max=400,StartLag=1)

fit2 <-Arima(SST, order=c(0,1,26), seasonal=list(order=c(0,1,1),period=12))
par(mfrow=c(3,2))
  acf(fit2$residuals,lag.max = 40)
  acf(fit2$residuals,lag.max = 400)
  LBQPlot(fit2$residuals,lag.max=40,StartLag=1)
  LBQPlot(fit2$residuals,lag.max=400,StartLag=1)
  plot(fit2$residuals)
  P2=spectrum(fit2$residuals,log="no")
    1/P2$freq[which.max(P2$spec)]
  plot(fit2$x,col="black",type="l") #original
  lines(fitted(fit2),col="red") #fitted
```

```
dev.off()
sarima.for(SST,5,0,1,26, 0,1,1,12)#sarima forecast 5 points

dev.off()
#Two Time Series analysis
ccf(SST,SOI,30, main="SST vs SOI", ylab="CCF")
B = c(0,1)
soi_1=filter(SOI,B,sides=1) #shift 1 lag back based on CCF
soi_1=soi_1[2:406]
sst_1=SST[2:406]

#Granger test
library(vars)
x = cbind(sst_1,soi_1)
var <- VAR(x, p = 30, type = "const")
causality(var, cause = "soi_1")$Granger

# AUTOMATICALLY SEARCH FOR THE MOST SIGNIFICANT RESULT
for (i in 1:30)
{
  print(causality(VAR(x, p = i, type = "const"), cause = "soi_1")$Granger)
  print(AIC(VAR(x, p = i, type = "const")))
}

# the other direction:
for (i in 1:30)
{
  print(causality(VAR(x, p = i, type = "const"), cause = "sst_1")$Granger)
}

#ARIMAX
mx = Arima(sst_1, order=c(0,1,26), xreg=soi_1 )
mx
par(mfrow=c(2,2))
  acf(mx$residuals,lag.max = 400)
  LBQPlot(mx$residuals,lag.max=400,StartLag=1)
  plot(mx$residuals)
  px = spectrum(mx$residuals,log="no")
  #plot(mx$x,col="black",type="l") #original
  #lines(fitted(mx),col="red") #fitted
  1/px$freq[which.max(px$spec)]
```

```
dev.off()
#SARIMAX
msx = sarima(sst_1, 0,1,26, 0,1,1, 12, xreg=soi_1, details=FALSE )
msx = Arima(sst_1, order=c(0,1,26), seasonal=list(order=c(0,1,1),period=12), xreg=soi_1)
msx
par(mfrow=c(2,2))
  acf(msx$residuals,lag.max = 400)
  plot(msx$residuals)
  psx = spectrum(msx$residuals,log="no")
  LBQPlot(msx$residuals,lag.max=400,StartLag=1)
  #hsx = spectrum(msx$fit$residuals,spans=c(20,20))

par(mfrow=c(2,1))
  plot(fit2$x,col="black",type="l") #original
  lines(fitted(fit2),col="red",lty=2) #fitted
  plot(msx$x,col="black",type="l") #original
  lines(fitted(msx),col="blue",lty=2) #fitted


install.packages("forecast")
library(forecast)


#VAR model
install.packages("vars")
library(vars)
x = cbind(SST,SOI)
summary(VAR(x, p=1))
# p = 1 indicates lag = 1
summary(fit2<-VAR(x, p=2, type="both"))
#  type = c("const", "trend", "both", "none")
acf(resid(fi2t), 52)
prediction = predict(fit, n.ahead = 24, ci = 0.95)
fanchart(prediction)
\end{document}
```