Cross Correlation Analysis between Sea Surface Temperature and Global Air Temperature

Minoru Tanaka

Department of Network and Information, School of Network and Information, Senshu University, Kawasaki 214-8580, Japan

Abstract. The purpose of this paper is discussing the estimation of a sample cross correlation function among the time series, global and regional annual average *sea surface temperature (SST) deviations of the Indian Ocean, the North Pacific, the South Pacific, the North Atlantic and the South Atlantic*. In order to show if a cross correlation estimate is significantly different from zero, the prewhitening one of the time series is very important. We use SARIMA model fitting for the prewhitening and estimate the cross correlation functions. Furthermore, we consider the lagged linear regression model with ARIMA errors by using the series of SST as a leading indicator of the global air temperature.

Keywords: cross correlation analysis, prewhitening, SARIMA model, sea surface temperature, global air temperature

1. Introduction

In the previous paper (Tanaka [7]), we have discussed the cross correlation analysis for the estimation of a sample cross correlation function (CCF) between two series. In this paper we shall discuss the estimation of CCF between the global air temperature and the sea surface temperature (SST) as a continuation of the paper [7] .

2. Cross Correlation Analysis and Prewhitening

2.1 Global Air Temperature and Global Sea Surface Temperature

Both the world annual average air-temperature and the world annual average oceanic-surface-temperature deviations are going up, and it turns out in the past 100 years that +0.55 degrees for the air temperature, and +0.3 for the sea surface temperature are rising (both series, 1891-2019, were obtained from Japan Meteorological Agency [5], and see Hansen [3]). Each average become the maximum in 2016 shown in Figure 2.1.1(a). Although the strong correlation is seen at the lag 0 from the graph of the sample cross correlation function of the two original series in Figure 2.1.1(b), the true correlation coefficient may not be obtained under the influence of a trend component.

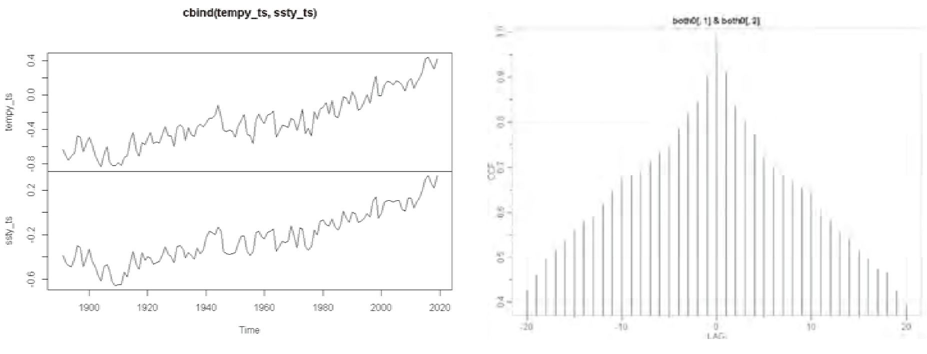

Figure 2.1.1. (a) global air temperature deviations (top left-side) and global sea surface temperature deviations (bottom), (b) the sample CCF between the two series (right-side)

Then we should need prewhitening the two series to estimate the true cross-correlation (Shumway [6] and Tanaka [7]), and we employ an ARIMA model fitting to each series for the prewhitening. Since each fitted model is ARIMA (1,1,3) and p-value in Ljung-Box test of each residuals is 0.7577 and 0.7047 below, we can say that the each residuals of the model will be white noise (see for example, Blockwell and Davis [1]). Figure 2.1.2 shows the two residuals of the ARIMA(1,1,3) models and the sample cross-correlation function CCF between the two series. There are significant correlations at lag 0 (+0.8), at lag 1 (+0.25) and at lag 18 (+0.24) which indicates that the SST might lead the global air temperature by one year and 18 years.

```
Series: tempy_ts
ARIMA(1,1,3) with drift
Coefficients:
          r1      ma1      ma2      ma3    drift
     -0.9285  0.5923  -0.6401  -0.3922  0.0081
s.e.  0.0829  0.1094   0.0927   0.0782  0.0025
        Ljung-Box test
data:  Residuals from ARIMA(1,1,3) with drift
Q* = 2.624, df = 5, p-value = 0.7577

Series: ssty_ts
ARIMA(1,1,3) with drift
Coefficients:
          r1      ma1      ma2      ma3    drift
     -0.9413  0.8695  -0.6248  -0.5827  0.0056
s.e.  0.0434  0.0938   0.1008   0.0914   0.0021
sigma^2 estimated as 0.004741:  log likelihood=162.85
AIC=-313.69   AICc=-313   BIC=-296.58
        Ljung-Box test
data:  Residuals from ARIMA(1,1,3) with drift
Q* = 2.9693, df = 5, p-value = 0.7047
```

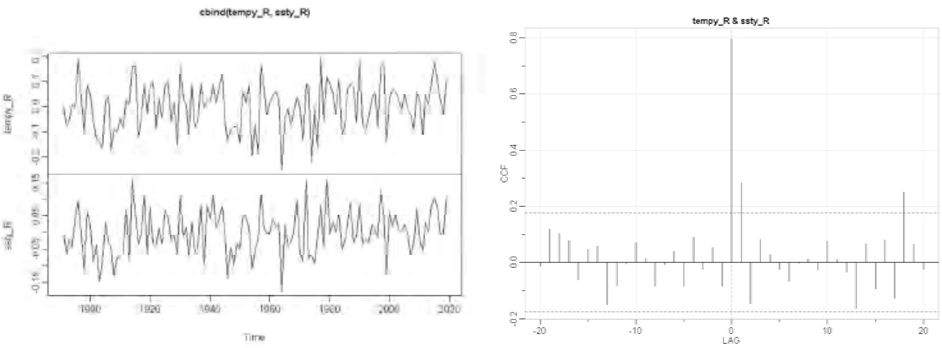

Figure 2.1.2. (a) Prewhitening series of the tempy_ts and of the ssty_ts (left-side)   (b) The sample CCF between the two series (right-side)

Next we consider the following lagged linear regression model with ARIMA errors by using the series of SST as a leading indicator of the global air temperature.

$$T_t = \beta_0 + \beta_1 S_t + \beta_2 S_{t-1} + \beta_3 S_{t-18} + X_t, \qquad (2.1)$$

where T_t is the global air temperature and S_t is the global SST in year t, and X_t is assumed white noise with mean zero and standard error σ. Using the R package dynlm, the linear model was fitted by ordinary least squares (see [2] and [6]). The following output for the model (2.1) shows that the parameter coefficients $\beta_0 = 0$ and $\beta_3 = 0$ are accepted in significance level 0.1.

```
Call:
dynlm(formula = tempy_ts ~ ssty_ts + L(ssty_ts, 1) + L(ssty_ts, 18))
Residuals:
       Min        1Q         Median      3Q         Max
  -0.118773 -0.032515  0.008488  0.031758  0.127839
Coefficients:
                   Estimate    Std. Error   t value   Pr(>|t|)
(Intercept)       0.004540   0.009121   0.498     0.6197
ssty_ts           1.173657   0.061727  19.014    <2e-16 ***
L(ssty_ts, 1)     0.102745   0.057580   1.784     0.0772 .
L(ssty_ts, 18)    0.011036   0.043472   0.254     0.8001
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.04507 on 107 degrees of freedom
Multiple R-squared:  0.9758,  Adjusted R-squared:  0.9751
F-statistic:  1436 on 3 and 107 DF,  p-value: < 2.2e-16
```

Then we can get the fitted model (2.2) using generalized least squares procedure.

$$\hat{T}_t = 1.18\, S_t + 0.11\, S_{t-1} + \hat{X}_t\,, \tag{2.2}$$

where $\hat{X}_t$ is an AR(2) process with $\hat{\sigma} = 0.04$ and the Adjusted R-squared is 0.986. Also from the Ljung-Box test of the residuals from AR(2) model, the p-value = 0.287 suggests that the residuals may be white noise. Therefore the lagged variable S_{t-1} will be a sufficient variable to predict the global air temperature series.

```
Call:
arima(x = resid(reg00), order = c(2, 0, 0), include.mean = FALSE)
Coefficients:
         ar1      ar2
       0.3182  0.2493
s.e.  0.0857  0.0878
sigma^2 estimated as 0.00198:  log likelihood = 216.59,  aic = -427.19

        Ljung-Box test
data:  Residuals from ARIMA(2,0,0) with zero mean
Q* = 9.7028, df = 8, p-value = 0.2865
Model df: 2.   Total lags used: 10
```

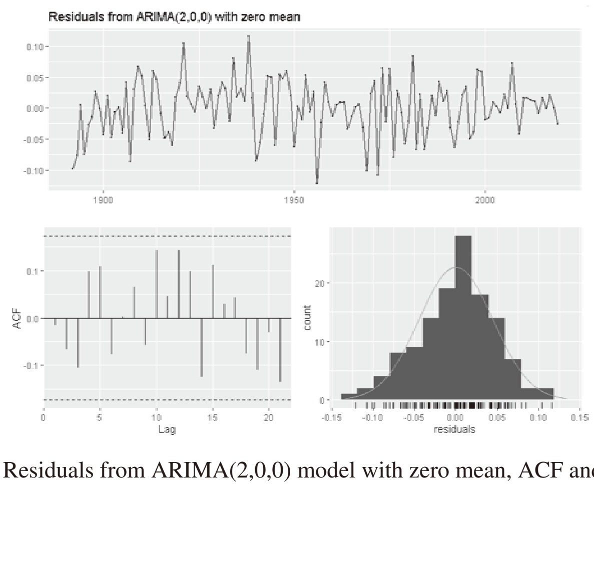

Figure 2.1.3. Residuals from ARIMA(2,0,0) model with zero mean, ACF and Histogram of the series.

2.2 Global Air Temperature and Five Sea Surface Temperatures

We here consider the regional annual average *sea surface temperature (SST) deviations of the Indian Ocean, the North Pacific, the South Pacific, the North Atlantic and the South Atlantic (1891-2019)(Figure 2.2.0)* which are obtained from Japan Meteorological Agency [5]. In order to show if a cross correlation estimate is significantly different from zero, the prewhitening one of the time series is very important. We also employ SARIMA model fitting for the prewhitening and estimate the cross correlation functions (see Shumway [6] and Tanaka [7]). Furthermore, we consider a lagged linear regression model with ARIMA errors by using the series of SST as a leading indicator of the global air temperature.

[1] Global air temperature and SST of South Pacific Ocean:

We consider the estimation of the cross correlation function between the global air temperature (G-Temp) and the SST of the South Pacific ocean (see Figure 2.2.1). Prewhitening the two series by fitting ARIMA model, we can estimate the CCF between the two series.

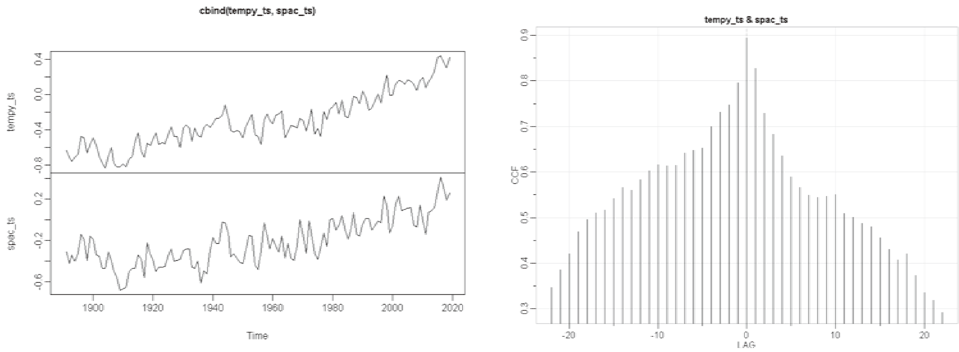

Figure 2.2.1. (a) G-Temp (top) and SST of South Pacific (bottom) (b) the sample CCF between the two series.

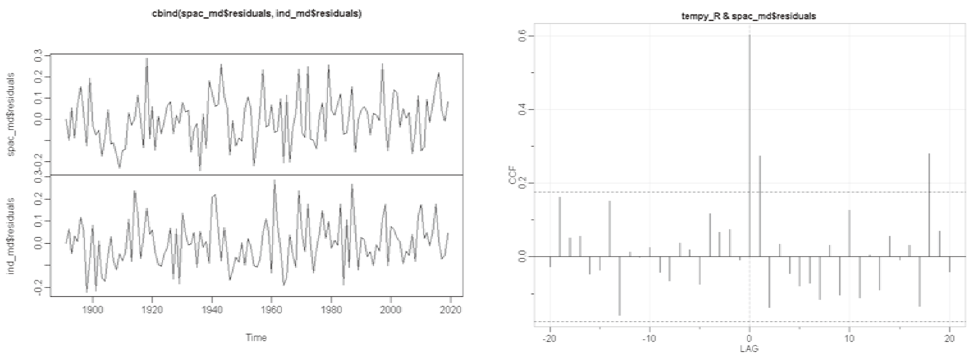

Figure 2.2.2. (a) Prewhitening series of the G-Temp and of the SST (left-side) (b) The sample CCF between the two series (right-side)

Figure 2.2.2 shows that the sample CCF peaked at lag $h = 0$ ($\hat{\rho}_{xy}(0) = 0.8$), at lag $h = 1$ ($\hat{\rho}_{xy}(1) = 0.24$) and at lag $h = 18$ ($\hat{\rho}_{xy}(18) = 0.25$). These results indicate that the SST might lead the air temperature by one year and 18 years.

[2] Global air temperature and SST of North Pacific Ocean:

Using the similar way of [1], we show the result of the sample CCF between the global air temperature and the

North Pacific SST.

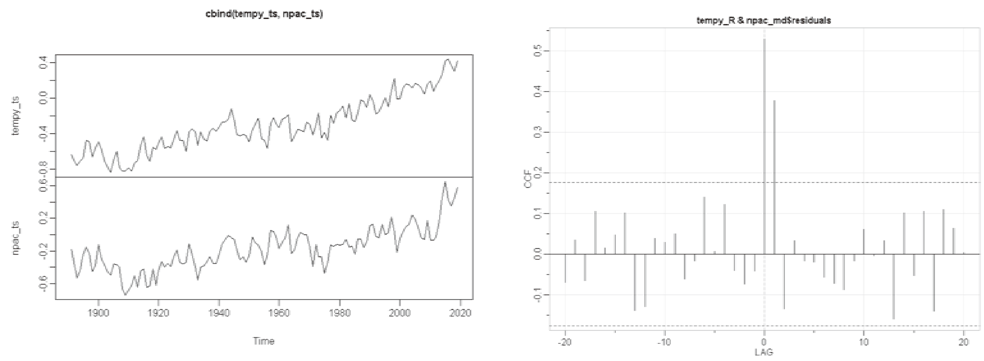

Figure 2.2.3. (a) G-Temp (top) and SST of North Pacific (bottom) (b) The sample CCF between the two prewhitening series.

The sample CCF peaked at lag $h = 0$ ($\hat{\rho}_{xy}(0) = 0.53$), at lag $h = 1$ ($\hat{\rho}_{xy}(1) = 0.38$), and these indicate that the SST might lead the air temperature by one year.

[3] Global air temperature and SST of South Atlantic Ocean:

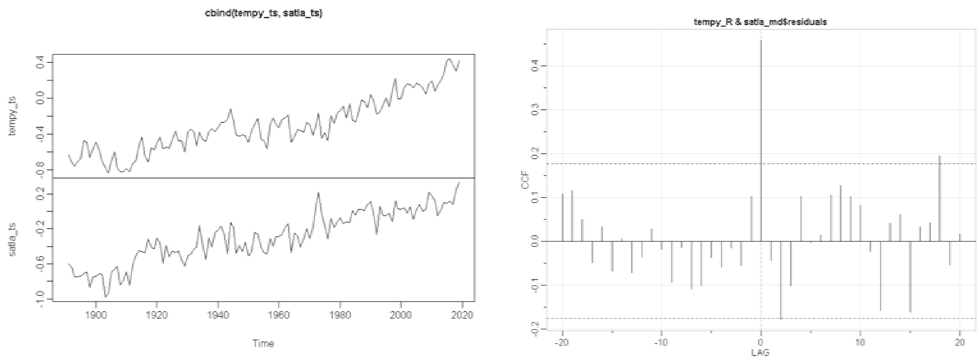

Figure 2.2.4. (a) G-Temp (top) and SST of South Atlantic (bottom) (b) The sample CCF between the two prewhitening series.

The sample CCF peaked at lag $h = 0$ ($\hat{\rho}_{xy}(0) = 0.45$), and this shows that the SST of South Atlantic Ocean measured at time t (years) is associated with the temperature deviations at same time *t*.

[4] Global air temperature and SST of North Atlantic Ocean:

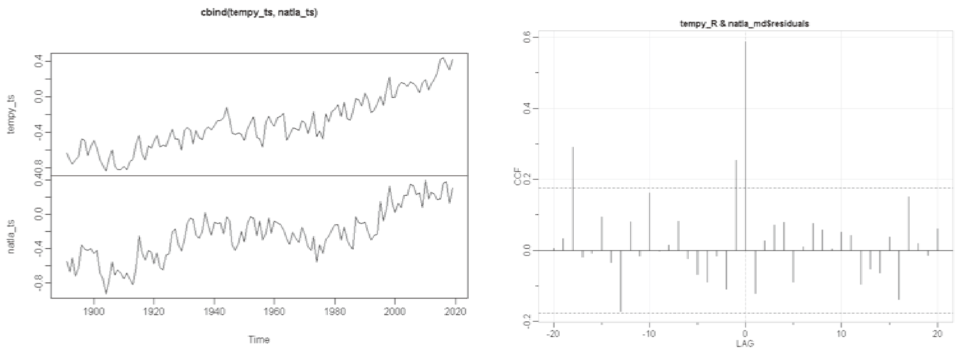

Figure 2.2.5. (a) G-Temp (top) and SST of North Atlantic (bottom) (b) The sample CCF between the two prewhitening series.

The sample CCF peaked at lag $h = 0$ ($\hat{\rho}_{xy}(0) = 0.59$), and this shows that the SST of North Atlantic Ocean measured at time t (years) is associated with the temperature deviations at same time t.

[5] Global air temperature and SST of Indian Ocean:

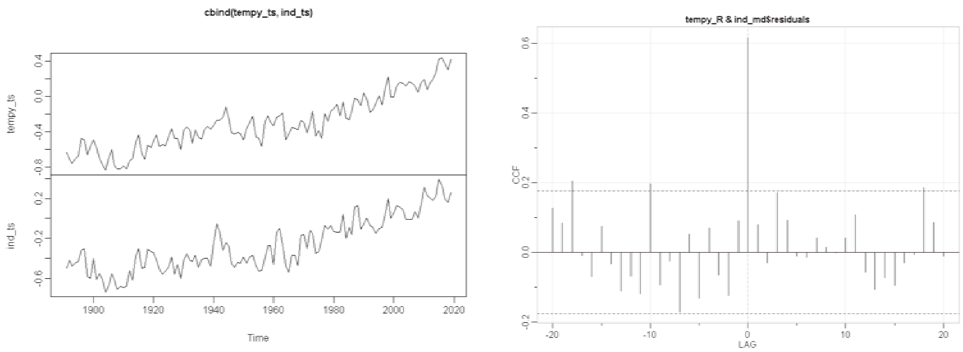

Figure 2.2.6. (a) G-Temp (top) and SST of Indian Ocean (bottom) (b) the sample CCF between the two prewhitening series.

The sample CCF peaked at lag $h = 0$ ($\hat{\rho}_{xy}(0) = 0.61$), and the SST of Indian Ocean measured at time t (years) is associated with the air temperature deviations at same time t.

In order to check the effect of each SST on the air temperature, we consider a lagged linear regression model with ARIMA errors by using the SST as a leading indicator of the global air temperature:

$$T_t = \beta_0 + \beta_1\, S1_t + \beta_2\, S2_t + \beta_3\, S2_{t-1} + \beta_4\, S2_{t-18} + \beta_5\, S3_t + \beta_6\, S3_{t-1} + \beta_7\, S4_t + \beta_8\, S5_t + X_t, \qquad (2.3)$$

where the variable T_t is the global air temperature, $S1_t$ is the SST of Indian Ocean (ind_ts), $S2_t$ is the SST of South Pacific (spac_ts), $S3_t$ is the SST of North Pacific (npac_ts), $S4_t$ is the SST of South Atlantic (satla_ts), $S5_t$ is the SST of North Atlantic (natla_ts) and X_t is the ARIMA(0,0,3) errors with mean zero and standard error σ. At first, assuming the X_t sequence is white noise, and using the R package dynlm, we will get the following output :

```
Call:
dynlm(formula = tempy_ts ~ ind_ts + spac_ts + L(spac_ts, 1) +
        L(spac_ts, 18) + npac_ts + L(npac_ts, 1) + satla_ts + natla_ts)
```

```
Residuals:
      Min         1Q        Median       3Q        Max
-0.115772 -0.029948  0.008407  0.029849  0.096885
Coefficients:
                    Estimate    Std. Error    t value    Pr(>|t|)
 (Intercept)        -0.002734   0.008171   -0.335    0.73864
 ind_ts              0.275675   0.052930    5.208    9.95e-07 ***
 spac_ts             0.141806   0.049337    2.874    0.00493 **
 L(spac_ts, 1)       0.051722   0.051207    1.010    0.31486
 L(spac_ts, 18)      0.066527   0.035253    1.887    0.06199 .
 npac_ts             0.249737   0.042204    5.917    4.41e-08 ***
 L(npac_ts, 1)       0.030254   0.049741    0.608    0.54439
 satla_ts            0.167467   0.038268    4.376    2.93e-05 ***
 natla_ts            0.264976   0.029435    9.002    1.31e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.0454 on 102 degrees of freedom
Multiple R-squared:  0.9766,         Adjusted R-squared:  0.9747
F-statistic: 531.1 on 8 and 102 DF,  p-value: < 2.2e-16
```

It is seen that the coefficients of β_0 and β_6 may be equal zero in significant level 0.1 and the variables $S3_{t-1}$ should be deleted on the model (2.3). We have refitted the model using GLS. Then we can get the following best fitted model (2.4) in terms of the maximum Adjusted R-squared value (0.986).

$$\hat{T}_t = 0.26\ S1_t + 0.14\ S2_t + 0.07\ S2_{t-1} + 0.07\ S2_{t-18} + 0.26\ S3_t + 0.17\ S4_t + 0.27\ S5_t + \hat{X}_t, \qquad (2.4)$$

where $\hat{X}$ is ARIMA(0,0,3) = MA(3) model. The output of the Ljung-Box test and Figure 2.2.7 below suggest that the residuals from ARIMA(0,0,3) are white noise with $\sigma = 0.05$ since the p-value is 0.268.

```
Call:
 dynlm(formula = tempy_ts ~ ind_ts + spac_ts + L(spac_ts, 1) +
                                        L(spac_ts, 18) + npac_ts + satla_ts + natla_ts)
Residuals:
     Min          1Q            Median       3Q        Max
-0.112259  -0.030438    0.004423   0.030776  0.100677
Coefficients:
                    Estimate    Std. Error  t value    Pr(>|t|)
 ind_ts             0.26293    0.04857    5.413     4.00e-07 ***
 spac_ts            0.14409    0.04844    2.975     0.00365 **
 L(spac_ts, 1)      0.07319    0.03885    1.884     0.06239 .
 L(spac_ts, 18)     0.07482    0.02608    2.869     0.00499 **
 npac_ts            0.26330    0.03283    8.019     1.67e-12 ***
 satla_ts           0.16591    0.03724    4.455     2.12e-05 ***
 natla_ts           0.27134    0.02723    9.966     < 2e-16 ***
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.04506 on 104 degrees of freedom
Multiple R-squared:  0.9865, Adjusted R-squared:  0.9855
F-statistic:  1082 on 7 and 104 DF,  p-value: < 2.2e-16
```

```
[2]  residuals
 arima(x = resid(reg2), order = c(0, 0, 3), include.mean = FALSE)
 Coefficients:
           ma1      ma2       ma3
       0.3625  0.2147  -0.1033
 s.e.  0.0919  0.0912   0.0879
 sigma^2 estimated as 0.001619:  log likelihood = 199.02,  aic = -390.03
 Ljung-Box test
 data:  Residuals from ARIMA(0,0,3) with zero mean
 Q* = 8.7878, df = 7, p-value = 0.2683
 Model df: 3.   Total lags used: 10
```

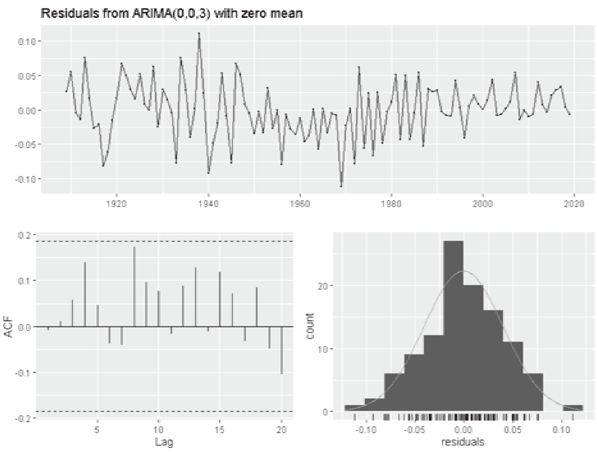

Figure 2.2.7. Residuals from ARIMA(0,0,3) model with zero mean, ACF and Histogram of the series.

Conclusions

We have considered the effect of the sea surface temperature (SST) deviations on the global air temperature deviations from 1891 to 2019. The sample CCF between the SST and the air temperature was estimated by use of the cross-correlation analysis with prewhitening method of ARIMA model fitting.

(1) CCF between global air temperature and global sea surface temperature deviations:

The sample CCF between the two series peaked at lag h = 0 ($\hat{\rho}_{xy}(0) = 0.8$), at lag h = 1 ($\hat{\rho}_{xy}(1) = 0.25$) and at lag h = 18 ($\hat{\rho}_{xy}(18) = 0.24$). These results indicate that the SST might lead the air temperature by one year and 18 years. We next considered the lagged linear regression model of the air temperature on the SST's. It is seen that the significant variables are SST and one year lagged SST.

(2) CCF between global air temperature and each SST of South Pacific Ocean, North Pacific Ocean, South Atlantic Ocean, North Atlantic Ocean and Indian Ocean:

In order to get the effect of the each SST on the global air temperature, we have considered a lagged linear regression model with ARIMA errors and the independent variables of five SST's as the leading indicators of the air temperature. It is seen that the significant variables are five SST's with lag zero, and SST of the South Pacific Ocean with lags one and 18 years and also that of North Pacific Ocean with lag one year as the leading indicators.

It turns out that the global air temperature is strongly subject to the influence of the sea surface temperature. Then, from what kind of factor is the sea surface temperature subject to strong influence? It will be a future subject to explore the factor which has given influence strong against the sea surface temperature.

References

[1] Blockwell,P.J. and R.A. Davis, 2002, *Introduction to Time Series and Forecasting.* Springer Verlag, New York.

[2] Cowpertwait.P.S.P. and A.V. Metcalfe, 2009, *Introductory Time Series with R*, Springer Verlag, New York.

[3] Hansen, J., Sato, M., Ruedy, R., Lo,K., Lea,D.W. and M. Medina-Elizade, 2006, "Global temperature change", *Proceedings of the National Academy of Sciences,* 103(39): 14288-14293.

[4] He,Y., 1995, *Time Series Pack for Mathematica*, Wolfram Research.

[5] Japan Meteorological Agency. "*The long-term change tendency of sea surface temperature*". Retrieved on Feb. 20, 2021, from http://www.data.jma.go.jp/gmd/kaiyou/data/shindan/a_1/glb_warm/glb_warm.html.

[6] Shumway, R.H. and D.S. Stoffer, 2019, *Time Series: A Data Analysis Approach Using R*, CRC Press.

[7] Tanaka, M., 2019, "Notes on Estimation of Cross Correlation Function between two Time Series", *Information Science and Applied Mathematics*, 27, 21-30.