Supporting Information for 'Analysis of air quality time series of Hong Kong with graphical modeling'

F. Hu, Z. Lu, H. Wong and T. P. Yuen

Introduction

This is the supporting information for 'Analysis of air quality time series of Hong Kong with graphical modeling'. The document is organized by first summarizing all the figures or tables displayed. The tables and figures are then displayed after the summaries. In the summaries, there is a hyper-link to the corresponding figure or table. Readers can simply click on the hyper-link, say S1, directing to the corresponding table or figure.

Tables

Table S1 shows the selected parameters in the Box-Cox transformation (Box and Cox, 1964) on the series using the TRANSREG procedure in the SAS/STAT package.

Table S2 summarizes the testing results of using partial coherence and partial cross-correlations. The last column of the table shows whether the two agree or not. The percentage of matching is 86.3%.

Figures

Figure S1 shows the graphs of coherency and partial coherency of the full model, which contains all 12 variables. The lower diagonal part shows the test statistics for partial coherency and the upper diagonal shows the test statistics for coherency. The dotted lines in the upper diagonal part represent the error bound for coherency, which is given by the 95% quantile of the $F(2, 106)$ distribution, with a value of 3.08. The dotted lines in the lower diagonal part represent the error bound for partial coherency, which is given by the 95% quantile of the $F(2, 88)$ distribution, with a value of 3.10.

Figure S2 shows the graphs of cross-correlation and partial cross-correlation of the full model using the VAR method. The lower diagonal part shows the partial cross-correlation and the upper diagonal shows the cross-correlation. The dotted

lines represent the approximate error bound of $\pm 2/\sqrt{n}$, namely $\pm 2/\sqrt{1941} = 0.0518$.

Figures S3 – S5 show the ACF plots of the deseasonalized series. The dotted lines represent the Bartlett's large lag standard error.

Figure S6 shows the ACF plot of the first-order difference of the deseasonalized SO_2 series at Tap Mun. The dotted lines represent the Bartlett's large lag standard error.

Matlab Code

The Matlab program to generate the plots may be obtained in the online supplements of this article at the publisher's website, inside a rar file ('MatlabCode.rar'). User can extract the rar file and run the main function to generate the figures.

Dateset

The dataset analyzed may be obtained in the online supplements of this article at the publisher's website, a csv file ('Data.csv'). There is a very small amount of missing data which are replaced using linear interpolation. Extreme values (values outside ±3.5 standard deviations) are replaced with 7 points moving averages using the EXPAND procedure in the SAS/ETS package. The data attributes are summarized as follows.

Attribute	Data type	Description
date	Date	Date.
city	Character	City.
cityCode	Character	City code.
pollutant	Character	Pollutant.
value	Numeric	Daily average of pollutant after replacing missing values using linear interpolation and replacing extreme values with 7 points moving averages.
valueBoxCox	Numeric	Value after Box-Cox transformation on 'value'
valueBoxCoxDs	Numeric	Deseasonalized value on 'valueBoxCox'
valueBoxCoxDsDiff	Numeric	First order difference on 'valueBoxCoxDs' for the SO_2 at Tap Mun. Others remain the same as 'valueBoxCoxDs'.

Figure S1: Plot of test statistics for coherency and partial coherency. The lower diagonal part shows the test statistics for partial coherency and the upper diagonal shows the test statistics for coherency.

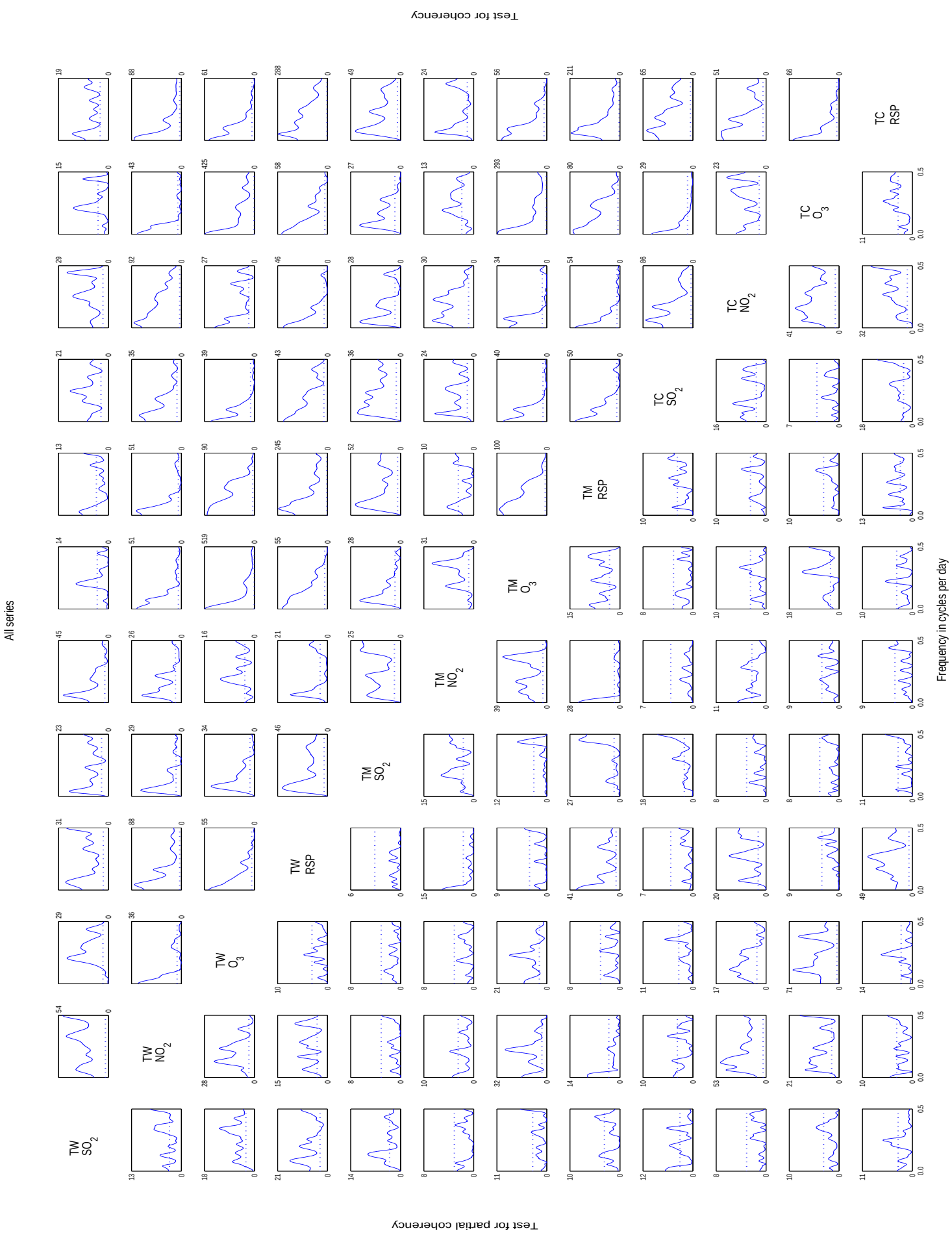


Figure S2: Plot of cross-correlation and partial cross-correlation. The lower diagonal part shows the partial cross-correlation and the upper diagonal shows the cross-correlation.

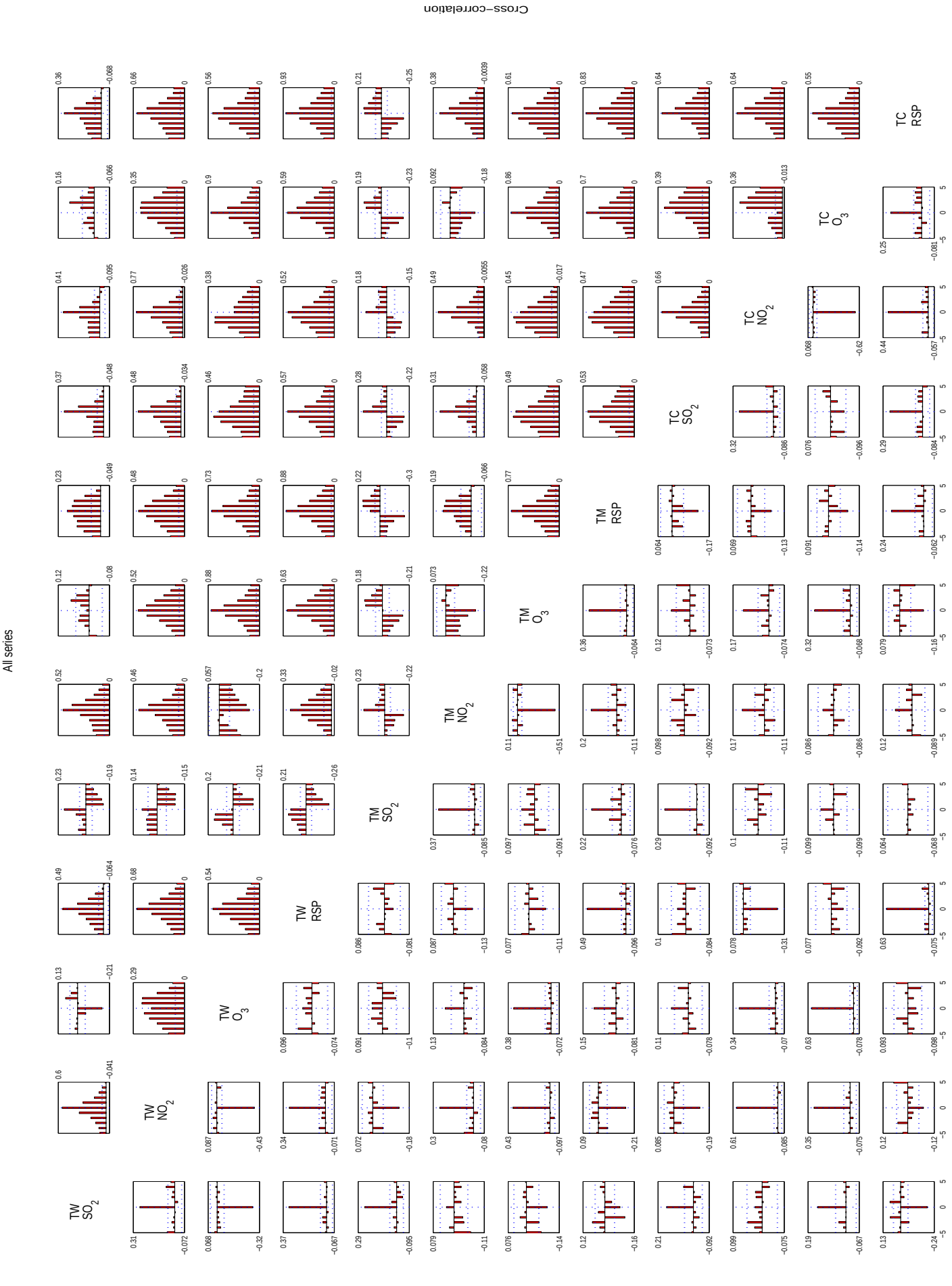
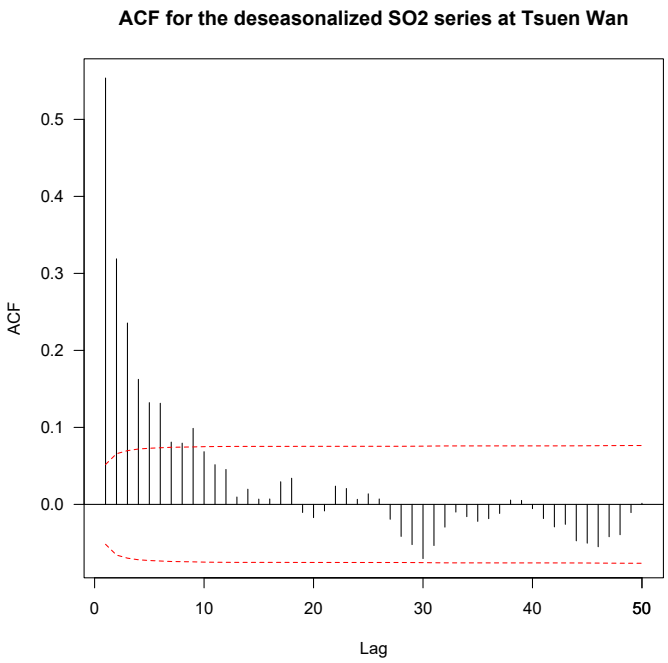
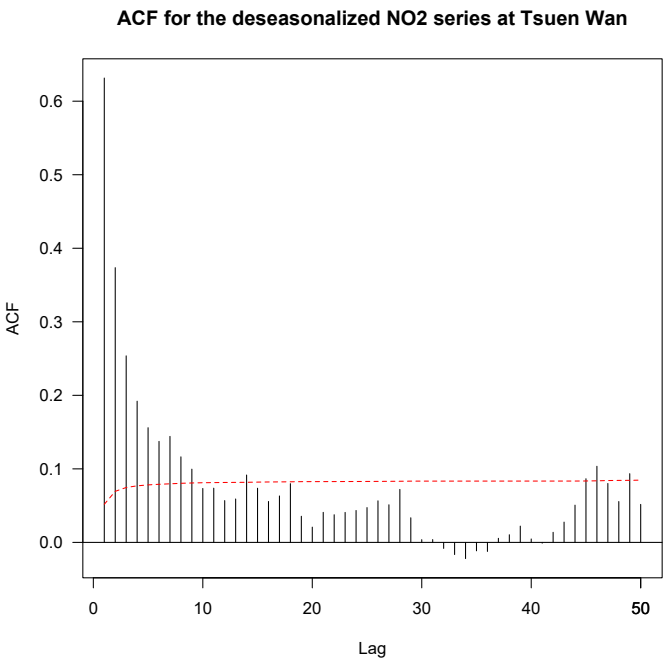


Figure S3: ACF plot of the deseasonalized series at Tsuen Wan. The dotted lines represent the Bartlett's large lag standard error.

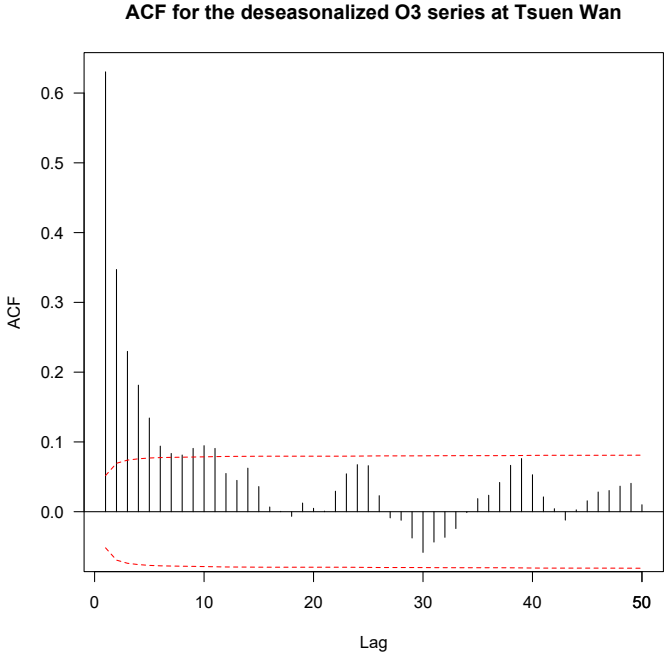
(a) SO_2



(b) NO_2



(c) O_3



(d) RSP

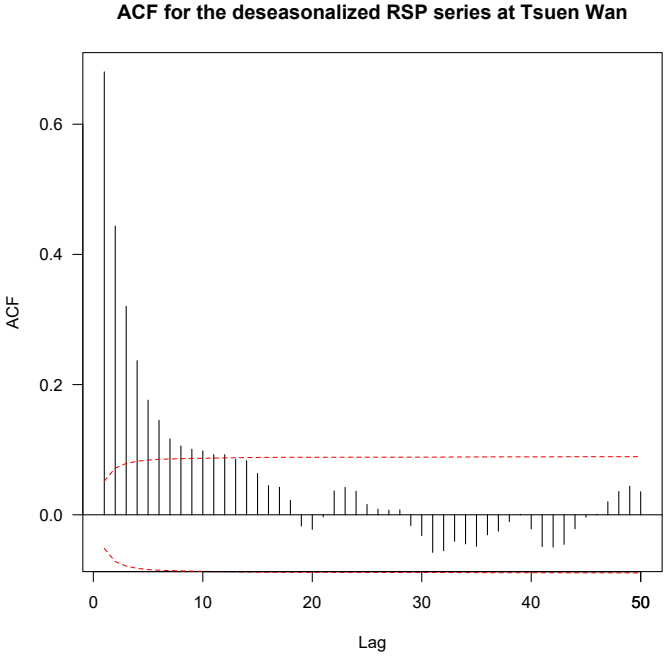
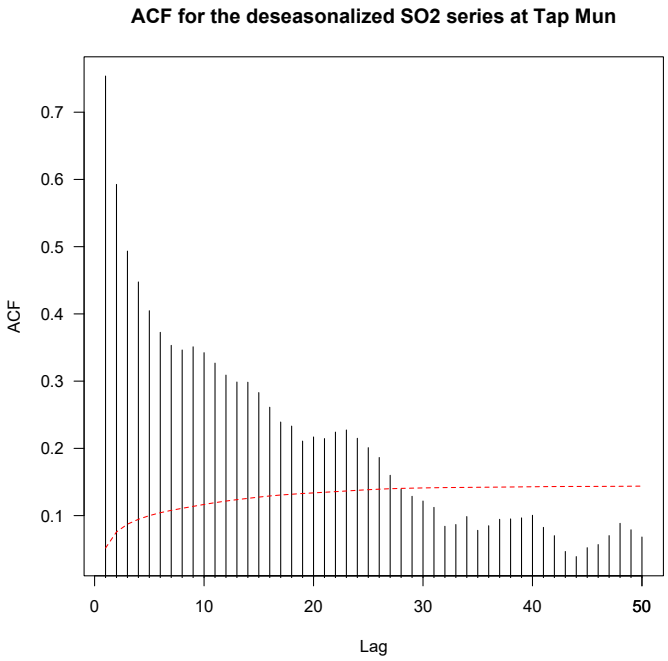
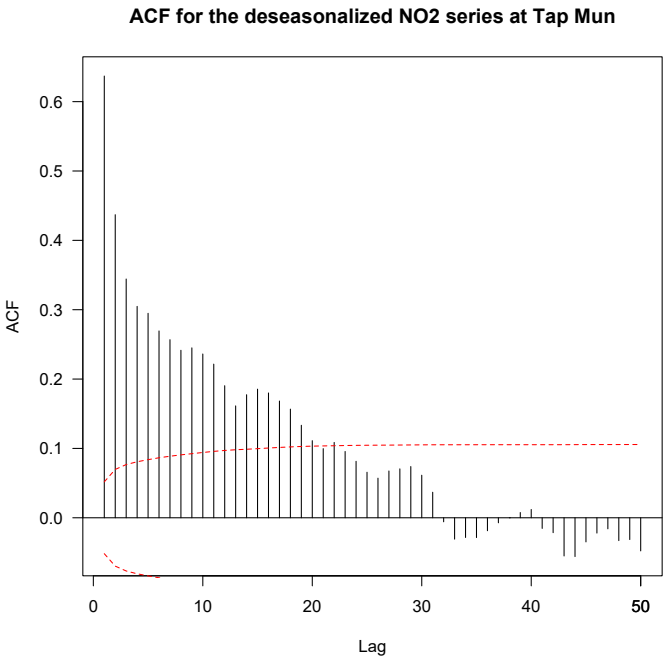


Figure S4: ACF plot of the deseasonalized series at Tap Mun. The dotted lines represent the Bartlett's large lag standard error.

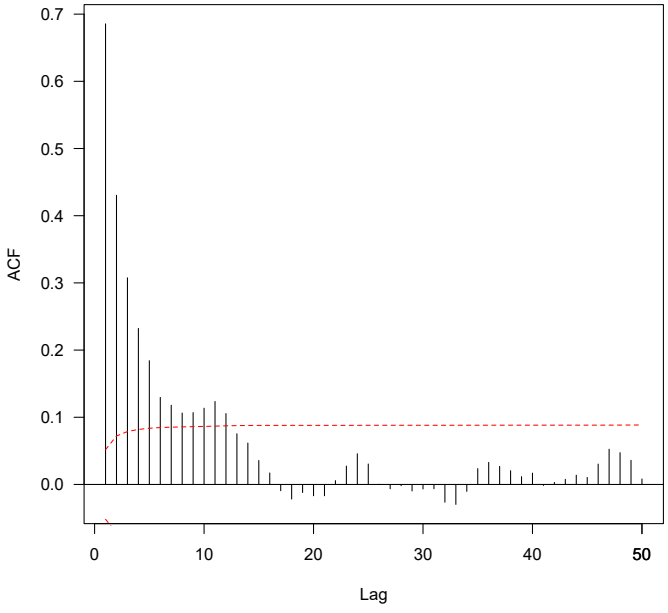
(a) SO_2



(b) NO_2



(c) O_3



(d) RSP

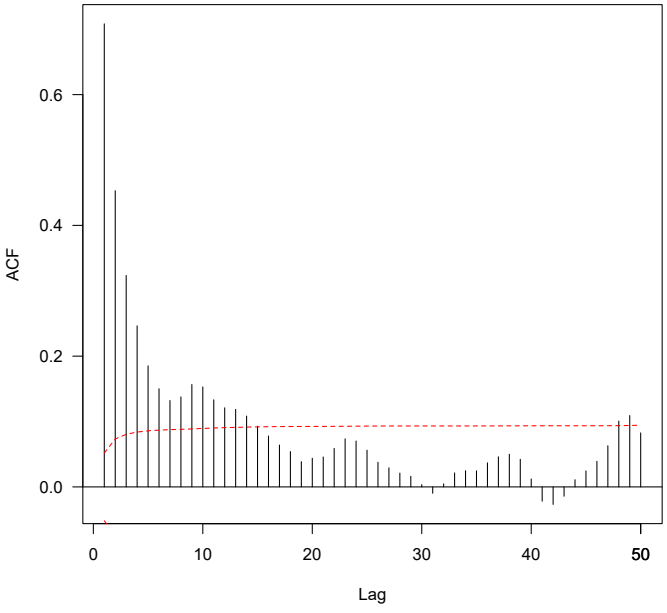
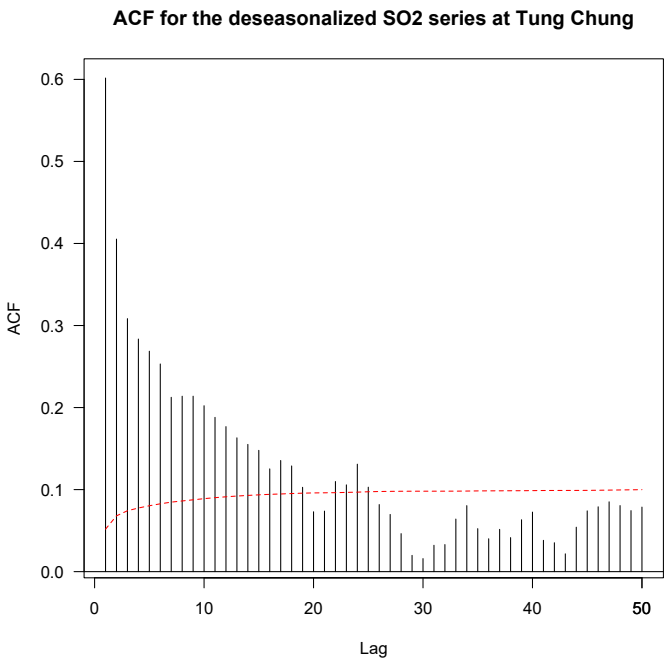
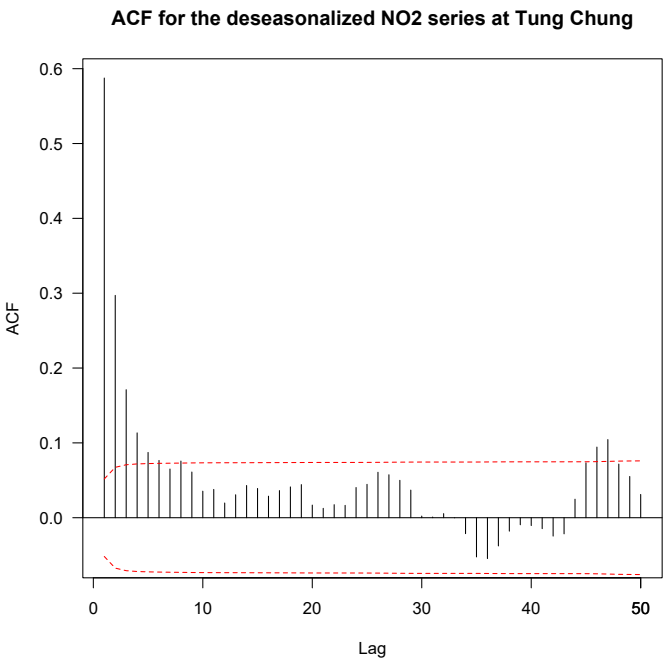


Figure S5: ACF plot of the deseasonalized series at Tung Chung. The dotted lines represent the Bartlett's large lag standard error.

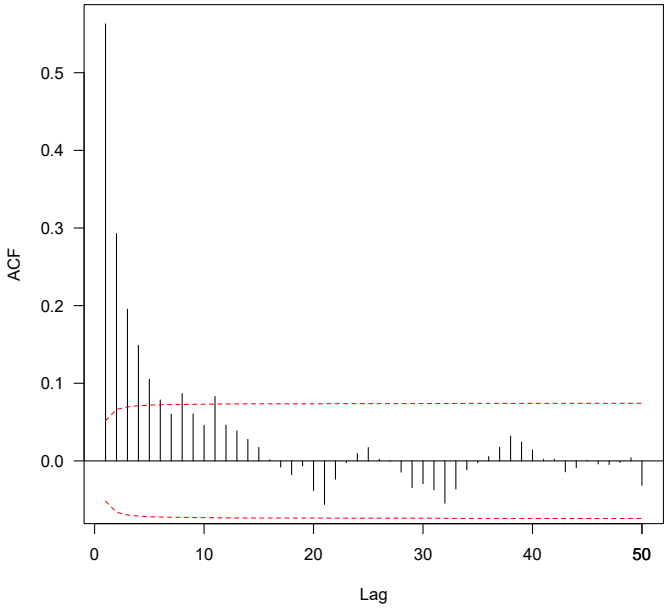
(a) SO_2



(b) NO_2



(c) O_3



(d) RSP

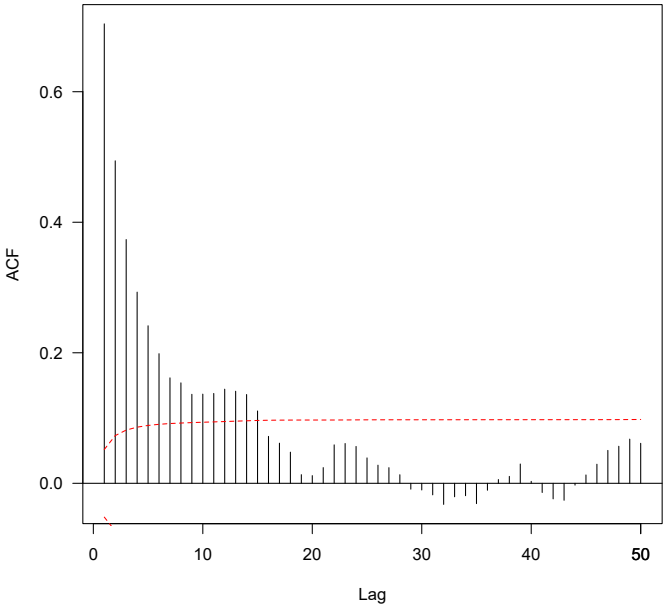
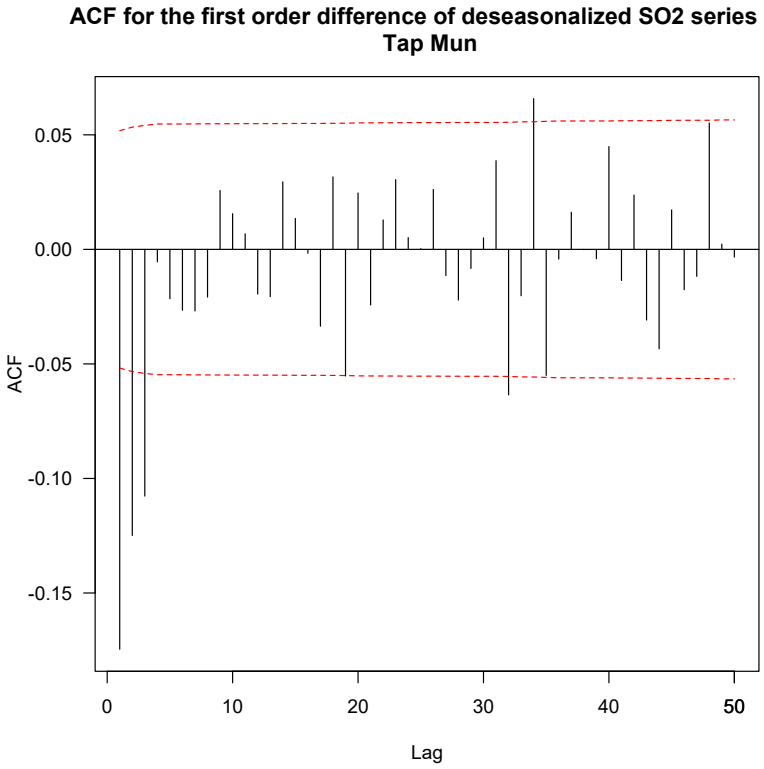


Figure S6: ACF plot of the first-order difference of deseasonalized SO_2 series at Tap Mun. The dotted lines represent the Bartlett's large lag standard error.



Tables

Table S1: Box-Cox transformation parameters for the 12 time series

Series	λ
Tap Mun SO2	0.5
Tap Mun NO2	0.25
Tap Mun O3	0.25
Tap Mun RSP	0.25
Tsuen Wan SO2	0.25
Tsuen Wan NO2	0
Tsuen Wan O3	0.25
Tsuen Wan RSP	0
Tung Chung SO2	0
Tung Chung NO2	0.5
Tung Chung O3	0.5
Tung Chung RSP	0.25

Table S2: Comparison of testing results of partial coherence and VAR methods

Series 1	Series 2	Partial Coherence	VAR	Match
Tsuen Wan SO2	Tsuen Wan NO2	Yes	Yes	Yes
Tsuen Wan SO2	Tsuen Wan O3	Yes	Yes	Yes
Tsuen Wan SO2	Tsuen Wan RSP	Yes	Yes	Yes
Tsuen Wan NO2	Tsuen Wan O3	Yes	Yes	Yes
Tsuen Wan NO2	Tsuen Wan RSP	Yes	Yes	Yes
Tsuen Wan O3	Tsuen Wan RSP	Yes	No	No
Tsuen Wan SO2	Tap Mun SO2	Yes	Yes	Yes
Tsuen Wan SO2	Tap Mun NO2	No	No	Yes
Tsuen Wan SO2	Tap Mun O3	Yes	Yes	Yes
Tsuen Wan SO2	Tap Mun RSP	Yes	Yes	Yes
Tsuen Wan NO2	Tap Mun SO2	No	Yes	No
Tsuen Wan NO2	Tap Mun NO2	Yes	Yes	Yes
Tsuen Wan NO2	Tap Mun O3	Yes	Yes	Yes
Tsuen Wan NO2	Tap Mun RSP	Yes	Yes	Yes
Tsuen Wan O3	Tap Mun SO2	No	No	Yes
Tsuen Wan O3	Tap Mun NO2	No	Yes	No
Tsuen Wan O3	Tap Mun O3	Yes	Yes	Yes
Tsuen Wan O3	Tap Mun RSP	No	Yes	No
Tsuen Wan RSP	Tap Mun SO2	No	No	Yes
Tsuen Wan RSP	Tap Mun NO2	Yes	Yes	Yes
Tsuen Wan RSP	Tap Mun O3	No	Yes	No
Tsuen Wan RSP	Tap Mun RSP	Yes	Yes	Yes
Tsuen Wan SO2	Tung Chung SO2	Yes	Yes	Yes
Tsuen Wan SO2	Tung Chung NO2	No	No	Yes
Tsuen Wan SO2	Tung Chung O3	Yes	Yes	Yes
Tsuen Wan SO2	Tung Chung RSP	Yes	Yes	Yes
Tsuen Wan NO2	Tung Chung SO2	Yes	Yes	Yes
Tsuen Wan NO2	Tung Chung NO2	Yes	Yes	Yes
Tsuen Wan NO2	Tung Chung O3	Yes	Yes	Yes
Tsuen Wan NO2	Tung Chung RSP	Yes	Yes	Yes
Tsuen Wan O3	Tung Chung SO2	Yes	Yes	Yes
Tsuen Wan O3	Tung Chung NO2	Yes	Yes	Yes
Tsuen Wan O3	Tung Chung O3	Yes	Yes	Yes
Tsuen Wan O3	Tung Chung RSP	Yes	No	No
Tsuen Wan RSP	Tung Chung SO2	No	No	Yes
Tsuen Wan RSP	Tung Chung NO2	Yes	Yes	Yes
Tsuen Wan RSP	Tung Chung O3	No	No	Yes
Tsuen Wan RSP	Tung Chung RSP	Yes	Yes	Yes
Tap Mun SO2	Tap Mun NO2	Yes	Yes	Yes
Tap Mun SO2	Tap Mun O3	Yes	No	No
Tap Mun SO2	Tap Mun RSP	Yes	Yes	Yes
Tap Mun NO2	Tap Mun O3	Yes	Yes	Yes
Tap Mun NO2	Tap Mun RSP	Yes	Yes	Yes
Tap Mun O3	Tap Mun RSP	Yes	Yes	Yes
Tap Mun SO2	Tung Chung SO2	Yes	Yes	Yes
Tap Mun SO2	Tung Chung NO2	No	No	Yes
Tap Mun SO2	Tung Chung O3	No	No	Yes
Tap Mun SO2	Tung Chung RSP	Yes	No	No
Tap Mun NO2	Tung Chung SO2	No	No	Yes

Table S2 – *Continued from previous page*

Series 1	Series 2	Partial Coherence	VAR	Match
Tap Mun NO2	Tung Chung NO2	Yes	Yes	Yes
Tap Mun NO2	Tung Chung O3	No	No	Yes
Tap Mun NO2	Tung Chung RSP	Yes	Yes	Yes
Tap Mun O3	Tung Chung SO2	No	Yes	No
Tap Mun O3	Tung Chung NO2	Yes	Yes	Yes
Tap Mun O3	Tung Chung O3	Yes	Yes	Yes
Tap Mun O3	Tung Chung RSP	Yes	Yes	Yes
Tap Mun RSP	Tung Chung SO2	Yes	Yes	Yes
Tap Mun RSP	Tung Chung NO2	Yes	Yes	Yes
Tap Mun RSP	Tung Chung O3	Yes	Yes	Yes
Tap Mun RSP	Tung Chung RSP	Yes	Yes	Yes
Tung Chung SO2	Tung Chung NO2	Yes	Yes	Yes
Tung Chung SO2	Tung Chung O3	No	No	Yes
Tung Chung SO2	Tung Chung RSP	Yes	Yes	Yes
Tung Chung NO2	Tung Chung O3	Yes	Yes	Yes
Tung Chung NO2	Tung Chung RSP	Yes	Yes	Yes
Tung Chung O3	Tung Chung RSP	Yes	Yes	Yes

References

Box GEP, Cox DR, 1964. An analysis of transformations. *Journal of the Royal Statistical Society: Series B* **26**(2): 211–252.