# Report - SF2943 Time Series Analysis

Group TS 6: Buqing Cao Titing Cui Xin Huang Zeyu Chen

May 16, 2018

## Part 1

### Introduction to the Data

The Mauna Loa carbon dioxide record is an iconic symbol of the human capacity to alter the planet. From the source: Time Series Data Library (citing: Hipel and McLeod

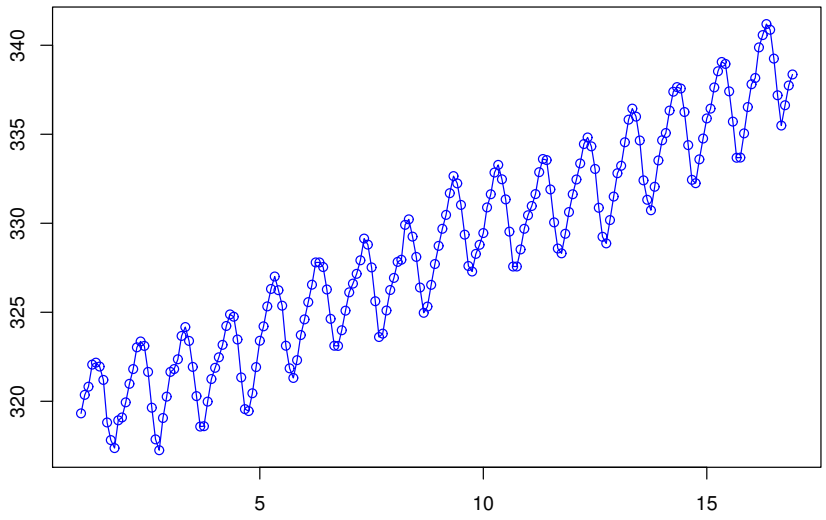


Figure 1: $CO_2$ (ppm) mauna loa, 1965-1980

(1994)), we can retrieve the dataset containing the measurements for $CO_2$-concentration in the atmosphere of Mauna Loa, Hawaii. From 1965 to 1980, a measured value of concentration (with unit ppm) is recorded for each month, and a total number of 193 data points

are available in this dataset. In the project, we mainly applied the R language with its advantageous packages *itsmr*, *forecast*, and *tseries*.

## Preprocessing

Since the original time series data exhibits a quite obvious characteristic of trend and seasonality, we made an assumption that this $CO_2$ series follows the classical additive model:

$$Y_t = T_t + S_t + X_t$$

where $T_t$ is the trend , $S_t$ is the seasonality , and $X_t$ is the stationary component. By applying a differencing operator with lag d = 12, because $S_t$ has a period d = 12, we can obtain:

$$\nabla_d Y_t = Y_t - Y_{t-d} = (1 - B^d)Y_t = T_t - T_{t-d} + X_t - X_{t-d}$$

Now the polynomial trend term $T_t - T_{t-d}$ can be eliminated by applying a power of the differencing operator $\nabla$. Specifically we find that the trend can be eliminated with the power of the differencing operator which equals to 1. In R, the differencing operator is realized with function *diff*.

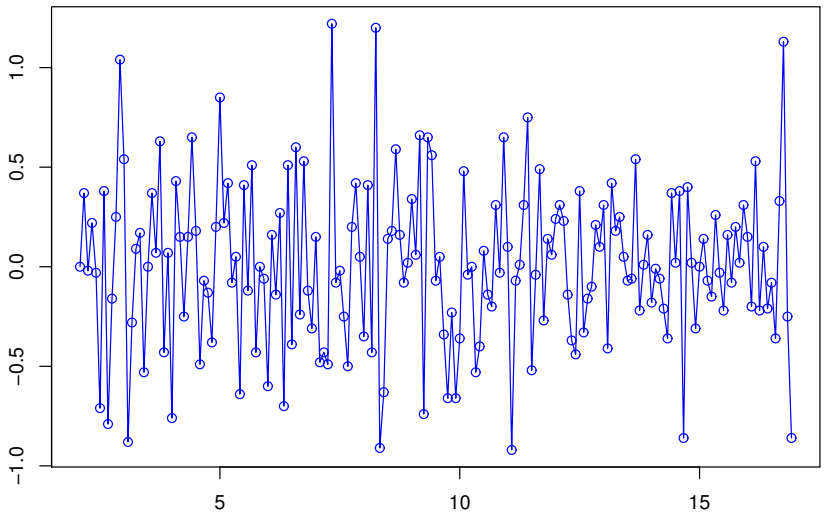


Figure 2: Pre-processed time series with detrend and deseasonality

After the differencing, we get the preprocessed series $\hat{X}_t$. Then we use the Augmented

Dickey-Fuller test (ADF test) to check the the stationarity of the series. The ADF test is conducted with the function *adf.test()* of package *tseries*. The result of the ADF test can sufficiently reject the null hypothesis of a unit root at level 0.05. As also can be seen from Figure 2, the pre-processed time series doesn't show serious deviation from stationarity.

## Model Analysis and Evaluation

In this section, first we compute the ACF and PACF values for the preprocessed time series $\hat{X}_t$, which are realized with the functions *acf* and *pacf*. From Figure 3, we can see

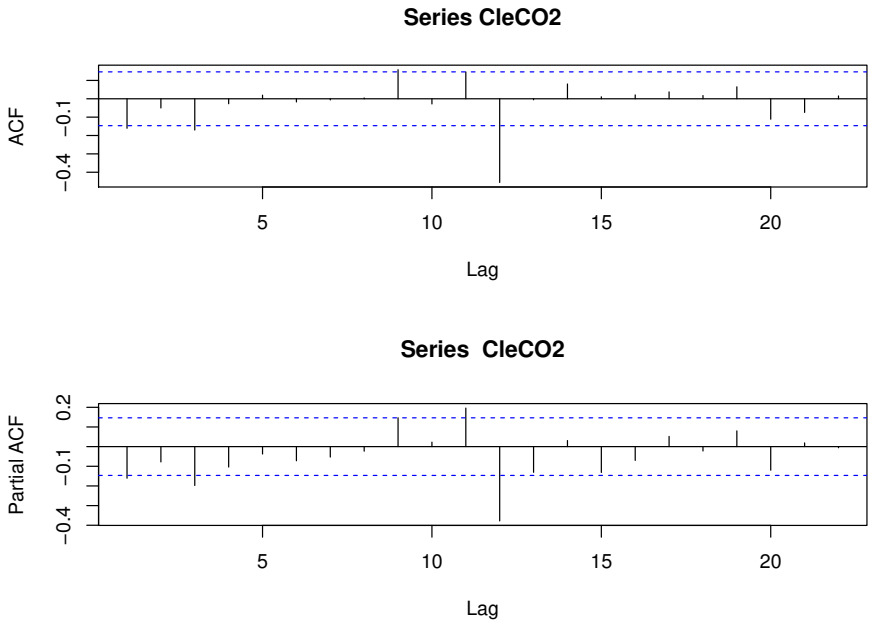


Figure 3: ACF and PACF of $\hat{X}_t$

that the ACF value is significant when lag = 12, which implies a seasonal ARMA model may be suitable. With the help of the function *auto.arima()* in *forecast* package, we can obtain that a model of $(1,0,1) \times (0,0,1)_{12}$ with mean zero has the smallest AICC value for our pre-differenced data. The searching loop for the best model is implemented with parameters p and q starting from 0 and their maximum is set as 5, P and Q are from 0 to 2. Combining the pre-differencing procedure $(1-B)(1-B^{12})Y_t$, our final model is SARIMA model with parameters $(1,1,1) \times (0,1,1)_{12}$. With maximum likelihood method, the coefficients for the final model are estimated as:

| Coefficients | Estimate | Standard error |
|---|---|---|
| $ar_1$ | 0.3714 | 0.1746 |
| $ma_1$ | -0.6709 | 0.1378 |
| $sma_1$ | -0.8293 | 0.0871 |

Using the function *tsdiag*, we can receive some diagnostics of this final SARIMA model. From Figure 4, we can observe that for sample ACF of residuals, no values fall outside the significant level bounds, and also the Ljund-Box test is comfortably passed at the significant level 0.05. The Gaussian QQ plot of residuals is also consistent with the normal distribution assumption.

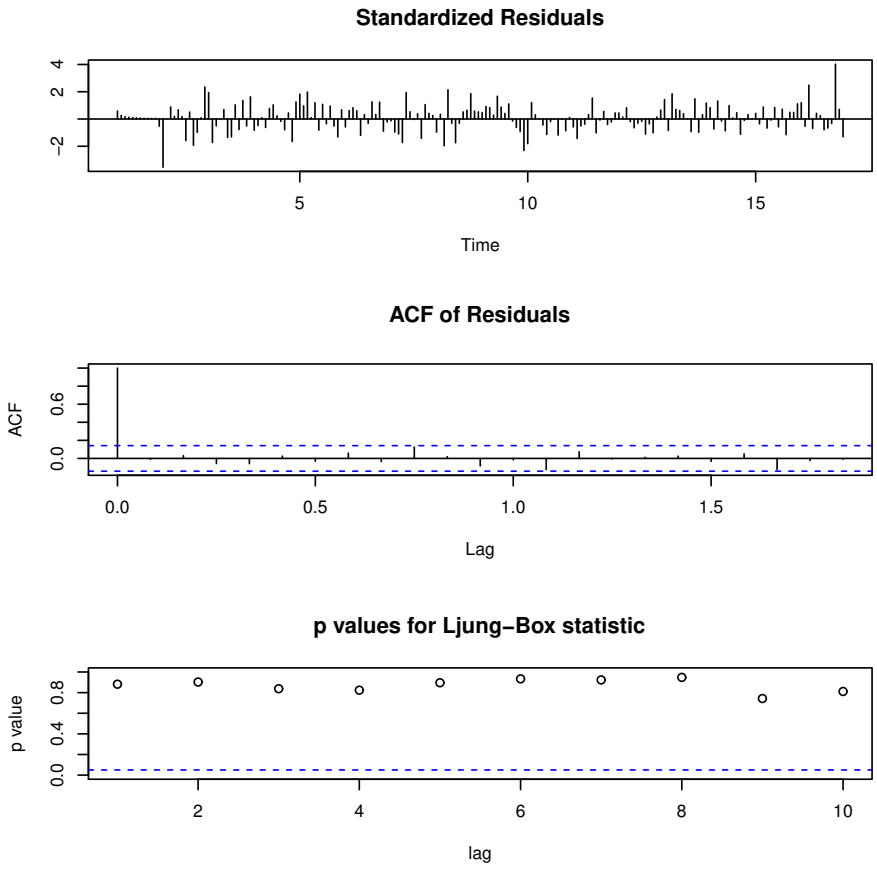


Figure 4: Diagnostics for the residuals of the final SARIMA model

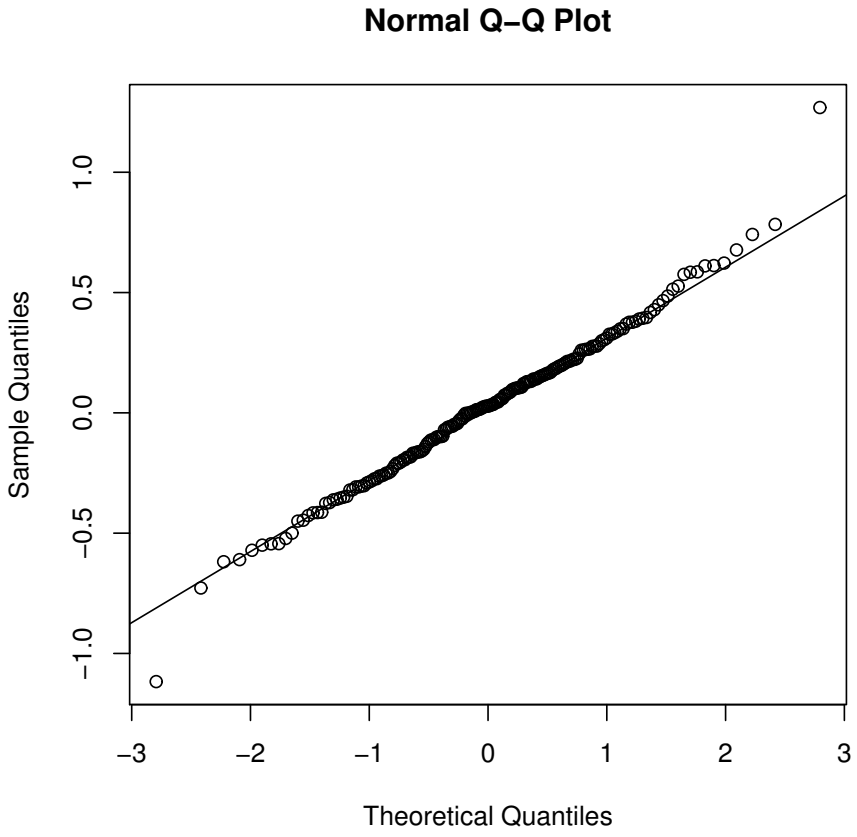


Figure 5: Q-Q plot for the residuals of the final SARIMA model

## Forecast

For the $(1,1,1) \times (0,1,1)_{12}$ SARIMA model, we use the function *forecast* in the package *forecast* to predict the $CO_2$-concentration in the following year. The forecast is applied with default parameters. The results are shown in Figure 6, with the gray intervals showing an 80% and 95% prediction intervals for the forecast. It can bee seen that, the forecast of the final model is relatively stable.

## Conclusion

From the graph of the original $CO_2$ data, we can see that the additive model seems applicable. First, we eliminate the trend and seasonality of the original data with differencing method. Second, we use a $(1,0,1) \times (0,0,1)_{12}$ SARIMA model to characterize the station-

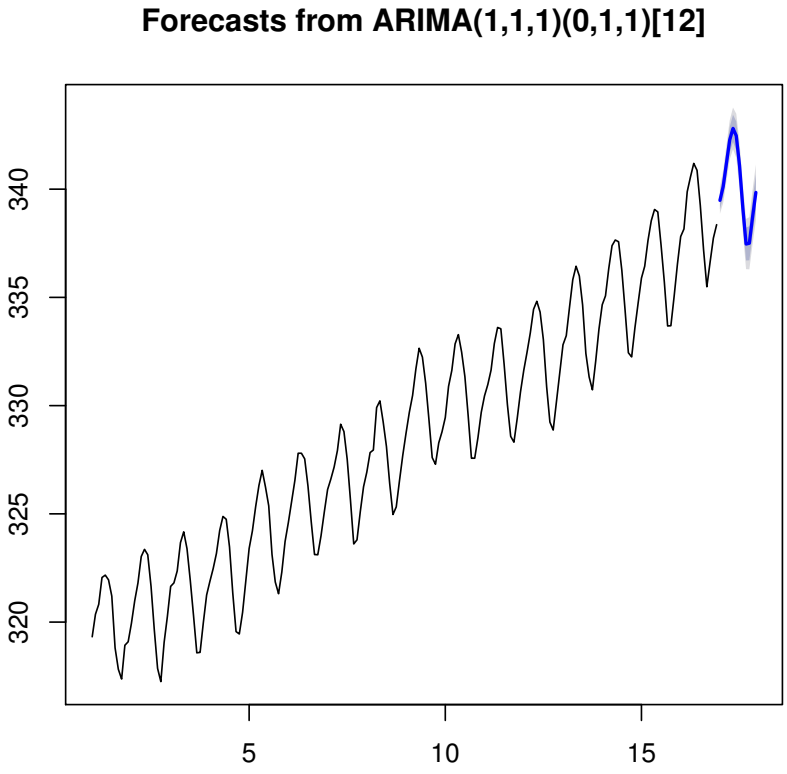


Figure 6: The forecasted CO2 concentration for the next 12 months from the final SARIMA model

ary component. The tests of the residuals show that our model has achieved preferable results.

## Peer Review

The project report conducted a comprehensive and through analysis of the presented data, showing good knowledge of many helpful methods and skills. Especially, the Augmented Dickey-Fuller test is applied to determine the differencing order, which perfectly follows the recommended work flow introduced in the course. Also, the ACF values are presented for preliminarily differenced series, which provided a convincing evidence for suitability of model assumptions. Further, the well-acknowledged Akaike Information Criterion and Bayesian Information Criterion are implemented in the model selection process, which

gives a quite validated result of ARMA (1, 2) model. The forecast also gives a reasonable prediction based on the ARIMA (1,1,2 ) model.

There are just a few small problems. From the graph of the original data, we guess that a quadratic trend may be more suitable. Especially after 1980, a growing trend is more significant. Therefore, considering the differencing with order = 2, i.e. $(1B)^2$ may be helpful. Besides, the diagnostics of the residuals can be performed with more applicable methods. A qq-norm test and Ljung-Box test may provide more supportive information for the model.

# Part 2

# Part 3

# References

[1] Brockwell, P. J. and Davis R. A. *Introduction to Time Series and Forecasting,* 2nd edition, 8th corrected printing. Springer, 2010.