TMA4285 Timeseries, Exercise 3

Marius Dioli, Amir Ahmed and Andreas Ferstad

September 2019

Abstract

This article presents the basics of time series analysis, and applies it to model monthly average atmospheric CO2 levels observed at Mauna Loa Observatory, Hawaii, U.S.

Introduction

Carbon dioxide levels have been monitored since 1958 at the Mauna Loa Observatory, located in the Pacific Ocean on top of Hawaii's biggest volcano. CO2 is a green-house gas, and as humans burn fossile fuels, more is added to the atmosphere. As a consequence we have since the industrial revolution seen an increase in global average temperature. In the Paris agreement one targeted to stay below a 2C increase, commonly that has been seen to be equivalent to about 450ppm in atmospheric CO2. In this article we will use theory on time series to model CO2 levels measured at the observatory, with the goal of forecasting future levels, and to predict when, if development continues as today, when will reach the 450ppm roof. For analysis, we will focus on flask measurements taken on a monthly basis as we are interested in predicting the long-term trend.

Theory

We note that large parts of the theory presented here, is based on Brockwell and Davis.

A timeseries can be considered as a stochastic process, a set of random variables indexed in accordance to their occurrence relative to the other measurements. We will organize the finite timeseries we will deal with into vectors: $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$. A realization of a timeseries is an observation of the the real values of the time series, and will be denoted as $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$

The covariance function of two elements of $\mathbf{X}$ is defined as:

$$\gamma_X(r, s) = Cov(X_r, X_s) = E((X_r - E(X_r))(X_s - E(X_s)))$$

In general the covariance function tells us how observations correlate to each other.

A time series is weakly stationary if:

- $E(X_t)$ is independent of t

- $\gamma_X(t + h, t)$ is independent of t at each h value.

In a weakly stationary timesries the correlation between two observations would only be depended the relative time between the observation of the repsective observations.

The auto correlation function at a selected lag h is:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

Where lagging a time-series is shifting it forward by a selected amount of indexes.

The estimator of the autocovariance of a realized time-series, referred to as the sample co-variance is:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x})$$

For white noise time-series sample autocorrelation $\gamma(h) \sim N(0, \frac{1}{n})$ distribution. Under general assumptions this estimator is for large samples close to unbiased.

The time series we will observe in the data-analysis later, is classically decomposed in the following manner:

$$X_t = m_t + s_t + Y_t, \quad t = 1, \ldots, n$$

Where $EY_t = 0$, $s_{t+d} = s_t$ and $\sum_{j=1}^{d} s_j = 0$. m is referred to as the trend of the time-series, and s_t is the seasonal trend. The noise Y_t is often assumed to be identically independently distributed (iid) white noise:

$$Y_t \sim WN(0, \sigma^2)$$

This assumption and this estimator can be used when creating autocorrelation (ACF) and partial autocorrelation plots (PACF),

Subtracting the trend estimation, will yield as timeseries without a trend. A trend can also be eliminated from a series by differencing:

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

This is referred to as lag-1 difference, where B is the backward shift operator.

$$BX_t = X_{t-1}$$

Can difference n-times by:

$$\nabla^n X_t = (1 - B)^n X_t$$

Some data, as stated earlier, show a seasonal trend, this can e.g. be that months of different years correlate, applying a lag-d difference operator is a way of reducing such a case to a stationary process, a lag-d operator can be defined as:

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t$$

A time series **X** is a linear process if:

$$X_t = \sum_{j=-\infty}^{\infty} \phi_j Y_j$$

with $\sum_{j=-\infty}^{\infty} |\phi_j| < \infty$. Where **Y** is white noise.

ARMA and ARIMA

This brings us to the ARMA and SARIMA models, the latter which will be a central part in our analysis. An $ARMA(p, q)$ model is defined as the following:

$$\phi(B)X_t = \theta(B)Z_t$$

where Z_t is white noise, iid with variance σ^2

and the SARIMA;

$$\phi(B)\Phi(B)X_t = \theta(B)\Theta(B)Z_t$$

More specifically we have:

$$ARMA(p, q): \quad Y_t = (1 - B)^d X_t$$
$$SARIMA(p, d, q) \times (P, D, Q)_s: \quad Y_t = (1 - B)^d (1 - B^s)^D X_t$$

Where:

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$
$$\Theta(z) = 1 + \Theta_1 z + \cdots + \Theta_Q z^Q$$
$$\phi(z) = 1 - \phi_1 z + \cdots + \phi_p z^p$$
$$\Phi(z) = 1 - \Phi_1 z + \cdots + \Phi_P z^P$$

Multiplying the polynomials together, this is equivalent to a $ARMA(p + sP, q + sQ)$ model, and fitting it would thus in essence be equivalent to fitting an ARMA model.

The innovations algorithm

To fit the model one can use the innovations algorithim. Suppose $\{X_t\}$ is zero mean, with $E(X_t)^2 < \infty$ for all t and suppose

$$E(X_i X_j) = \kappa(i, j)$$

Denote the best linear predictors, and their mean squared erros as:

$$\hat{X}_n = \begin{cases} 0, n = 1 \\ P_{n-1} X_n, n \geq 2 \end{cases}$$

$$v_n = E(X_{n+1} - P_n X_{n+1})^2$$

Where P_i is the projection onto the space spanned by the i first elements of the timeseries in the $L^2(\Omega)$ hilbert-space with inner-product:

$$< X_i, X_j >= E(X_i X_j)$$

Define the innovations as:

$$U_n = X_n - \hat{X}_n$$

and let $\mathbf{U}$ be a vector with the innovations ranging from 1 to n as its elements. As all X_n are assumed to be the result of some linear combination of the preceding values there exists a matrix A such that

$$\mathbf{U}_n = A_n \mathbf{X}_n$$

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{11} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1,-n-1} & a_{n-1,-n-2} & a_{n-1,-n-3} & \cdots & 1 \end{bmatrix}$$

This matrix is invertible, as it is is unit lower triangular (it has a non-zero determinant of 1). The inverse will be on the following form:

$$C_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,-n-1} & \theta_{n-1,-n-2} & \theta_{n-1,-n-3} & \cdots & 1 \end{bmatrix}$$

$$A_n C_n = C_n A_n = I_n$$

This yields:

$$\hat{\mathbf{X}}_n = \mathbf{X}_n - \mathbf{U}_n = C_n \mathbf{U}_n - \mathbf{U}_n = C_n \mathbf{X}_n - C_n \hat{\mathbf{X}}_n - \mathbf{X}_n + \hat{\mathbf{X}}_n = (C_n - I_n)(\mathbf{X}_n - \hat{\mathbf{X}}_n)$$

Define

$$\Theta_n = C_n - I_n$$

$$\Rightarrow \hat{\mathbf{X}}_n = \Theta_n(\mathbf{X}_n - \hat{\mathbf{X}}_n)$$

Combining this with the fact that:

$$\mathbf{X}_n = C_n(\mathbf{X}_n - \hat{\mathbf{X}}_n)$$

solving for the coefficients above can easily done numerically, Brockwell and Davis suggest calculating recursively yielding the innovation algorithm:

$$v_0 = \kappa(1, 1)$$

3

$$\theta_{n,n-k} = v_k^{-1}(\kappa(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta$$

$$v_n = \kappa(n+1,n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j$$

ARIMA and the innovation algorithm Forecasting ARMA process using the innovation algorithim. Assume we have an ARMA process with known (p,q) values. We transform the process into:

$$\begin{cases} W_t = \sigma^{-1}X_t, & t = 1,\ldots,m \\ W_t = \sigma^{-1}\phi(B)X_t, & t > m \end{cases}$$

$m = max(p,q)$ The autocovariance can then be found using:

$$\kappa(i,j) = \begin{cases} \sigma^{-2}\gamma_X(i-j), & 1 \le i,j \le m \\ \sigma^{-2}(\gamma_x(i-j) - \sum_{i=1}^{p}\phi_r\gamma_X(r-|i-j|)), & min(i,j) \le m < max(i,j) \le 2m \\ \sum_{r=0}^{q}\theta_r\theta_{r+|i-j|}, & min(i,j) > m, \\ 0 \end{cases}$$

Where applying the innovation algorithim to W yields:

$$\begin{cases} \hat{W}_{n+1} = \sum_{j=1}^{n}\theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}), & 1 \le n < m \\ \hat{W}_{n+1} = \sum_{j=1}^{q}\theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}), & n \ge m \end{cases}$$

This enables us to write X_t as a linear combination of the W_t-s vice versa.

We note:

$$\hat{W}_{n+1} = P_n W_{n+1}$$

$$\hat{X}_{n+1} = P_n X_{n+1}$$

Where P is a projection in the hilbert space $L^2(\Omega)$.

Using this we get

$$\begin{cases} \hat{W}_t = \sigma^{-1}\hat{X}_t, & 1 \le t \le m \\ \hat{W}_t = \sigma^{-1}(\hat{X}_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}), & t > m \end{cases}$$

Combining this we get:

$$X_t - \hat{X}_t = \sigma(W_t - \hat{W}_t)$$

which in turns yield:

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n}\theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & 1 \le t < m \\ \phi_1 X_n \cdots + \phi_p X_{n+1+p} + \sum_{j=1}^{q}\theta_{nj}X & t \ge m \end{cases}$$

Model comparison, finding values for p and q

Selecting different values for p and q would yield different models. Using AICC and the AIC criterion is one way of selecting among these. The methods both base themselves on the idea of scoring on the likelihood of seeing the realized values given the model and then to punish the complex models more. We prefer simple models as they are less prone to over-fitting the data. I.e. will mean square error not be able to pick up upon this, so more complex models will always be favored, which is why we should use other criteria as AIC or AICC.
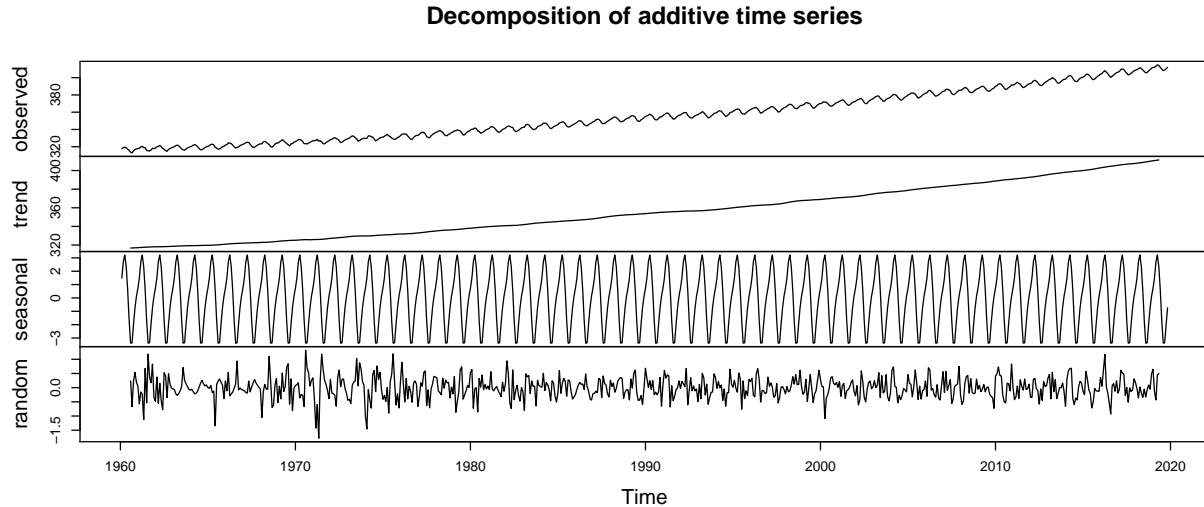
Data Analysis

We start by creating simple time series plot to get an overview.

Observed monthly CO2 levels
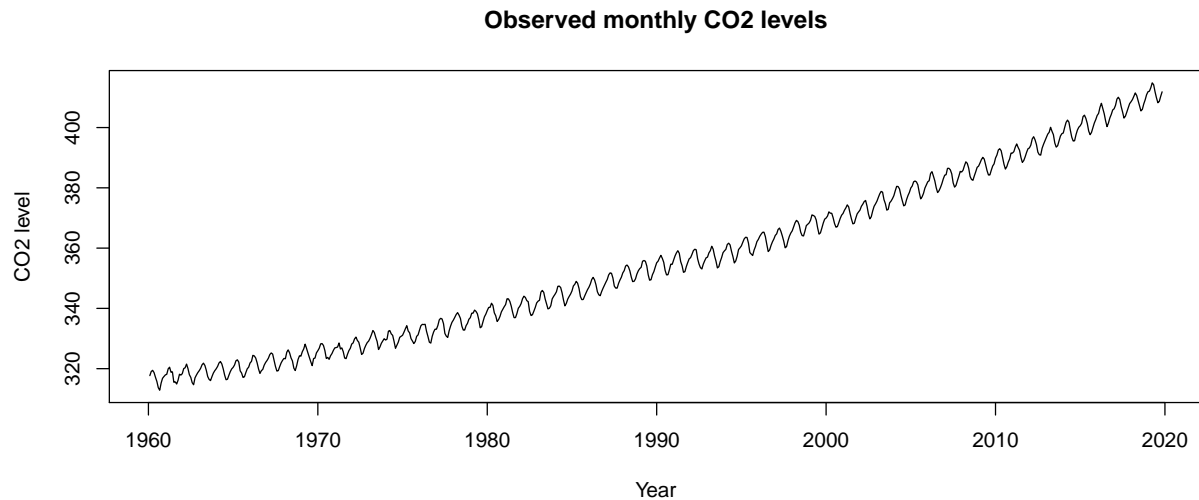
As expected we see an increase. There also seems to be a seasonal trend.

We can decompse the timeseries to get a closer look.

Decomposition of additive time series

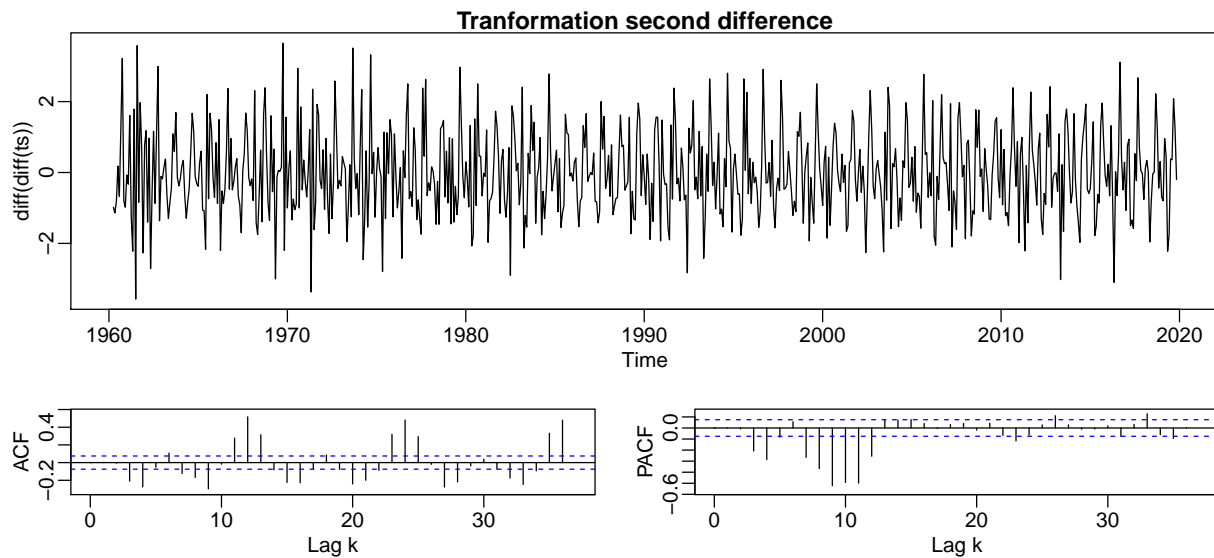With this we see a much clearer sesonal trend.

The data is definitely not stationary, thus we apply differencing (∇) to see if it gets any better.

Tranformation first difference

Applying ∇^2:

Tranformation second difference
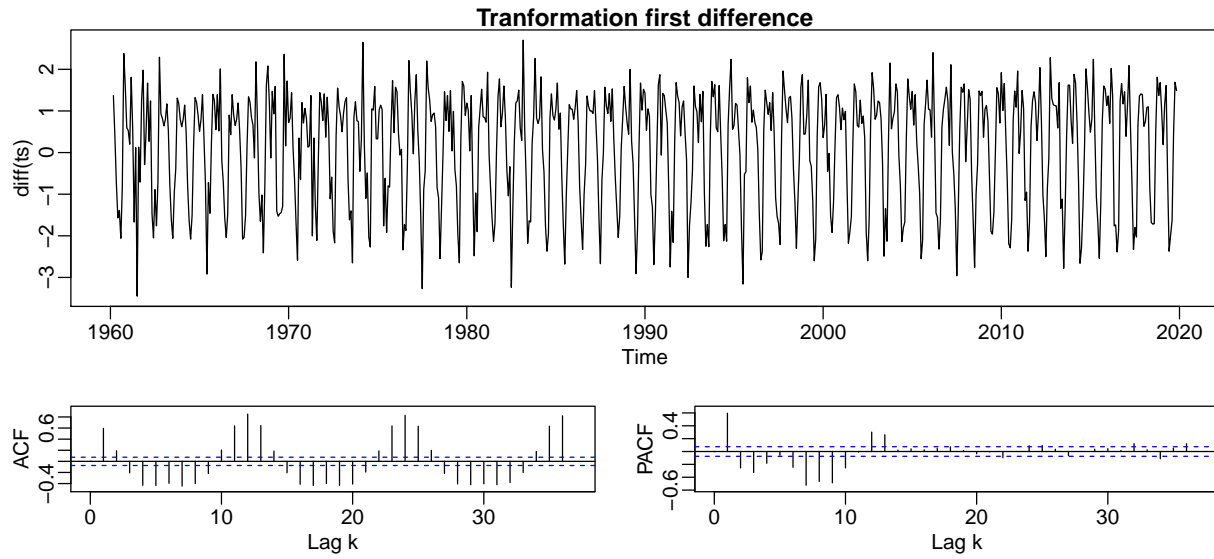
Does not seem to get any better.

We try capturing the seasonal trend. We expect that a season last for a year, so 12 months seems like a good starting point for differencing. Applying ∇^{12} we get:

Tranformation 12th−difference
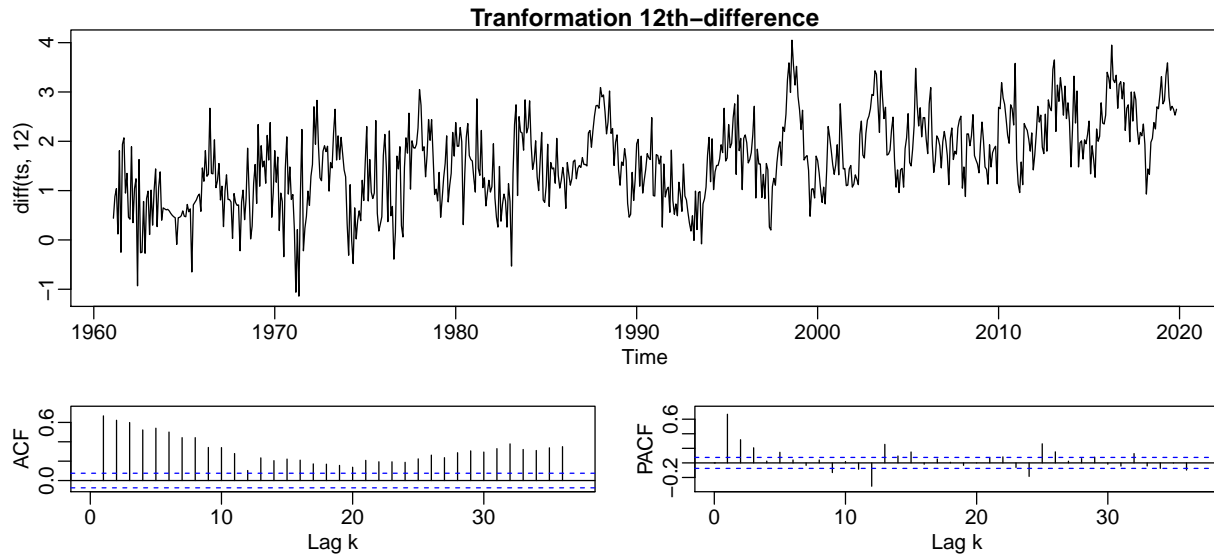
With this it seems we have removed the seasonal component. But it is clearly not completly stationary. We try applying $\nabla\nabla^{12}$ to capture the lag:

Tranformation 12th−difference
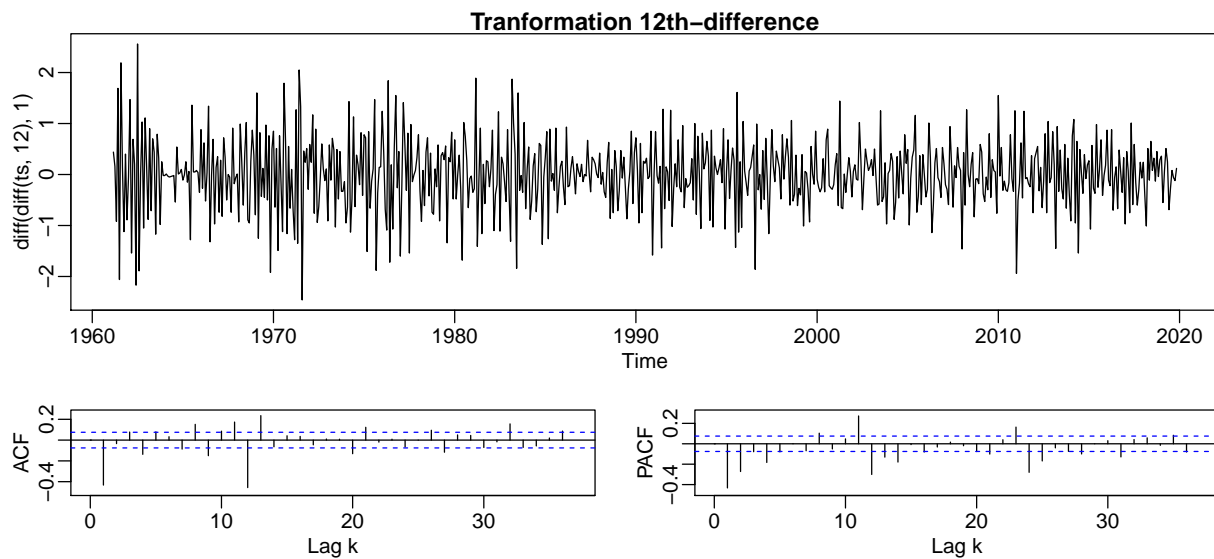
Which seems somewhat stationary, we could increase the lag to get a better fit, but as we prefer simple models, and we want to avoid overfitting, we decide to use $\nabla\nabla^{12}$ to tranform.

We will later see that the acf plot improve form.

There parameters results in a SARIMA model.

(Coefficient can i.e. be fit using innovations algoritihm. But that might not be the method impletemented in the package).

To find the best choice of parameters we compare AIC and AICC of different models.

and $p = 4, d = 1, P = 5, Q = 1$ seems to be best choice of the remaining parameters, where we get AIC and AICcc values of 1.27

With this chose of parameters our model would be:

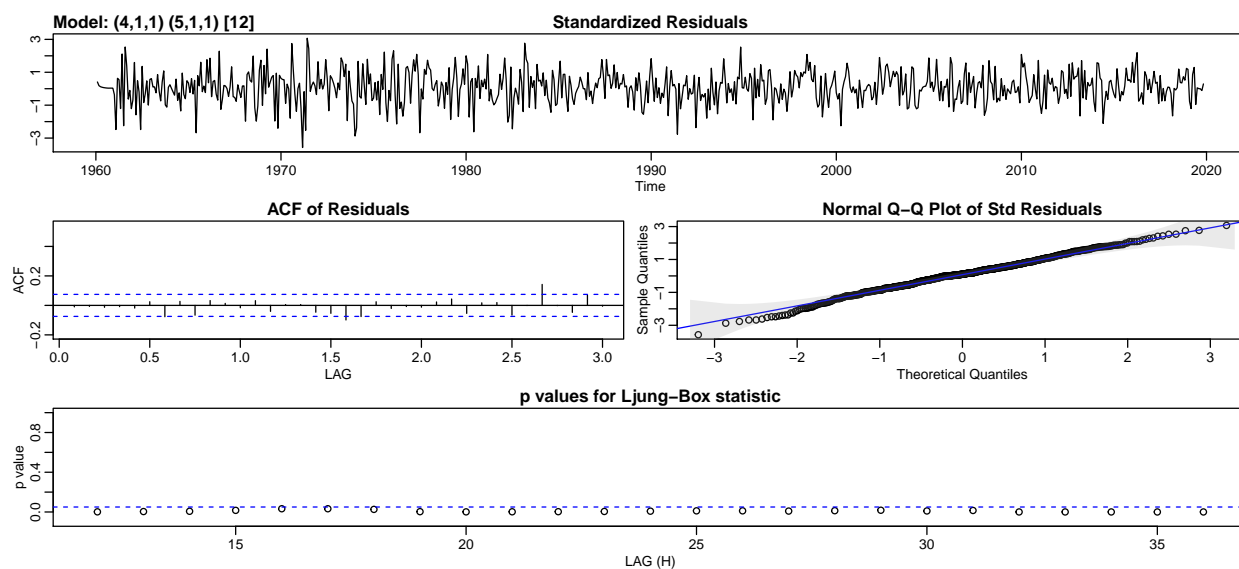$$\phi(B)\Phi(B)X_t = \theta(B)\Theta(B)Z_t$$

$$\theta(z) = 1 + \theta_1 z$$

$$\Theta(z) = 1 + \Theta_1 z$$

$$\phi(z) = 1 - \phi_1 z + \cdots + \phi_4 z^4$$

$$\Phi(z) = 1 - \Phi_1 z + \cdots + \Phi_5 z^5$$

Fitting the model with the selected parameters, we get the following diagnostic:

And the following coefficients:

```
##        phi1        phi2        phi3        phi4      theta1        Phi1
## -0.05851883 -0.06976959 -0.02473830 -0.11904962 -0.51345326 -0.01650475
##        Phi2        Phi3        Phi4        Phi5      Theta1
## -0.09926963  0.01414698 -0.11440507 -0.07581437 -0.82967466
```

The standardized residuals seems to be random, and there are not clear trend. The residuals seems homoscedastic and are centered at zero, which is a good indcation that our models assumptions are held by the data.

Looking at the autocorrolation plot, there does not seem to be any lag corrolating that much. Still, there are a few outliers, but that is often on 12+ months and can be considered to be well within the range of some family wise error rate.

Points in the qq-plot also seems to align with the normal line quite well, which indicate a good fit for our model.

All in all, the model seem to fit the data quite well.

Confidence interval for coefficients

We use bootstrapping to create confidence intervals for the coeffiecients. Using bootstraping we create a simulated data-set by drawing at random from our original data set, and fit a model using this. We repeat this, and in the end get a set of possible coefficient values, which we can use to confidence intervals. We will

use a a blocked bootstrap method. Blocked bootstraping is to select adjunct sections from the timeseries at random and join the selections together (Hyndman and Athanasopoulos 2018).

For the coefficients we get the following bootstrapped confidence intervals (95%) using 200 samples:

```
##                  phi1        phi2        phi3       phi4      theta1
## lower 2.5% -0.1822252 -0.28560066 -0.27060160 -0.2218933 -0.7799848
## upper 2.5%  0.4505167 -0.05617251 -0.03621915 -0.0521043 -0.2397842
##                  Phi1        Phi2        Phi3        Phi4        Phi5
## lower 2.5% -0.1527267 -0.17364720 -0.12560727 -0.11921476 -0.10586522
## upper 2.5%  0.0596933  0.05461404  0.07821557  0.05451245  0.04177681
##              Theta1
## lower 2.5% -0.9480253
## upper 2.5% -0.7909041
```

Forecasting with bootstrapping

Bootstrapping can also be used to obtain prediction interval as the ones gained with traditional methods is often too narrow.(Hyndman and Athanasopoulos 2018). We again used blocked bootstrapping.

The methods used in this section is adapted from Hyndman and Athanasopoulos's bootstrapping and bagging chapter.

We the fit different models, on 200 bootstrap samples:

Modelled flask CO2 levels

Modelled flask CO2 levels
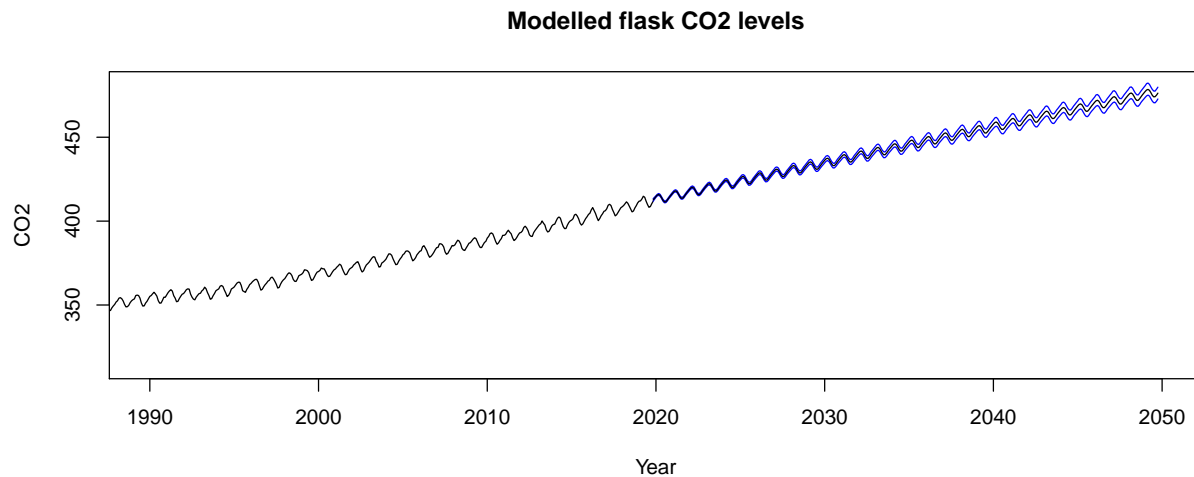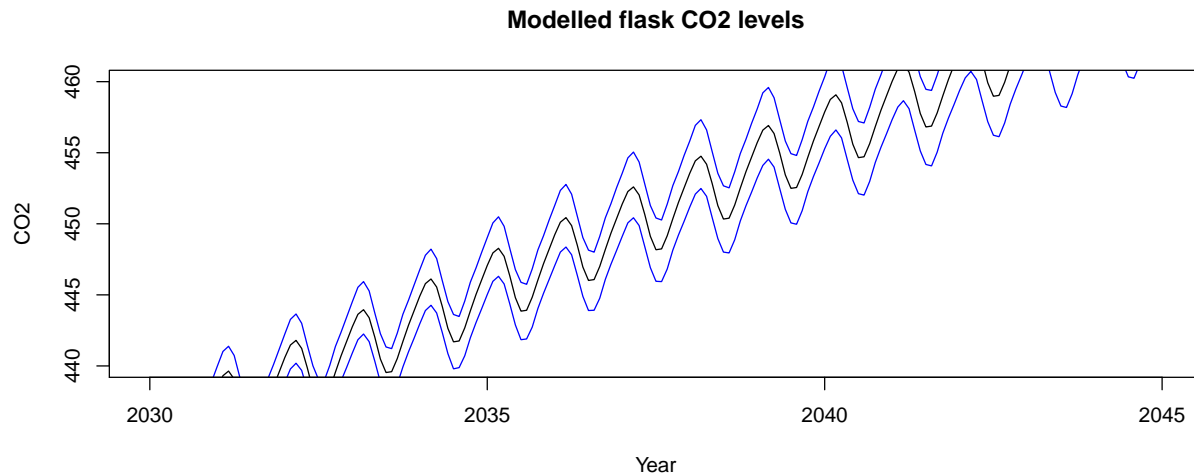
Discussion and Conclusion

We chose differencing over detrending to remove trend elements in our data as our goal is to coerce the data to stationarity for the purposes of fitting a model (2017 Time Series analysis and Its applications P.56)

Since our data was non-stationary, an ARIMA model seemed appropriate. Our data displays a clear seasonal pattern. The earth's ecosystem in general, and weather in particular, are seasonal, and it is reasonable to assume that co2 levels at a measuring station is correlated with changing weather. Therefore, a model that takes this into account is suitable, leading to our choice of a SARIMA model. Selecting the parameters for our model is done by testing all combinations of parameter values up to 5 and then picking the one with the smallest AICC score. It would seem intuitive to some to set very high values for our paramters to take into account large numbers of past states, but this would lead to an overly complex model and overfitting. It is therefore prudent to explore parameters including only a few previous states in the model.

Our model seems quite robust when inspecting various statistics. The normal QQ plot has only slight deviations at the tails. The standardized residuals appear staionary, and the p values for the Ljung-Box statistic suggest a good fit. The earth's atmosphere is a large enough system that any changes in the trend of CO2 will be gradual, barring some major event. Therefore we would expect accurate predictions a few years into the futre, with decreasing certainty at the 5 years and further out in the future. This seems to be the case in our model.

Our models inidicate that CO2 levels will keep rising into the future, increasing by approximately 30ppm 10 years into the future from where the data ends. The rate of increase is approximately linear with some seasonal variation depending on which month of the year we are considering.

The model seem to predict an increase of the CO2 levels in the next few years, and the data seems to fit the model assumptions. If we follow the same development as today, and if we assume the model assumption to be true, we will at upper 97.5% in confidence interval reach 450ppm CO2 in 2035, and will at lower 0.025% reach 450ppm in 2038.

We can conclude from our analysis and model that co2 levels will keep increasing into the future at a rate similar to the current rate.

References

C. D. Keeling, S. C. Piper, R. B. Bacastow, M. Wahlen, T. P. Whorf, M. Heimann, and H. A. Meijer, *Exchanges of atmospheric CO2 and 13CO2 with the terrestrial biosphere and oceans from 1978 to 2000.* I.

Global aspects, SIO Reference Series , No. 01-06, Scripps Institution of Oceanography, San Diego, 88 pages, 2001.

Brockwell, Peter J., and Richard A. Davis. 2016. *Introduction to Time Series and Forecasting.* Springer.

Hyndman, Rob J., and George Athanasopoulos. 2018. *Forecasting: Principles and Practice.* OTexts.

Shumway, Robert H., and David S. Stoffer. 2017. *Time Series Analysis and Its Applications: With R Examples.* Springer.