Report: The Global Carbon Budget
NEKN81, Time series analysis
16th February, 2020

Meera Damaraju, Oskars Niks Mālnieks, Sebastian Mårtensson

Abstract
The aim of this report is to use CO_2 changes during the last 60 years and provide alternative estimates of the amount of anthropological carbon dioxide emission change in the atmosphere until 2100. Our forecasts are then compared to data on RCP2.6 and RCP8.5. We see that both of our estimates indicate a significant increase of carbon dioxide in the atmosphere, that exceeds CO_2 concentration in 2100 nearly two times of the highest RPC estimate, i.e., RCP8.5.

1. Introduction

Using data provided by The Global Carbon Project (GCP) we estimate the total atmospheric CO_2 concentration until 2100 and then compare it to the Representative Concentration Pathways (RCP) from the Intergovernmental Panel on Climate Change (IPCC). RCP2.6 and RCP8.5 consider global surface temperature scenarios of 1.5 and 4.5 celsius degrees warming above pre industrial level by the year 2100. (Meinshausen et al., 2011). The RCP is measured as Radiative Forcing Values where higher values reflect an increase in global temperature. The difference between these estimates is that in the RCP8.5 scenario CO_2 concentration increases exponentially and assumes that the concentration of CO_2 will continue to grow exponentially with high population growth, continued use of fossil fuels and decreasing land sinks (due to deforestation).

In RCP2.6 the accumulated emissions are concave and leads to CO_2 post-estimation concentrations returning to 360 ppm by 2300. This divergence is important given that there exists a positive feedback between CO_2 emissions and global surface temperature, which when paired with reduced sink rates of oceans and land (Bennedsen, et al. 2019), might lead to irreversible damage to the Earth's ecosystem. Forecasting the amount of emissions in the atmosphere, and thereby the degrees of warming, is therefore of great importance and relevance to us all.

2. Estimation process

In order to forecast into the year 2100 we must first formalize the CO_2 emission and absorption process to replicate Earth's carbon cycle. For that we use the GCP proposed concept of Global Carbon Budget (GCB) that shows the main channels of CO_2 emission and absorption process allowing us to quantitatively assess interaction between variables.

Measuring and forecasting these human made emissions and natural sinks enables us to assess whether and to what extent CO_2 concentration grows annually, possibly indicating that natural absorption processes cannot keep up with rising CO_2 levels.

The growth in atmospheric CO_2 can be estimated using Equation (1) in which the emissions and the sinks put and absorb CO_2 from the atmosphere. To allow for measurement errors and miscellaneous "extreme" events such as volcanic eruptions, an error term is also added in order for the identity of Equation (1) to hold.

$$G^{ATM} = E^{FE} + E^{LUC} - S^{LND} - S^{OCN} - \varepsilon_t \quad (1)$$

G^{ATM} is the growth in atmospheric carbon dioxide concentration;
E^{FE} is the carbon dioxide emissions from fossil fuels;
E^{LUC} is the carbon dioxide emissions from land use change;
S^{LND} is the estimated emissions absorbed by the biosphere;
S^{OCN} is the amount of carbon dioxide absorbed by the ocean;
ε_t is the budget imbalance (and is recorded as the variation, error in the model).

With the data from GCP, we can already observe that within Formula 1 the emission side is larger than the absorption side, leading to the loss of natural equilibrium and accumulation of carbon dioxide in the atmosphere.

To predict G^{ATM} we use two slightly different approaches. The first approach relies completely on already available G^{ATM} data which we use to predict future values and the other approach is based on estimating and forecasting the components of G^{ATM} . This implicitly means that we are simplifying the estimation in the first approach given that G^{ATM} consists of 4 different components, thus potentially disregarding information that could have been used to provide us with more accurate estimation. The first approach will be used as our baseline. We will later extend the forecast to all of the component parts, as given by equation (1) and forecast these 4 variables separately, summing them up and comparing with the baseline model and the RCP forecasts.

For both models we use the following method to arrive at our estimates. First, we both graphically, and using the Augmented Dickey Fuller (ADF) test, verify whether our data is stationary. If a variable is non-stationary, we introduce first-differences accordingly to achieve stationarity. Nonetheless, as it will be discussed further on, whenever we face ambiguity while using these tests, we rely on our previous experience and general rationale on whether the data should be differenced or not.

Secondly, as our main tool of estimation we use the autoregressive integrated moving average model - ARIMA (p,d,q). To determine the lags for the AR and MA parts we analyse the ACF, PACF graphs and then compare different Akaike Information Criteria (AIC) for models with different lags. This is done because ACF, PACF tests for complex models can be ambiguous which is why, after establishing our "base-model" value of AIC, we compare the tests´ values for lags above and below our initial ARIMA model. In addition, the AIC test could have been replaced by the BIC test, however, given AIC's simplicity and availability in our R package we opt for it. One might argue that the BIC test is better because it hinders overfitting of an estimation model. The amount of lags that we arrive at is, however, not indicative of overfitting.

Lastly, to evaluate the general fitness of our models we split existing G^{ATM} data into two parts: 80% that are used to estimate the ARIMA models and 20% for which we use ARIMA to compare the models´ accuracy between the true and predicted value from 2007 to 2018. We would like to point out that this allows us to continue this research beyond this report - we can carefully adjust our estimation window or create models that gradually includes all data, resulting in ARIMA that eventually uses 100% of available data.

2.1. Data description

We are given two data sets - the first one, as already mentioned, contains data on G^{ATM} and other GCB components for the time period 1959-2018 (some data for GCB components are available since 1750)[1]. There also are various alternative models given within this data, however, we completely rely on GCP estimates. Let's briefly analyze each component of GCB separately.

Our main variable of interest is G^{ATM}. The atmospheric growth rate is estimated directly from the atmospheric CO_2 concentration measurements. Initially time series are shown as stationary using ADF test, however, from Figure 1 we conclude that G^{ATM} is actually non-stationary due to the changing mean which is trending upwards. Taking the first difference of G^{ATM} makes the series stationary. For the graph of the differenced atmospheric growth, see Appendix A 2.1.1.

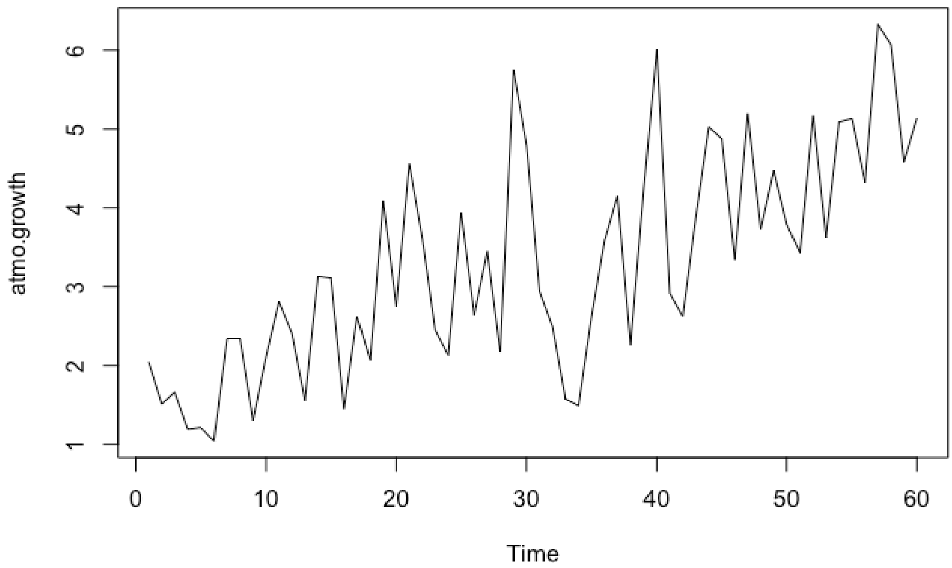


Figure 1. The atmospheric CO_2 growth rate[2]

Emissions from fossil fuel combustion and industrial processes are also trending, which is why we use first-difference to achieve stationarity (see A 2.1.2.). We take the first difference of all remaining variables for the same reason of the prevalence of a trend: emissions from land-use change, the ocean sink and the land sink (see A 2.1.3 - A 2.1.5 accordingly). It is worthwhile to note that some of these variables are not directly observed, and are estimates[3], thus potentially making forecast errors larger.

[1] All values provided in the data set comprise of gigatonnes of carbon or 3.664 billion tonnes of carbon. The data set itself comes from a conglomeration of different sources.
[2] Here and further on (if not indicated otherwise): **Time 0 = 1959.**
[3] The ocean sink is estimated as the average of several global ocean models that reproduce observed mean ocean sink of the 1990s. The land sink is estimated from various models that reproduce observed mean land sink in the 1990s.

Our second data set contains RCP2.6 and RCP8.5 estimates for the time period 2019-2100. Given that we do not use them for regressions, we will not analyze them further, but will briefly comment on what assumptions they rest on. RCP2.6 relies on the assumption that oil consumption declines whilst bio energy and carbon capture technology have an impact in reducing CO_2 in the atmosphere as an effect of stringent climate policies (van Vuuren et al. 2011). The pathway set out by RCP8.5 is on the opposite side of the spectrum and forecasts the upper end of CO_2 in the atmosphere as a result of no climate policy, high population growth and a continued reliance on fossil fuels. Both of the estimates provide the upper and lower end of the forecast spectrum.

One could argue that setting the forecasted value equal to the mean of the stationary process is enough for such a long time horizon, due to mean reversion of stationary process. However, it is interesting to see how the process develops - for example 40 years into the future. That is why we forecast using the ARIMA model and not just by setting the forecasted value equal to the mean of the growth.

2.2. Approach 1: Forecasting using Naive forecast

To form the base model with which we will make comparisons, we use a naive forecast in which atmospheric growth is estimated using only itself. In order to arrive at a plausible ARIMA specification, we observed the graphs of the Autocorrelation (ACF) and Partial Correlation functions (PACF) for the differenced G^{ATM} as shown in Figure 2. From the ACF we observe that the first two lags are significant, indicating that there possibly exists 2 lags in the MA component. For the PACF we observe 3 spikes as being significant which indicates that the Autoregressive component should consist of 3 lags. The additional significant lags such as at lag 6 for PACF seems spurious due to the lags dropping off up until that one particular point. It is unlikely that such outliers would add more precision and we therefore ignore them.

Our "base-model" (BM) then becomes an ARIMA (3,1,2), due to the process being non-stationary and have the autorrelations functions as exhibited in Figure 2.

To check the robustness of this model, we examine the AIC values when increasing and decreasing the lags of the AR component by 1. We found that the lowest AIC value belonged to ARIMA (2,1,2) which is only marginally better than the AIC value of BM. To not misspecify the model and omit variables we settle on the initial lags, as given by the ACF and PACF as discussed above.

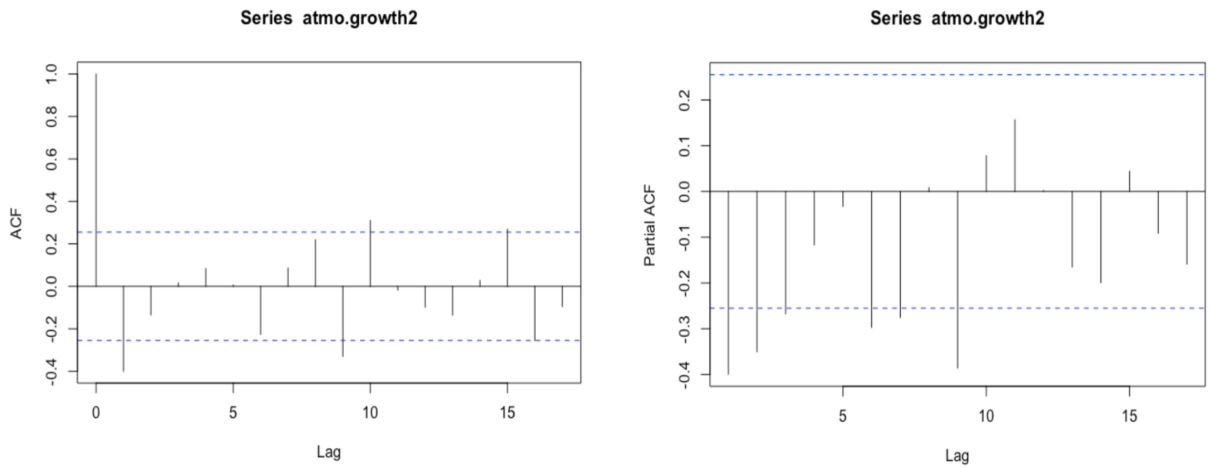


Figure 2. ACF and PACF test comparison for G^{ATM}

As can be seen from Figure 3, the predicted values of the BM model captures the trend that the actual data follow but does not capture the greater variance that the actual data exhibit.

Given that the forecast of interest is supposed to stretch over 80 years, deviations in year-to-yearly values should therefore be of little concern to our results. The values that stretch over longer horizons are more interesting and are the ones that our model captures better. Comparison of BM with RCP scenarios is included in part 2.4. Nevertheless, we should expect that general precision of our second approach should be more accurate given that predicting each value separately allows for greater adaptability and more attention to separate trends.

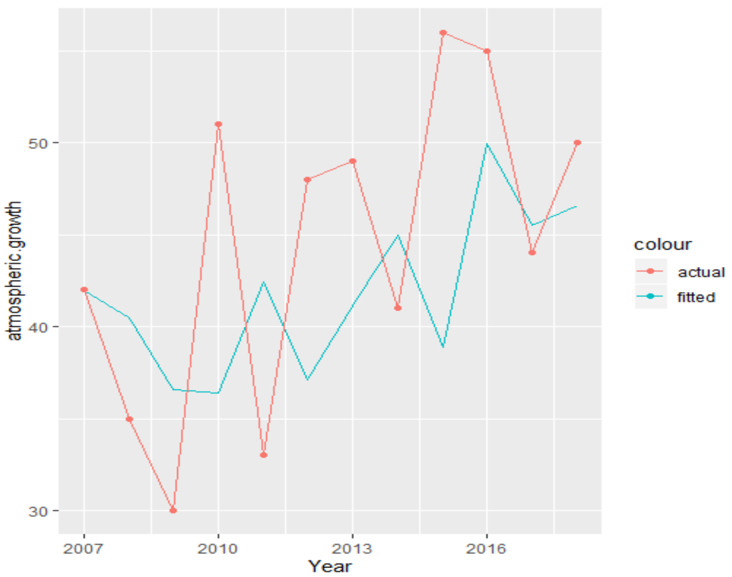


Figure 3. Comparison of BM and actual values between 2007 & 2018

2.3. Approach 2: Forecasting using Non-naive model

While the Naive forecast, regressed only using its own values, can make a plausible prediction, a non-naive forecast encompassing more elements, information and nuances may aid in producing a better forecast. This “non-naive” forecast is in turn composed of 4 separate processes which when combined form another way of forecasting CO_2 atmospheric growth. The models are estimated in the same way as the naive forecast, where we checked the stationarity and analyzed the plots of the ACF and PACF.

The estimated forecasts were then combined as in Equation (1) and converted into parts per million (ppmv) to forecast up to year 2100. Table 1 gives more detailed information on the lag length of each variable as well as the AIC value of the selected model.

Table 1 – Specifications of the regression models

VARIABLES	AR(p)	Difference	MA(q)	AIC
G^{ATM}	3	1	2	379
E^{FE}	5	1	1	162
ELUC	3	1	2	315
SLND	3	1	3	378
SOCN	2	1	0	252

From this we can observe that most of the variables required that the AR part to be larger or equal than 3 lags, indicating strong persistence and importance of historical values.

Moreover, when evaluating ARIMA model’s performance for each variable in our “20% interval” we notice that a similar pattern emerges as it happened with the Naive forecast - our estimates shows lower variation, yet equally well mimics the general trend movement in the long run (see Figures 4 and 5). “Ocean Sink” and “Land Use Change Emissions” show a strong upwards trend in this forecast window which is consistent with the growth from 1959.

The “Land Sink " in Figure 4, appears to be stationary, mostly due to the constant mean, which goes against our finding of non-stationarity. Looking at the entire sample from 1959 to 2018 we found however that the process was non-stationary based upon the plot and the ADF test. The strong upwards trend of the “Fossil fuel emissions” in Figure 5 is unrivaled by the other stronger upwards tending processes.

Figure 4. Land sink (left) and Ocean sink vs actual values between 2007 & 2018. Measured in gigatonnes of carbon in the atmosphere.

Figure 5. Land use emissions (left) and Fossil fuel emissions vs actual values between 2007 & 2018. Measured in gigatonnes of carbon in the atmosphere.

2.4. Results: 2019-2100

The results from the Naive and the Non-Naive models are presented in Figure 6 together with the highest and lowest estimate of the RCP forecasts. They were converted into Radiative Forcing Values using a conversion rate of 1 ppmv to 2.127 gigatons of carbon. As can be seen, the naive forecast is very simple in that it is only a straight line which is not surprising given the long time horizon and the assumptions behind the construction of the model. The non-naive forecast more closely resembles an exponentially growing process and forecasts lower values compared with the naive-forecast. The non-naive forecast loosely follows the forecasts made by the IPCC up until 2050, from which it starts to deviate and grows exponentially. The trend in both forecasts is not very surprising given that they both take a historical approach to emissions and does not take into account a decline in population growth, greener technologies or any other mechanisms that affect emissions and sinks.

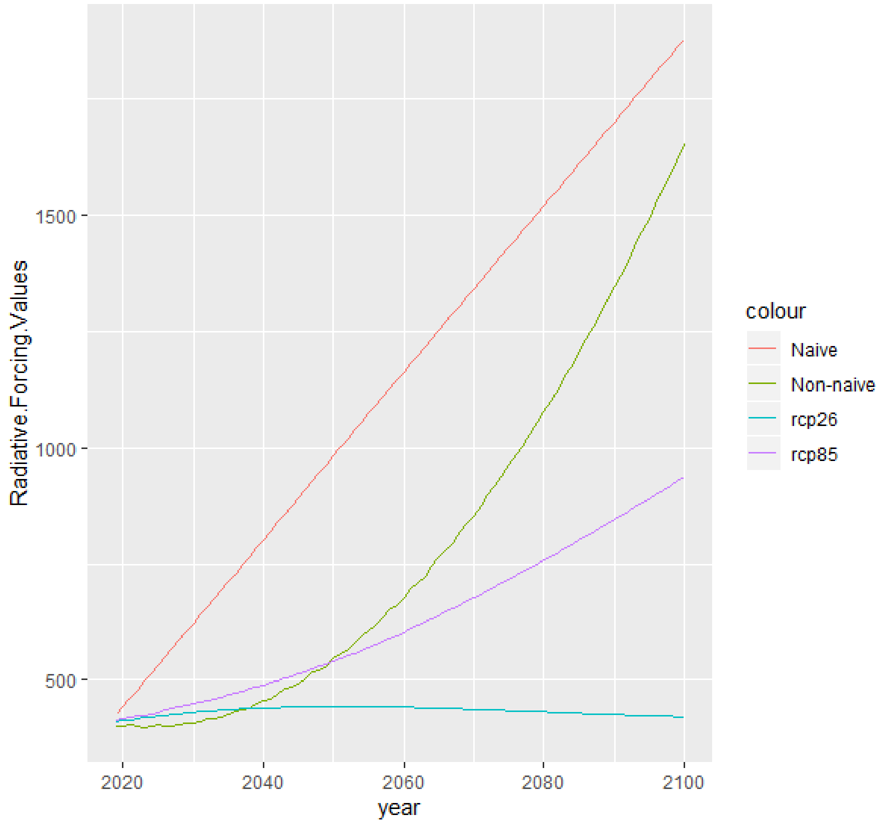


Figure 6. Our estimate comparison with RCP2.6 and RCP8.5 2019-2100

This result of our estimates could be mainly related to the fact that such complex phenomena can hardly be explained by these simple models and due to simultaneous feedback between variables G^{ATM} can experience steep increases within a short time.

3. Conclusions

Using only historical data on sources of emission and absorption of CO_2, we forecast the Radiative Forcing Values into the year 2100. Based on Figure 6 we conclude that if atmospheric carbon dioxide continues to increase, as it has done in the past, then it will overshoot the highest estimate produced by the IPCC. Our estimates should be seen as a maximum given the assumptions under which the models were formed, namely that we continue to burn fossil fuels at the current (or increasing) rates, population increases and have technologies of similar efficiency. The difference between the Naive forecast, in which only information from the actual level of atmospheric growth was used, and the non-naive forecast is striking not only for this application but for forecasting in general. Regressing on many more variables, and thereby including more information, creates more nuanced forecasts. Our result should therefore be seen as an approach in which human emissions stay roughly the same and our society develops as it has been for the last 60 years.

Works Cited

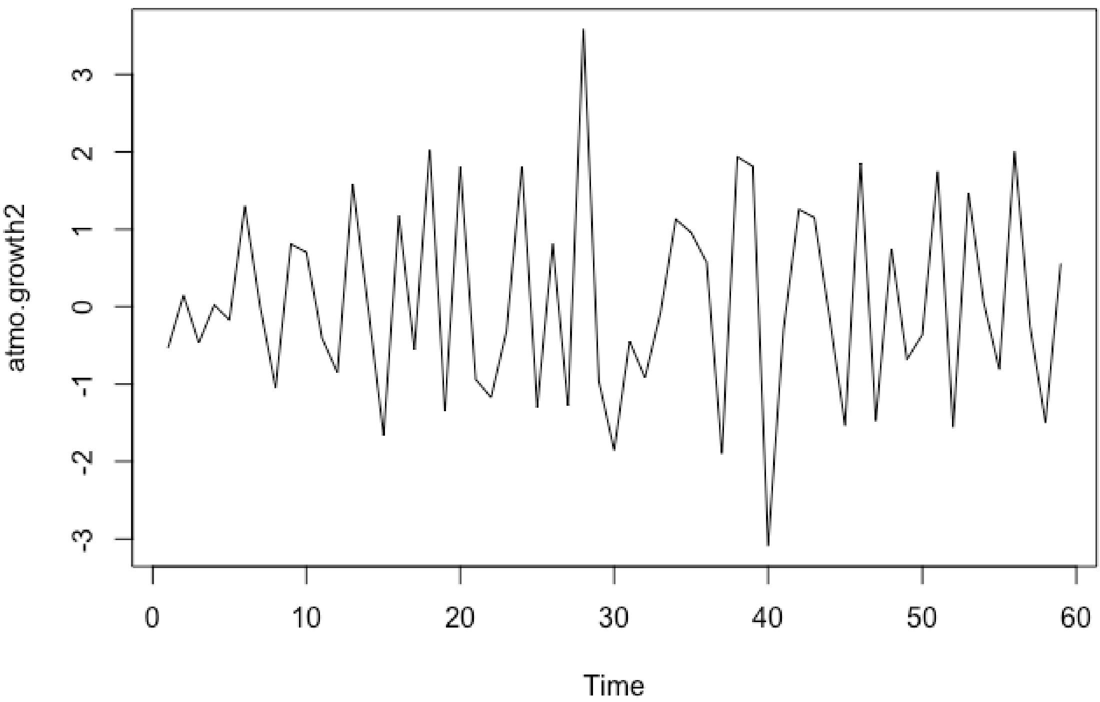
Bennedsen, Mikkel, et al. "Trend Analysis of the Airborne Fraction and Sink Rate of Anthropogenically Released CO2." *Biogeosciences*, vol. 16, no. 18, 2019, pp. 3651–3663., doi:10.5194/bg-16-3651-2019.

Meinshausen, Malte, et al. "The RCP Greenhouse Gas Concentrations and Their Extensions from 1765 to 2300." *Climatic Change*, vol. 109, no. 1-2, 2011, pp. 213–241., doi:10.1007/s10584-011-0156-z.

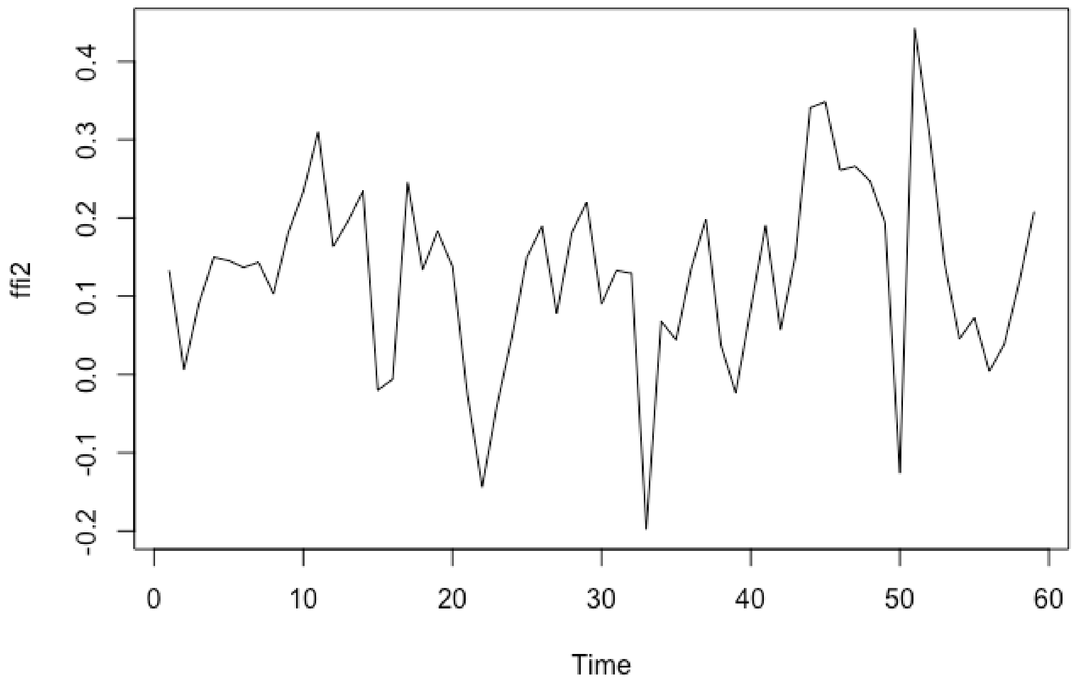
van Vuuren, Detlef P. et al "The representative concentration pathways: an overview" *Climate Change*, vol 109, no. 5-31, 2011, pp 4-31.. DOI 10.1007/s10584-011-0148-z

Appendix

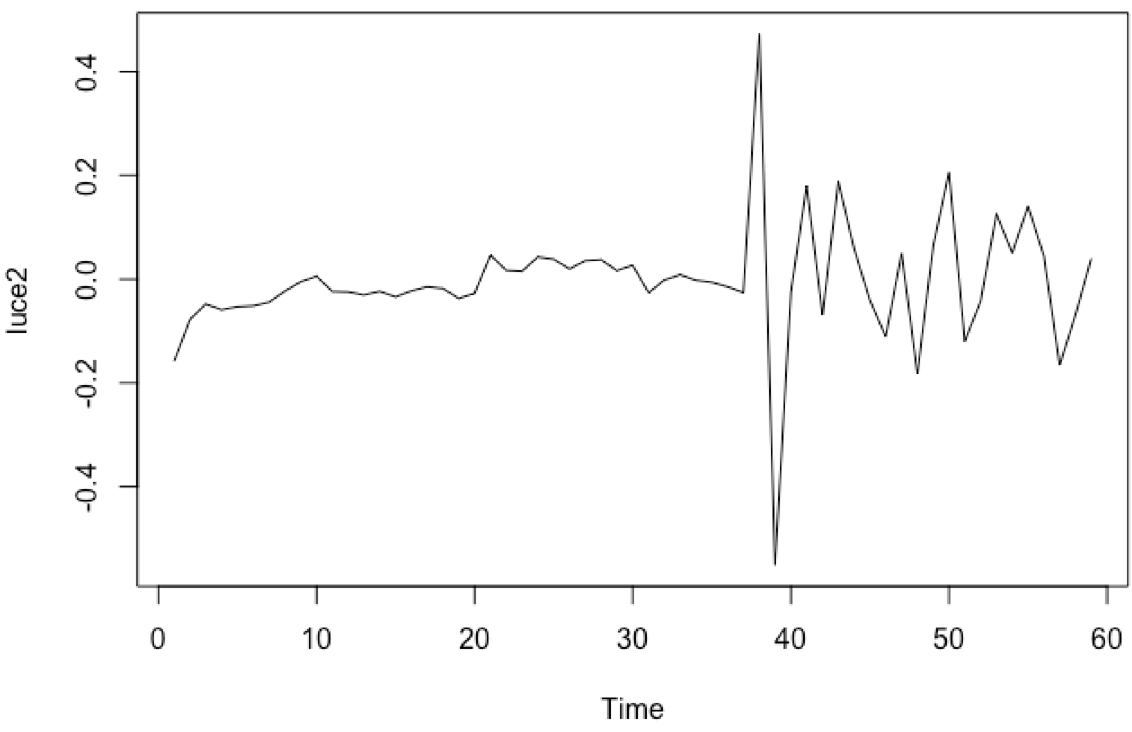
A 2.1.1. The differenced atmospheric CO_2 growth rate



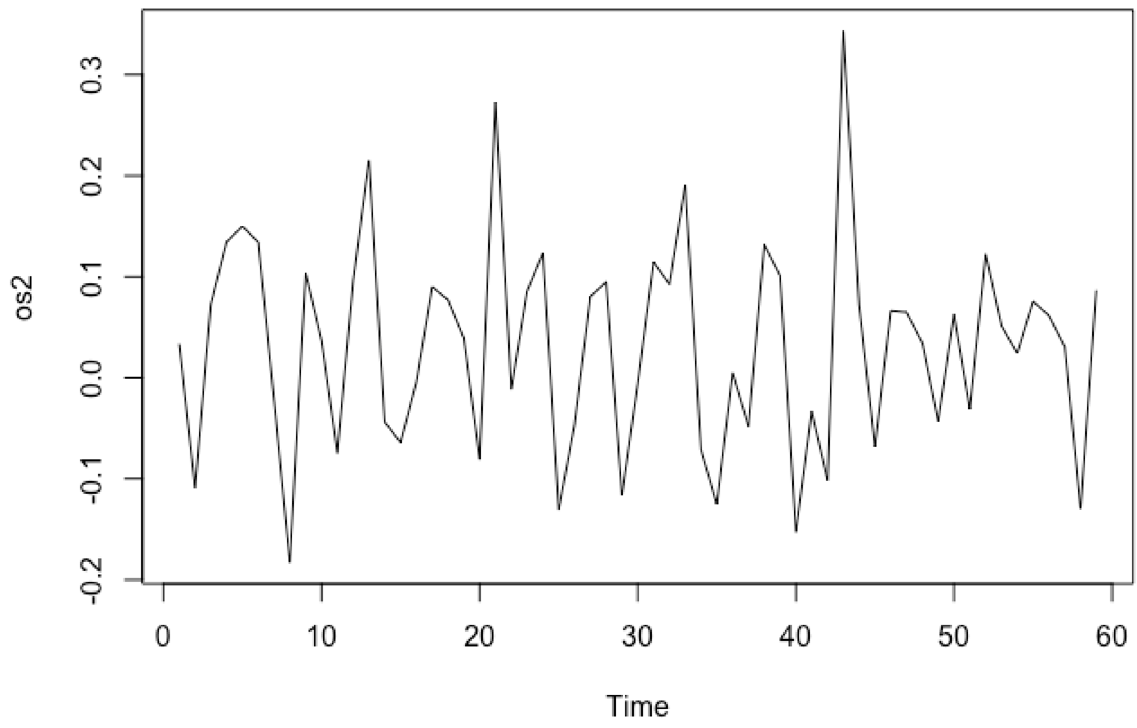
A 2.1.2. The differenced emissions from fossil fuel combustion and industrial processes



A 2.1.3. The differenced emissions from land-use change



A 2.1.4. The differenced ocean sink



A 2.1.5. The differenced land sink