Forecasting CO2 Atmospheric Concentration using a State Space Model

Team 25

Abstract

Climate change is nowadays a key topic of interest in political discussions. It is crucial to have proper models and perform accurate forecasts of climate-related variables to efficiently tackle the issues that climate change brings. This paper focuses on the forecast of carbon dioxide atmospheric concentration. We propose a state space approach that takes into account the measurement errors to which climate data is subject to. The model is based on the Global Carbon Budget Equation and allows to jointly estimate and forecast all the unobserved components driving the observed variables of the Budget Equation and CO2 atmospheric concentrations. We compare our forecast from 2019 to 2100 to Representative Concentration Pathways (RCP) that correspond to distinct scenarios leading to different temperature increases above pre-industrial levels. We find that if all variables keep on following current trends, the best and worst RCP scenarios are less likely to be achieved. We further conduct a simulation study where we let the level of CO2 emissions drops to its 1959 level for future periods, we find that our forecasts are tilted toward the best scenario. The contrary happens if we let CO2 emissions explode. These results stress the fact that there is a serious need for new global climate policies in order to save our planet.

Keywords: Global Carbon Budget Equation, CO2 Atmospheric Concentration, State Space Models, Forecasting.

Contents

1 Introduction

Over the last two decades, the Carbon Budget has been a problem of great interest, especially after the Earth Summit in Rio de Janeiro in 1992, where the member states of the United Nations decided to stick together and to handle the conflict relating to sustainability and global warming together. Awareness regarding climate change increased and it became clear that the need for understanding the relationship between carbon dioxide (hereafter CO2) emissions, land and ocean sinks and CO2 atmospheric growth is critical. The problems are again emphasized at the COP21 in Paris. The main result from these climate discussions is that we globally need to limit the temperature rise to a maximum of 2 degrees above pre-industrial level. This was formally captured in the Paris Agreement. To sufficiently tackle the climate problem, new policies need to be established and policy makers are interested in what these policies need to include to achieve a specific future goal. Hence, not only is it essential to understand historical and recent budgets, it is also necessary to perform reliable forecasts to compare potential future scenarios of various environmental variables.

It is necessary to have a proper understanding of the carbon cycle and to have accurate measurements of CO2 emissions and their redistribution among the atmosphere, ocean, and terrestrial biospheres. It can directly be concluded from the carbon cycle that the amount of emitted CO2 that is not absorbed by the land or the ocean sinks must end up in the atmosphere. Hence, the equilibrium condition is a such,

$$G_t = E_t^{FF} + E_t^{LU} - S_t^{LD} - S_t^{OC},$$

where G_t is the growth in atmospheric CO2 concentration, E_t^{FF} are CO2 emissions from fossil fuels combustion and industrial processes, E_t^{LU} are the CO2 emissions from land-use change (e.g. deforestation), S_t^{LD} is how much of these emissions are absorbed by the terrestrial biosphere and S_t^{OC} is the amount absorbed by oceans (Le Quéré et al., 2018). However, each variable is measured from different environmental models and the construction of each variable is hence subject to measurement errors. In practice, this leads to a significant discrepancy between the right and left hand sides of the identity, implying a disequilibrium in the carbon cycle. The Global Carbon Project aims at quantifying this discrepancy (denoted by ε_t) by means of the so called Global Carbon Budget Equation,

$$G_t = E_t^{FF} + E_t^{LU} - S_t^{LD} - S_t^{OC} + \varepsilon_t. \tag{1}$$

It is however impossible to disentangle whether a positive imbalance for instance is induced by an overestimation of emissions, an underestimation of the sinks absorption or an overestimation of the growth of atmospheric CO2 concentration.

The rise of interdisciplinary techniques and mostly the use of econometric methods have been of interest in the recent years as similar problems have been observed in economic applications. The aim of this paper is to model and forecast the concentration of CO2 in the atmosphere by means of an econometric approach. We propose a state space model for this purpose. It allows to jointly estimate and forecast the state variables of all the observable variables that appear in the Carbon Budget Equation and more. We indeed propose a model specification which allows to estimate and forecast directly also the unobserved components of CO2 concentration, although it does not appear directly in the Carbon Budget Equation. Moreover, the model has the flexibility of imposing the restrictions defined by the Carbon Budget Equation.

The paper is organized as follows, Section 2 highlights the motivations and the necessity of an econometric model for the Carbon Budget. Then, it describes the data set used in this research. Section 4 presents the construction of the state-space model based on the Global Carbon Budget. Estimation of the model between 1959 and 2017 as well as forecasts up to 2100 are performed in Section 5. Section 6 discusses the limitations and the possible further research that this paper sheds light on, before concluding in Section 7.

2 The Global Carbon Budget

The Global Carbon Budget states that all CO2 emissions must either end up in the Earth's sinks or in the atmosphere, while accounting for potential measurement errors. The traditional approach to model the growth of atmospheric CO2 concentration has been related to the field of physics and climatology (Le Quéré et al., 2018). Each variable in the Budget Equation (1) is measured with distinct – usually – climatology models and the imbalance is then simply constructed as the discrepancy between the measured atmospheric CO2 concentration growth and the one implied by the difference between emissions and Earth absorption.

Measurements errors associated with CO2 measurements and emission estimates still limit our confidence in calculating net carbon uptake from the atmosphere by the land and ocean (Ballantyne et al., 2015). As we enter into an era in which scientists are expected to provide an increasingly more detailed assessment of carbon concentrations in the atmosphere at increasingly higher spatial and temporal resolutions (Canadell et al., 2011), it is critical that we develop a framework less influenced by these measurement errors. Therefore, instead of solely relying on the budget equation and estimating each series separately to approximate the imbalance, the equation now builds the backbone of the statistical models used to predict the concentration level of CO2. Several approaches have been undertaken recently in the field to model the CO2 concentration atmospheric growth. Strassmann and Joos (2018) suggest a simple climate model, known as Bern Simple-Climate model (BernSCM), in order to capture the carbon cycle appropriately and measure the long-term effect of humans in the Earth's Budget Balance but they do not tackle the issue of measurement errors. The model links a climate component with an energy-economy model to simulate the emissions and corresponding consequences for the climate. Bennedsen et al. (nd) on the other hand recognize the problem of measurement errors that occurs with estimating the carbon concentration by modelling the airborne fraction and sink rate of CO2 released by human force in a state space model. Such models are based on the assumption that the variable of interest is mismeasured and that it is therefore unobserved. The authors present several ways to measure the impact of human behavior in the Earth's energy budget by using the budget equation as a part of the model construction.

Another concern with respect to modelling the global carbon cycle is the potential feedback effect between CO2 atmospheric concentration and climate change. Indeed, we expect an increase in CO2 concentration to lead to an increase in temperature, which may then lead to a reduction in land and ocean absorption (more arid soils or more water evaporation for instance). In the literature, this topic is widely discussed among scientists. However, no agreement has been found about the existence or the magnitude of such feedback effect. Among others, Friedlingstein (2015) compares eleven coupled climate-carbon models and while they find significant feedback effects, no consensus is found regarding the magnitude of such feedback or even to what it is attributed. Strassmann and Joos (2018) find that the feedback range includes zero and, hence, no statistical evidence is provided that makes it necessary to include it. Gloor et al. (2010) also point out that no statistically significant feedback effect can be found. For the sake of simplicity and from the lack of consensus regarding this feedback effect, we will assume that it is non-existent and will therefore omit it in the model construction.

The aim of this paper is to construct an econometric model to forecast the level of CO2 atmospheric concentration. To tackle the issue of measurement errors we will employ a state-space model inspired by Bennedsen et al. (nd).

3 Data

For this research we use a data set provided by http://www.icos-cp.eu/GCP/2018 which contains yearly time series data from 1959 to 2017, amounting to 59 observations. The data objects are anthropogenic carbon emissions from fossil fuels and cement production emissions (E^{FF}) and from land-use change, mostly deforestation and afforestation (E^{LU}) , different estimates of how much of these emissions are absorbed by the terrestrial biosphere (land sink, S^{LD}), and different estimates of the amount absorbed by the ocean (ocean sink, S^{OC}). By necessity, all emissions not absorbed by Earth's carbon sinks must end up in the atmosphere, and so the final object of interest is the growth in atmospheric concentrations (G). All observables are measured in billion tonnes of carbon per year (GtC/yr). For this research we will consider the combined variables emission ($E^{FF} + E^{LU}$) and sink ($S^{LD} + S^{OC}$). The reasoning behind this will be further elaborated in the next section. The plots of atmospheric growth, emission and sink in levels and in first differences are given in the Appendix. Furthermore a summary statistics table for these variables is given in the appendix.

Next, we continue the analysis by formally testing for unit roots in the series. In particular, we want to identify the order of integration of our series. In this context, one should address the concept of the Pantula Principle, which is especially relevant for the Augmented Dickey-Fuller (ADF) test. One should difference the series as many times as it deems appropriate for making the series stationary, which we specify to be $d = 2$. We test for a unit root in the differenced series. If the null of a unit root is rejected, we decrease d by one and repeat the unit root test. The procedure is stopped when the test cannot be rejected anymore. The order of integration is then assumed to be $I(d + 1)$. This procedure ensures that the differenced lags included in the ADF test are stationary. The following table depicts the results from the ADF test, where the null hypothesis is non-stationarity and the alternative hypothesis is stationarity. From this we can conclude that the emission and sink series are integrated of order one and that atmospheric growth is stationary in levels. However, the last result is not what we expect from the plotted time series. This can be due to the small sample size, since we only have 59 observations. Therefore, we rely on what the graphs tell us and conclude that all series are integrated of order one.

	Atmospheric Growth	**Emission**	**Sink**
2nd difference	0.01	0.01	0.01
1st difference	0.01	0.02831	0.01
Levels	0.03667	0.5422	0.06508

Table 1: p-values for the Augmented Dickey-Fuller test for stationarity

Since the CO2 atmospheric concentration is the variable of interest, we construct this variable in the following way,

$$C_t = C_{1959} + \sum_{\tau=1}^{t} G_\tau, \tag{2}$$

where C_t is measured in GtC/yr.

The order of integration of the variables will therefore be used in the construction of the model in the upcoming section.

4 A state space model approach

State space models allow to model variables of interest that are subject to measurement errors, which is the case for the Global Carbon Budget Equation, as the budget imbalance is generally different from zero. A state space model for the airbone fraction and the sink rate has already been proposed in Bennedsen et al. (nd). They model the trends of the above-mentioned variables in two different state space models. We build on their approach by proposing a single state space model that exploits the restrictions of the Global Carbon Budget Equation and jointly estimates the unobserved components of all the variables that appear in the equation. Moreover, our model allows to jointly estimate the state variables of both the atmospheric concentration of CO2 and its first difference. State space models take the following form:

$$y_t = Z\mu_t + \epsilon_t \tag{3}$$

$$\mu_t = T\mu_{t-1} + M\eta_t \tag{4}$$

Equation (3) represents the observation equation, that is, we observe variable y but the variable of interest is its state, μ, which is not observed. The error term in the observation equation hence represents the measurement error in the observed variable. The transition equation (4), represents the transition process of the state variable of interest, which in this example depends on its own lag.

The available sample size and the order of integration of each series limit estimations when the number of parameters becomes too large. Since this paper focuses on the effects of emission paths on the CO2 atmospheric concentration and does not make the distinction between the two sources of emission and the two sources of absorption, we make the following reduction of the model.

$$G_t = E_t - S_t + \varepsilon_t,$$

where $E_t = E_t^{FF} + E_t^{LU}$ and $S_t = S_t^{LD} + S_t^{OC}$ and $G_t = C_t - C_{t-1}$. Nonetheless, it is possible to extend our state space representation in order to model all of these four variables separately. As mentioned, the four variables of interest are subject to measurement errors. Furthermore, the CO2 atmospheric concentration growth (G_t) can on the one hand be measured from environmental models, but can also be approximated by the difference between emissions and absorption ($E_t - S_t$). That is, with different measurement errors, G_t and ($E_t - S_t$) should have the same state variable in the observation equations. Following the structure of Equation (3) and the assumptions made, our set of observation equations of interest is as follows,

$$\begin{bmatrix} C_t \\ G_t \\ E_t - S_t \\ E_t \\ S_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t^C \\ \mu_t^E \\ \mu_t^S \end{bmatrix} + \epsilon_t, \tag{5}$$

where ϵ_t is a vector stacking the five error terms of the observation equations, they are assumed independently normally distributed with mean zero and distinct variance: $\epsilon_t \sim N(0, H)$, $H = \text{diag}(\sigma^2_{C,\epsilon}, \sigma^2_{G,\epsilon}, \sigma^2_{E-S,\epsilon}, \sigma^2_{E,\epsilon}, \sigma^2_{S,\epsilon})$. The CO2 atmospheric concentration variable is *I(2)* and its first difference is equal to G_t, which is *I(1)* and depends on the state of E_t and S_t which are themselves *I(1)*. Hence, the state variables are assumed to be integrated of the same order as their corresponding observed variables and the transition equations of the three states of interest are as follows,

$$\begin{bmatrix} \mu_t^C \\ \mu_t^E \\ \mu_t^S \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1}^C \\ \mu_{t-1}^E \\ \mu_{t-1}^S \end{bmatrix} + \eta_t. \tag{6}$$

The three error terms η_t are also assumed independently normally distributed with mean zero and distinct variance: $\eta_t \sim N(0, Q)$, $Q = \text{diag}(\sigma^2_{C,\eta}, \sigma^2_{E,\eta}, \sigma^2_{S,\eta})$. M is equal to the identity matrix.

The hyperparameters of the model described above, i.e., all the variances of the error terms, are estimate by maximizing the following diffuse log-likelihood:

$$\ell_d = -\frac{Tn}{2}\log(2\pi) - \frac{1}{2}\sum_{t=d+1}^{T} \log\left(|F_t|\right) - \frac{1}{2} v_t' F_t^{-1} v_t,$$

where d is the number of non-stationary variables, and v_t and F_t are, respectively, the prediction errors and their covariance matrix. The latter are estimated, together with the state variables and their covariance matrices, via the Kalman filter recursions

$$\begin{aligned} v_t &= y_t - Z a_t \\ F_t &= Z P_t Z' + H \\ K_t &= T P_t Z' F_t^{-1} \\ a_{t|t} &= a_t + P_t Z' F_t^{-1} v_t \end{aligned} \qquad \begin{aligned} P_{t|t} &= P_t - P_t Z' F_t^{-1} Z P_t \\ a_{t+1} &= T a_t + K_t v_t \\ P_{t+1} &= T P_t \left(T - K_t Z\right)' + Q, \end{aligned}$$

$t = 1, \dots, T$. We use a diffuse initialization for all the state variables, as all of them are non-stationary (Durbin and Koopman, 2012).

5 Results

5.1 Estimation

We estimate our proposed state space model for the sample period starting in 1959 and ending in 2017. Figures 1 and 2 report, respectively, the filtered estimates of the state variables μ_t^C for C_t and $\mu_t^E - \mu_t^S$ of G_t, together with their 95% confidence intervals, which are constructed using the estimates for $P_{t|t}$ from the Kalman filter recursions. The figures show that the adequacy of the restriction on the state variable of G_t being equal to the one of $E_t - S_t$, as the estimated local trend does not deviate much from the observed series G_t. On the contrary, the occasional deviations of the estimated trend from $E_t - S_t$ are due to a deterioration of the model when imposing a common trend (Bennedsen et al., nd). See Figures 13-15 in the Appendix for the filtered estimates of the remaining state variables of the model, and Table 4 for the maximum likelihood estimates of the hyperparameters. We do not venture into diagnostic checking as our main interest relies on the forecast of the state variables, rather than on a correct model specification. For the same reason we do not use Kalman filtering instead of smoothing.

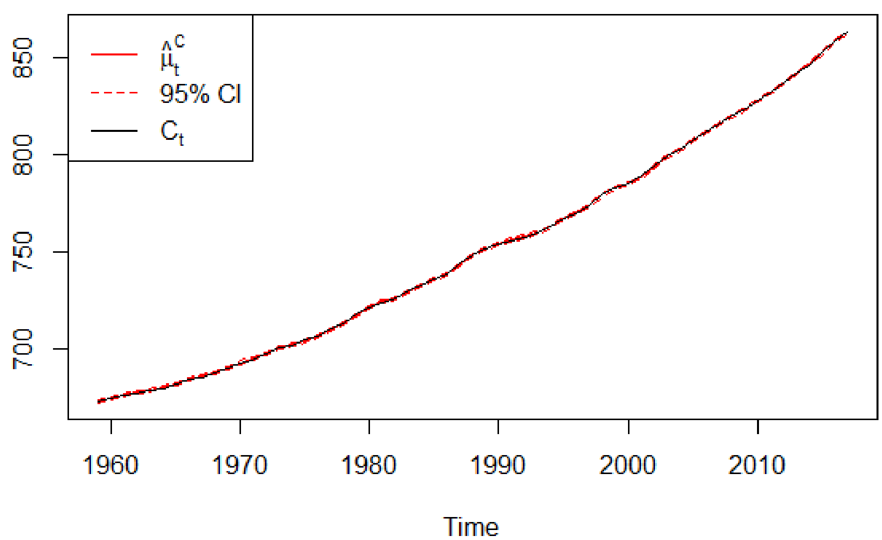


Figure 1: In-sample estimates of the state variable μ_t^C of C_t in GtC/yr, together with their 95% confidence intervals.

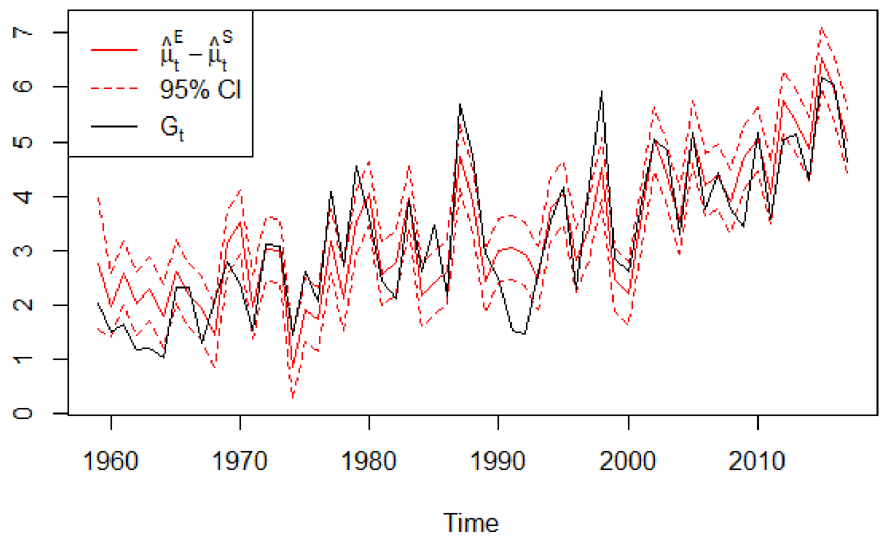


Figure 2: In-sample estimates of the state variable $\mu_t^E - \mu_t^S$ of G_t in GtC/yr, together with their 95% confidence intervals.

5.2 Forecast

To evaluate the accuracy of our constructed model, we perform in-sample and out-of-sample forecasts. In presence of missing observations, the Kalman filter is carried out by imposing $K_t = 0$ and $v_t = 0$. We based our forecast evaluation on the accordingly forecasted state variables and build prediction intervals using the forecasted variances of the state variables. Figure 3 depicts the one-step step ahead forecasts with expanding window. We include both the 50% and 99% confidence intervals. The width of the confidence intervals does not increase as for each one-step ahead forecast, we update the information that we have at the current time.

Such investigation is however not really representative of the forecast ability of the model for longer horizon. We then performed 56 (as we used a diffuse initialization for the estimation of the model and therefore had to exclude as many starting observations as many non-stationary state variables appear in the model, namely 3) in-sample recursive one-step ahead forecasts where each forecast point now depends on past forecasted points and not on realised values. Figure 4 depicts the recursively obtained 56 points ahead forecast. We again include both the 99% and 50% confidence intervals. While the confidence intervals significantly diverge as the forecast horizon increases, the predicted series accurately approximate the observed series, which lies within the 50% confidence interval.

However, models may perform well with in-sample forecasts but poorly with out-of-sample. Yet, the lack of data points does not allow us to evaluate its performance out-of-sample. We nonetheless perform 83 recursively obtained one-step ahead forecasts, from 2018 to 2100. Figure 5

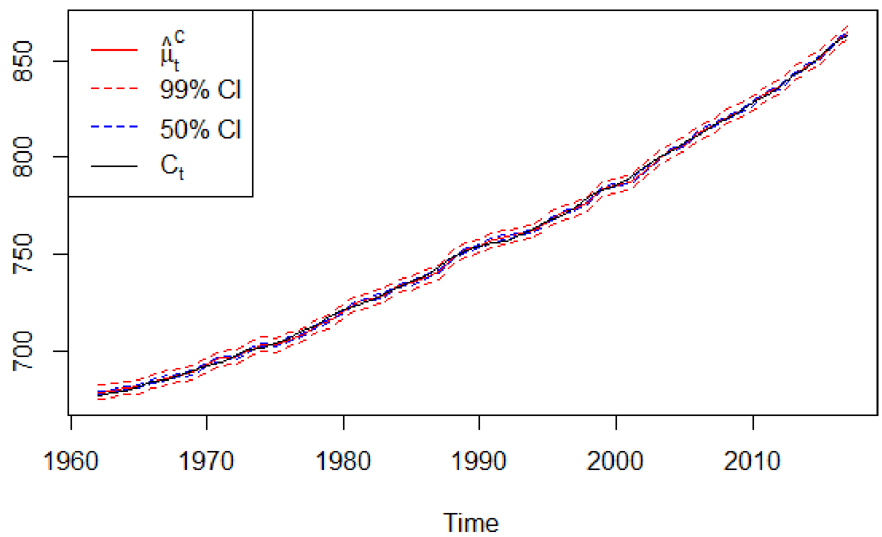


Figure 3: One step-ahead forecasts of μ_t^C of C_t in GtC/yr, together with their 99% and 50% confidence intervals.

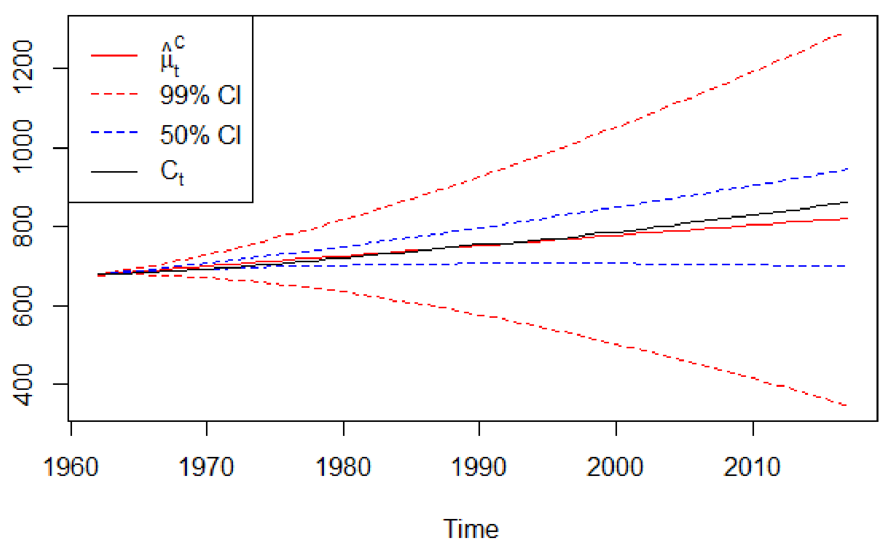


Figure 4: 56 step-ahead forecasts of μ_t^C of C_t in GtC/yr, together with their 99% and 50% confidence intervals, using the observed sample period.

shows the CO2 atmospheric concentration forecast with both the 99% and 50% confidence intervals. Due to the significantly far horizon and the recursively dependent forecasts, the confidence intervals significantly widen, leading to a possible CO2 atmospheric concentration between 500 and 2000 GtC/yr in 2100 at the 99% confidence interval.

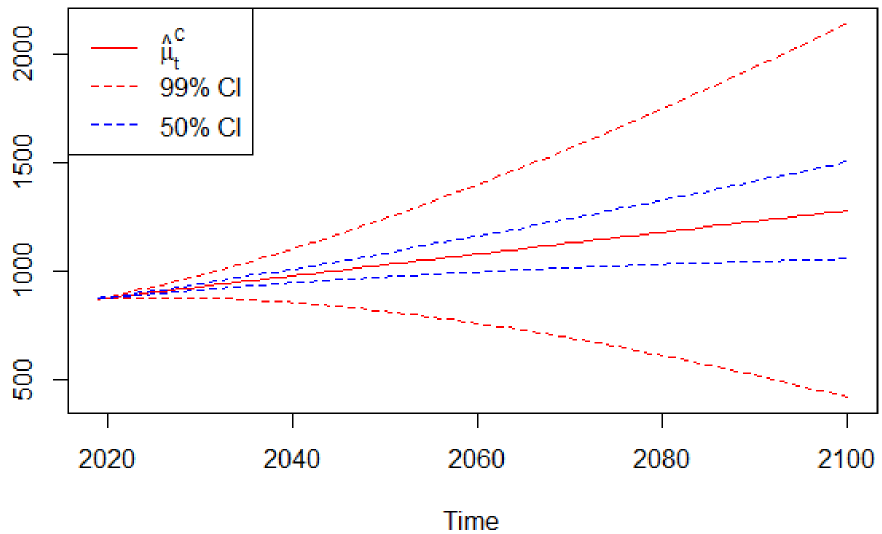


Figure 5: Out of sample forecasts of μ_t^C of C_t in GtC/yr, together with their 99% and 50% confidence intervals.

Scientists have already performed several in-depth analyses of how the CO2 concentration would evolve when different restrictions are imposed over the same forcast horizon (until 2100). This set of scenario forecasts of the atmospheric greenhouse gas concentrations are called Representative Concentration Pathways (RCPs) and are compared to our forecast in the upcoming section.

5.3 Forecasted scenarios

For each category of emissions (for instance, agricultural emissions and aviation emissions), an RCP contains a set of starting values and the estimated emissions up to the year 2100, based on assumptions about economic activity, energy sources, population growth and other socio-economic factors (Moss et al., 2010). We consider four of these pathways: RCP8.5, RCP6, RCP4.5 and RCP2.6. Table 2 specifies the details of each path. Radioactive forcing measures the influence of a factor in the change of the Earth's energy balance (in watt per square meters). The goal of working with these different scenarios is not to predict the future but to better understand uncertainties and alternative futures. In this way, it is easy to derive how robust different political decisions or options may be under a range of possible futures. In order to explore what the RCP concentration scenarios imply for our forecasted emission paths, we use the RCP data set obtained from http://www.iiasa.ac.at/web-apps/tnt/RcpDb/. We have 82 observations, starting from 2019 up to 2100. Note here that the values provided in this data set are slightly lower than the ones indicated for the CO2 concentration in table 2. We plot the RCP CO2 concentration paths for the four scenarios together with our forecasted series of CO2 concentration.

Name	Radiative Forcing	CO2 Concentration (in ppm)	Temperature Anomaly	Pathway
RCP8.5	8.5 Wm^2 in 2100	1370	4.9	Rising
RCP6	6 Wm^2 post 2100	850	3	Stabilization without overshoot
RCP4.5	4.5 Wm^2 post 2100	650	2.4	Stabilization without overshoot
RCP2.6	3 Wm^2 before 2100, declining to $6Wm^2$ 2100	490	1.5	Peak and decline

Table 2: RCP-Scenarios Explanation

Figure 6 compares our forecast with the 4 RCP scenario paths at the 99% confidence interval. At such confidence interval, all paths are within the possible range. Our forecast lies within RCP4.5 and RCP6, which corresponds to a maximum temperature increase of around 2.75 degrees above pre-industrial level (calibrated temperature from Table 2). Yet, since all possible scenarios are within the confidence range, temperature could potentially increase to a maximum of 4.9 degrees above pre-industrial level or to only a maximum of 1.5 degrees as depicted from the two extreme scenarios.

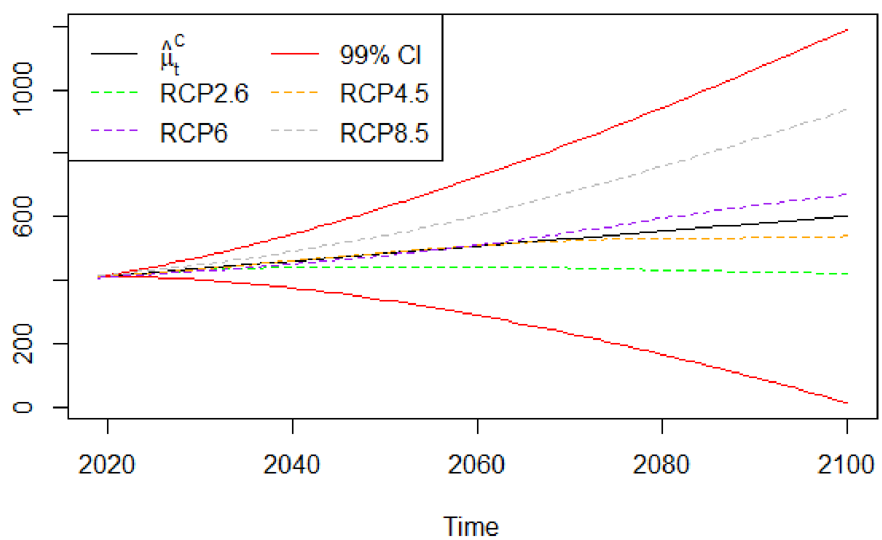


Figure 6: Out of sample forecasts of μ_t^C of C_t in ppmv, together with their 99% confidence intervals, compared to the RCP forecasted scenarios.

Reducing the confidence interval to 75% then renders unlikely the worst case scenario (RCP8.5) as shown in Figure 17 in the Appendix, and further decreasing the confidence interval to 50% leads to the additional exclusion of the best case scenario (RCP2.6) as shown in Figure 7. Those results are given that the future will follow the current trend of the growth of CO2 concentrations. Hence, we can be 50% certain to find ourselves in the medium cases if no changes are imposed on the emission pathway. This shows that in order to reach the goals of the COP21 in Paris, the need for new climate policies becomes urgent.

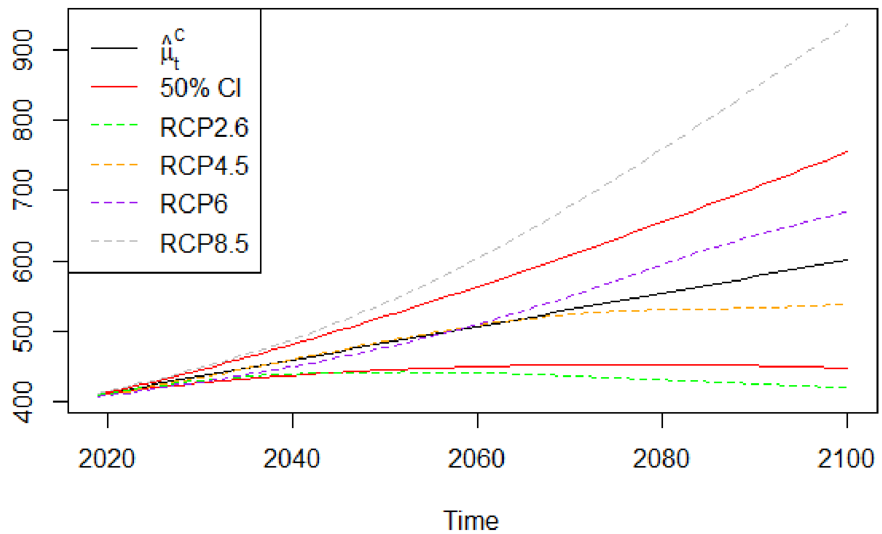


Figure 7: Out of sample forecasts of μ_t^C of C_t in ppmv, together with their 50% confidence intervals, compared to the RCP forecasted scenarios.

In order to evaluate the effects of potential future paths of CO2 emissions on our forecast and its implication regarding the temperature rise in the future, we simulated different scenarios (sensitivity analysis). First, we set E_t equal to its value in 1959, which is also the lowest value. Figure 8 shows how our forecast for the state variable of CO2 concentration gets closer to the best case scenario defined by RCP2.6. The two worst case scenarios are now excluded from the 50% prediction interval. Figure 18 shows that the worst case scenario is still excluded with a 75% confidence interval and Figure 19 shows that all scenarios are still possible with a 99% confidence interval but the latter is tilted towards the best scenario. We therefore conclude that if the overall emission level of CO2 would be immediately decreased to its level of 1959, then the RCP2.6 scenario would likely be reached. On the contrary, if we let CO2 emissions explode and set its value to 20 GtC/yr in the entire out-of-sample period, then we can see from Figure 9 how the estimation for CO2 concentration gets closer to the worst case RCP scenario and the 50% confidence intervals tend to exclude the three best ones. Figure 20 shows that in this case the 75% confidence interval now excludes the best scenario and Figure 21 shows again that all scenario are included within the 99% confidence interval, but the interval is tilted towards the worst scenario. This simulation study shows that the forecast of CO2 concentration based on changes in CO2 emissions is in line with our expectations and suggest that our model is indeed appropriate for this purpose, based on the

Carbon Budget Equation.

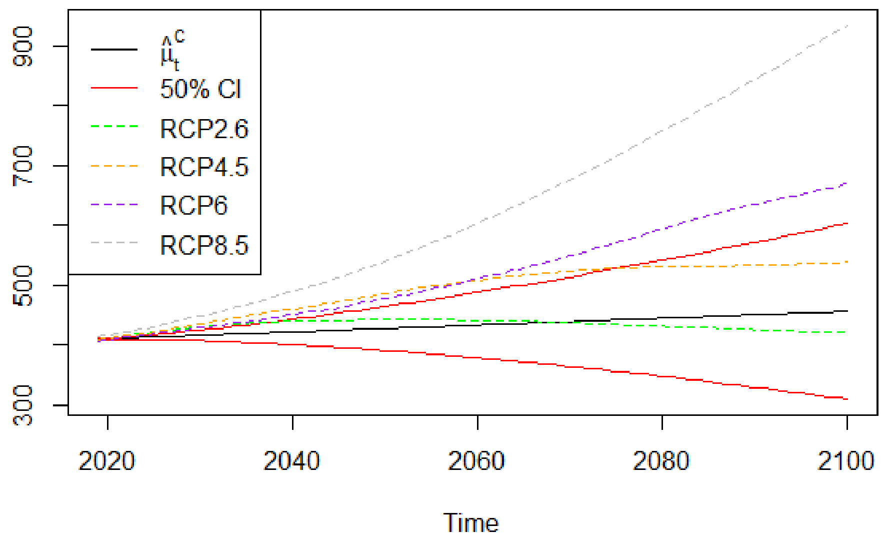


Figure 8: Out of sample forecasts of μ_t^C of C_t in ppmv, together with their 50% confidence intervals, compared to the RCP forecasted scenarios, assuming that E_t is equal to its value in 1959 in the out-of sample period.

6 Discussion

A key motivation to undertake this study has been to model CO2 atmospheric concentration using an econometric approach. The basis of our model is the Global Carbon Budget Equation and by using a state-space approach we intend to tackle the measurement errors that climate data are subject to. Following the critiques of Friedlingstein (2015) and Bennedsen et al. (nd), we do recognize that a drawback of the constructed model is that it does not incorporate the carbon-climate feedback. Since the discussion of these effects are still heated, we do realize that it is worth wile to expand the model and include these effects. Therefore, we suggest this for further research, following the idea of coupled carbon-climate models.

Moreover, due to having only yearly data at hand, the sample size is relatively small (59 data points). Hence, drawing reliable conclusions is very hard and limited, e.g. tests like the ADF-test may not be valid. We also acknowledge this problem in the forecasts. Predicting more than 80 out-of-sample forecasts with a model that is estimated using less than 60 observations leads to substantial uncertainty, especially the further in the future the forecasts are.

A third limitation is the number of variables used in the model. We decided to merge all emissions and sinks to reduce the dimensionality. In this way, we cannot distinguish the emissions and

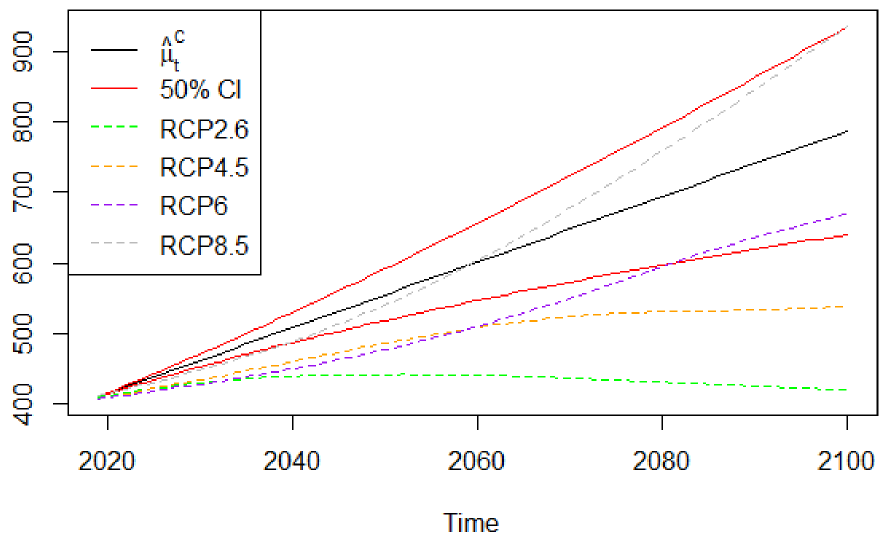


Figure 9: Out of sample forecasts of μ_t^C of C_t in ppmv, together with their 50% confidence intervals, compared to the RCP forecasted scenarios, assuming that E_t is equal to 20 GtC/yr in the out-of sample period.

sinks and therefore cannot observe the individual effects and cannot draw individual conclusions. Additionally, we limited our data set to those 5 variables while potentially other external factors could influence CO2 concentration, such as the direct, or indirect effects of other greenhouse gas emissions for instance.

Another limitation arises in the sensitivity analysis, where we set the value of emissions equal to the value that it had in 1959 (or to a high fixed value) and then keep it constant in the model. This is done in order to then evaluate the path of the carbon concentration after a policy intervention for example. However, this is not realistic due to two reasons. First of all, it is assumed that the value of emissions will stay constant over time. This is highly unlikely. Second, the model currently assumes a big jump in the value of emissions at the beginning to investigate the different scenarios. An extension of this simulation study would be to let the values gradually increase/decrease at a specific rate over time, hence find a decreasing time series of emissions such that the value of C_{2100} is equal to the desired value depending on which scenario we are investigating.

To put it in a nutshell, the model constructed in this paper is rather simplified and may lack dynamics or external factors. However, it tackles the problem of measurement errors, it is based on the Carbon Budget Equation, and its forecasts in are line with climate specialists scenarios forecasts. Yet, further research could extend the constructed state space model to incorporate the carbon-climate feedback effects and to model all of the environmental variables of interest individually.

7 Conclusion

Climate change is nowadays a key topic of interest in political discussions. It is crucial to have proper models and perform accurate forecasts of climate-related variables to efficiently tackle the issues that climate change brings. This gave rise to multidisciplinary approaches, such as using econometric methods to model climate series and their dynamics. This paper focuses on the forecast of carbon dioxide atmospheric concentration.

We propose a state space approach that takes into account the measurement errors to which climate data are subject to. This was motivated by the persistent Global Carbon Budget Equation imbalance observed in practice over the past years. The model is based on this equation and allows to jointly estimate and forecast all the unobserved components driving the observed variables of the Budget Equation and CO2 concentrations. Our analysis provides insights into whether the goals set during climate summits, such as the maximum of 2 degrees rise abover pre-industrial levels set by the COP21, are achievable under the current evolution of the emissions and sinks.

We compare our forecast of CO2 concentration from 2018 to 2100 to Representative Concentration Pathways that correspond to distinct scenarios leading to different temperature increases above pre-industrial levels. We find that if all variables follows the current trend, any scenario from a maximum increase of 1.5 degrees above pre-industrial level to a maximum increase of 4.9 degrees lies within the 99% confidence interval. Even though the worst case scenario is excluded in a 50% confidence interval, so is the best case scenario, which corresponds to a maximum increase of 1.5 degrees.

We also performed a sensitivity analyses by imposing future emission paths and investigate their effects on potential future temperature rise using the same RCPs scenarios. We find that if we drop CO2 emissions to their lowest level registered in 1959, then our forecasts are tilted towards the best-case scenarios and we can then say with 50% certainty that the two worst case scenarios will not happen. Instead, if we let CO2 emissions explode, the worst-case scenarios becomes the most likely one. Hence, we can recognize here the serious need of new climate policies in order to reach this goal. The sensitivity analysis also implies that our forecasts of CO2 concentration based on changes in CO2 emissions are in line with our expectations and suggest that our model is indeed appropriate for this purpose. Moreover, the paper presented several limitations of the data set and model encountered.

References

Ballantyne, A. P., Andres, R., Houghton, R., Stocker, B. D., Wanninkhof, R., Anderegg, W., Cooper, L. A., DeGrandpre, M., Tans, P. P., Miller, J. B., Alden, C., and White, J. W. C. (2015). Audit of the global carbon budget: estimate errors and their impact on uptake uncertainty. *Biogeosciences*, 12(8):2565–2584.

Bennedsen, M., Hillebrand, E., and Koopman, S. J. (n.d.). Trend analysis of the airborne fraction and sink rate of anthropogenically released co2.

Canadell, J., Ciais, P., Gurney, K., Le Quéré, C., Piao, S., Raupach, M., and Sabine, C. (2011). An international effort to quantify regional carbon fluxes. *Eos*, 92(10):81–82.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford university press.

Friedlingstein, P. (2015). Carbon cycle feedbacks and future climate change. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2054):20140421.

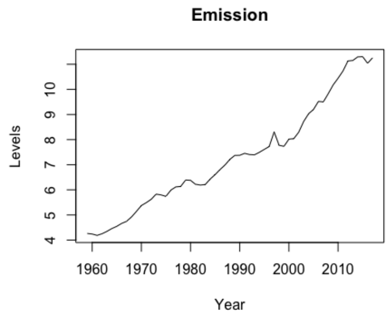
Gloor, M., Sarmiento, J. L., and Gruber, N. (2010). What can be learned about carbon cycle climate feedbacks from the co 2 airborne fraction? *Atmospheric Chemistry and Physics*, 10(16):7739–7751.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Arneth, A., Arora, V. K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Doney, S. C., Gkritzalis, T., Goll, D. S., Harris, I., Haverd, V., Hoffman, F. M., Hoppema, M., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Johannessen, T., Jones, C. D., Kato, E., Keeling, R. F., Goldewijk, K. K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., Wright, R., Zaehle, S., and Zheng, B. (2018). Global carbon budget 2018. *Earth System Science Data*, 10(4):2141–2194.

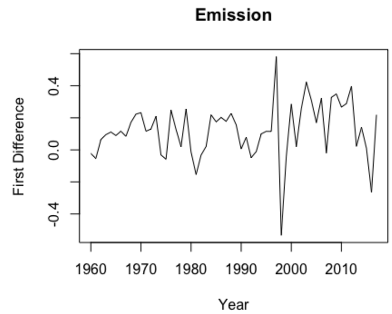
Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., et al. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, 463(7282):747.

Strassmann, K. M. and Joos, F. (2018). The bern simple climate model (bernscm) v1. 0: an extensible and fully documented open-source re-implementation of the bern reduced-form model for global carbon cycle–climate simulations. *Geoscientific Model Development*, 11(5):1887–1908.

Appendices

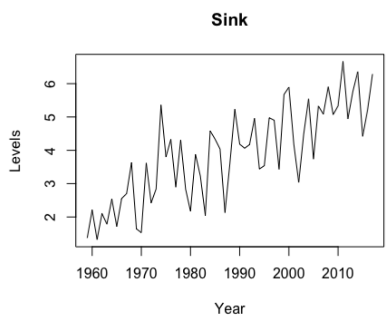


(a) Emission Level

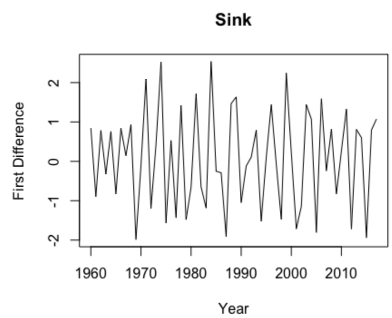


(b) Emission First Difference

Figure 10: Emission

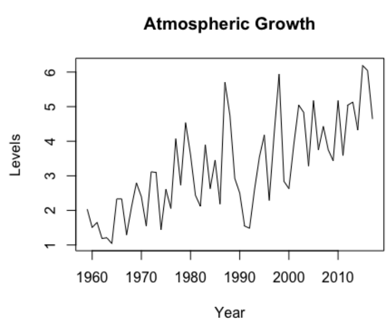


(a) Sink Level

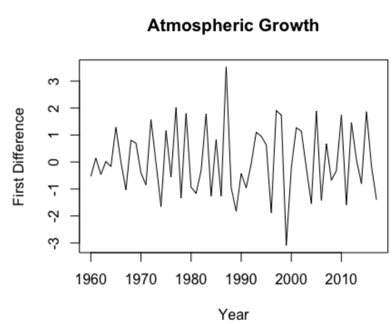


(b) Sink First Difference

Figure 11: Sink



(a) Atmospheric Growth Level



(b) Atmospheric Growth First Difference

Figure 12: Atmospheric Growth

	Atmospheric Growth	Emission	Sink
Mean	3.258692	7.281585	3.88693
Standard Deviation	1.381219	2.130644	1.41866

Table 3: Summary Statistics

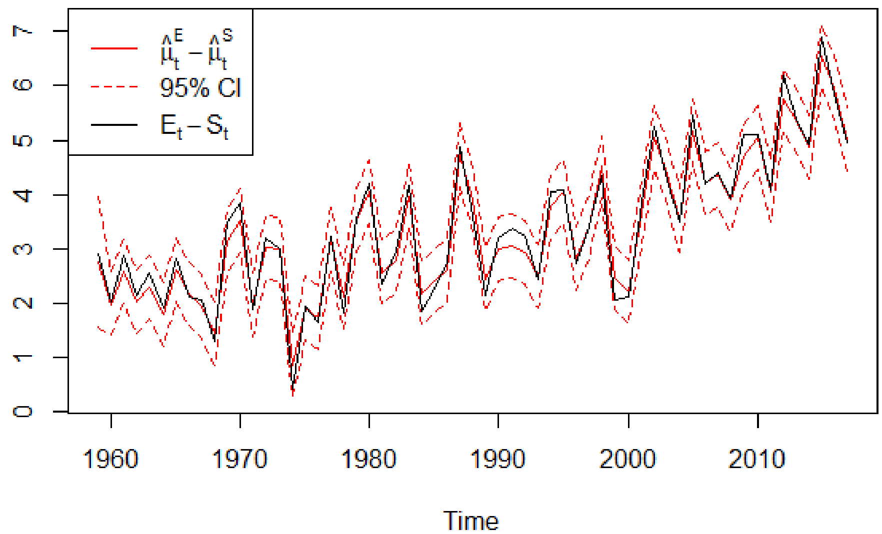


Figure 13: In-sample estimates of the state variable $\mu_t^E - \mu_t^S$ of $E_t - S_t$ in GtC/yr, together with their 95% confidence intervals.

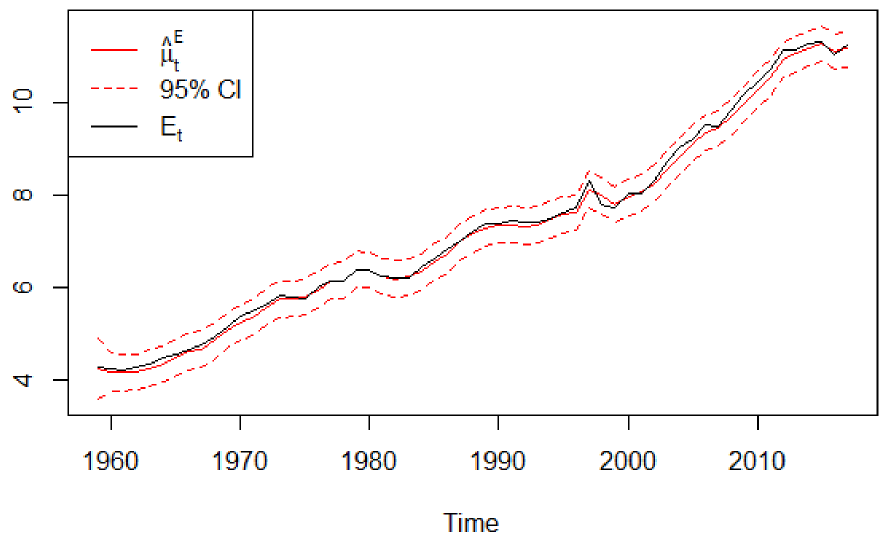


Figure 14: In-sample estimates of the state variable μ_t^E of E_t in GtC/yr, together with their 95% confidence intervals.

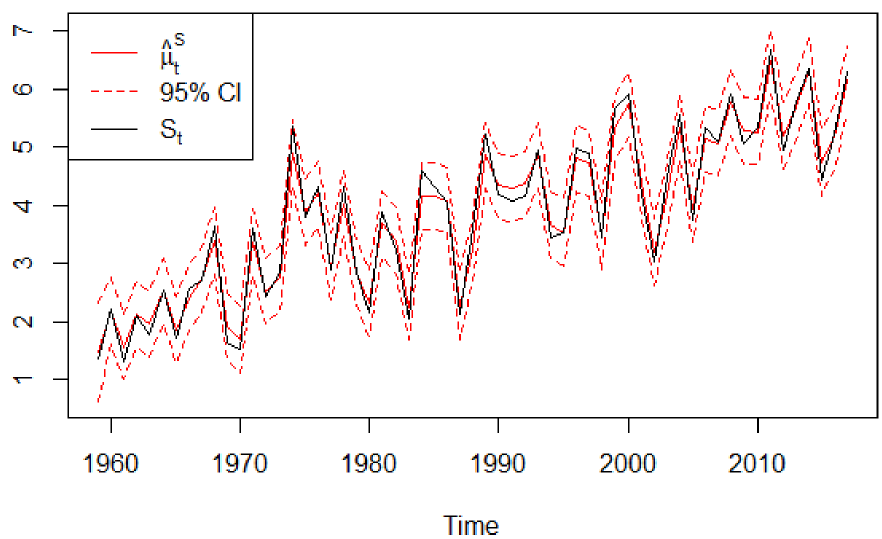


Figure 15: In-sample estimates of the state variable μ_t^S of S_t in GtC/yr, together with their 95% confidence intervals.

Parameters in H	Estimates	Parameters in Q	Estimates
$\sigma^2_{C,\epsilon}$	0.291	$\sigma^2_{C,\eta}$	0.403
$\sigma^2_{G,\epsilon}$	0.661	$\sigma^2_{E,\eta}$	0.084
$\sigma^2_{E-S,\epsilon}$	0.248	$\sigma^2_{S,\eta}$	0.505
$\sigma^2_{E,\epsilon}$	0.072		
$\sigma^2_{S,\epsilon}$	0.17		

Table 4: Maximum likelihood estimates of the hyperparameters in the state space model.

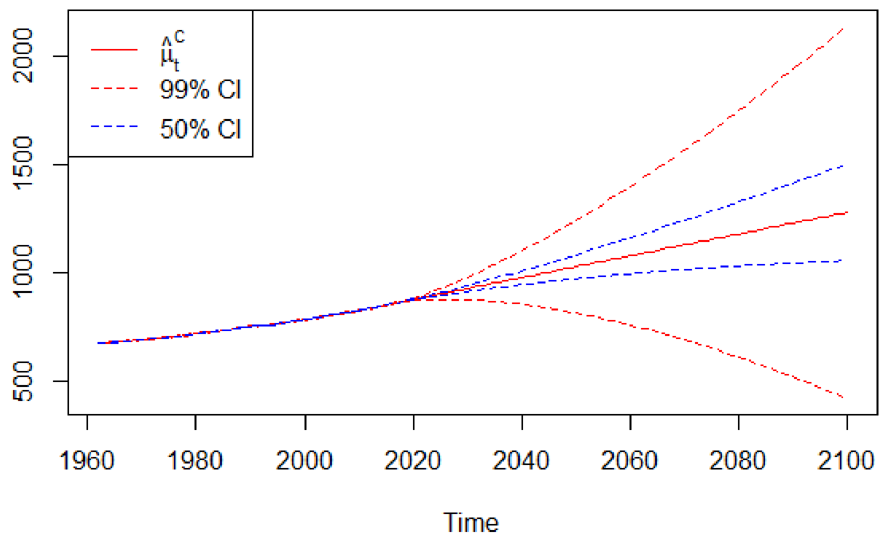


Figure 16: Out of sample forecasts of μ_t^C of C_t, together with their 99% and 50% confidence intervals.

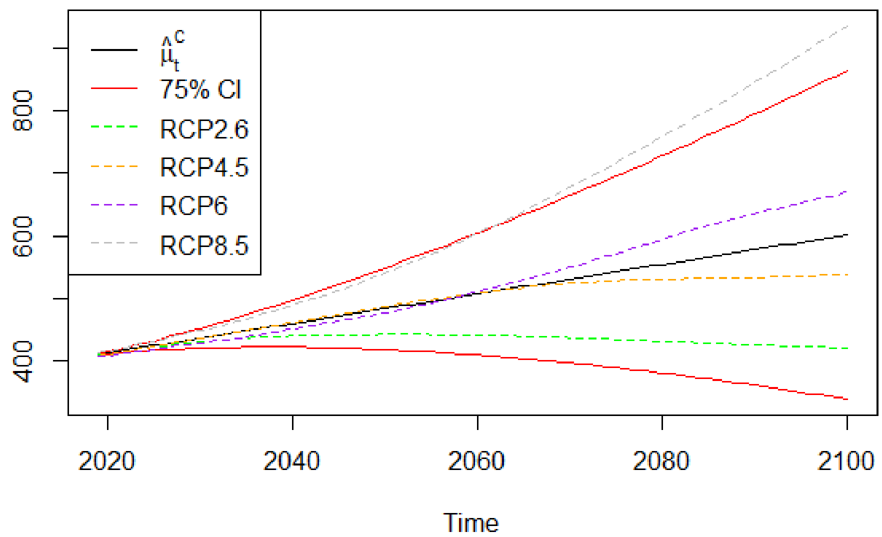


Figure 17: Out of sample forecasts of μ_t^C of C_t in ppmv, together with their 75% confidence intervals, compared to the RCP forecasted scenarios.

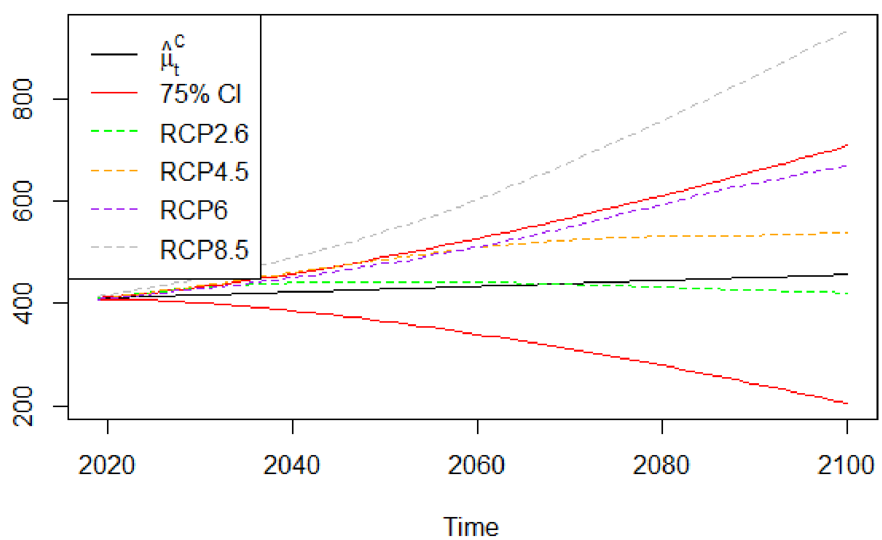


Figure 18: Out of sample forecasts of μ_t^C of C_t in ppmv, together with their 75% confidence intervals, compared to the RCP forecasted scenarios, assuming that E_t is equal to its value in 1959 in the out-of sample period.

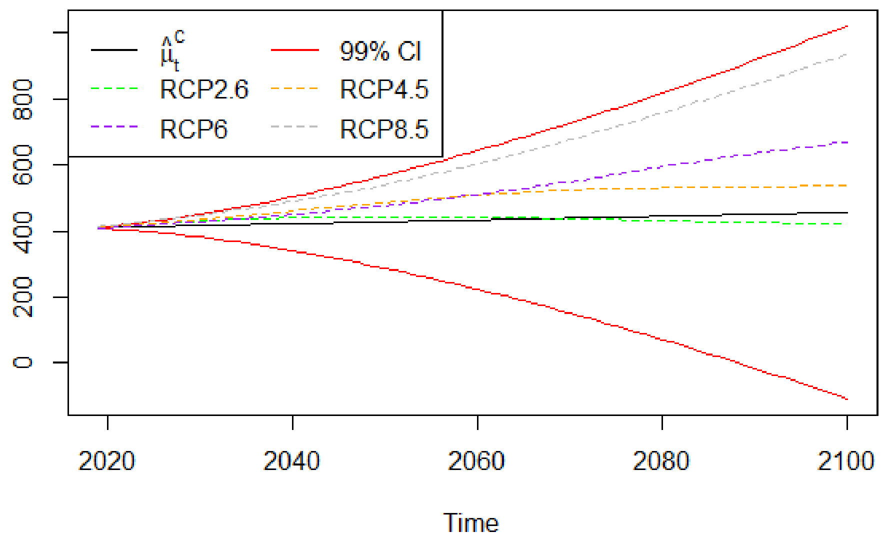


Figure 19: Out of sample forecasts of μ_t^C of C_t in ppmv, together with their 99% confidence intervals, compared to the RCP forecasted scenarios, assuming that E_t is equal to its value in 1959 in the out-of sample period.

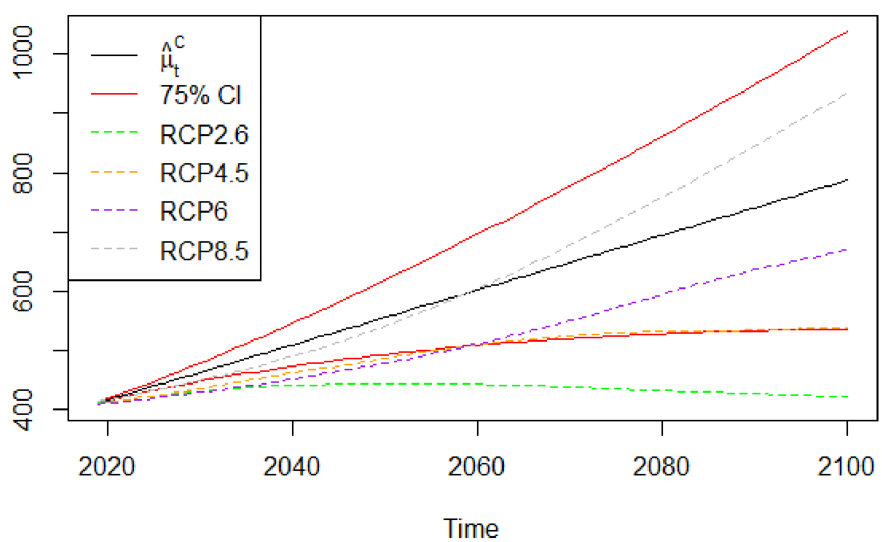


Figure 20: Out of sample forecasts of μ_t^C of C_t in ppmv, together with their 75% confidence intervals, compared to the RCP forecasted scenarios, assuming that E_t is equal to 20 GtC/yr in the out-of sample period.

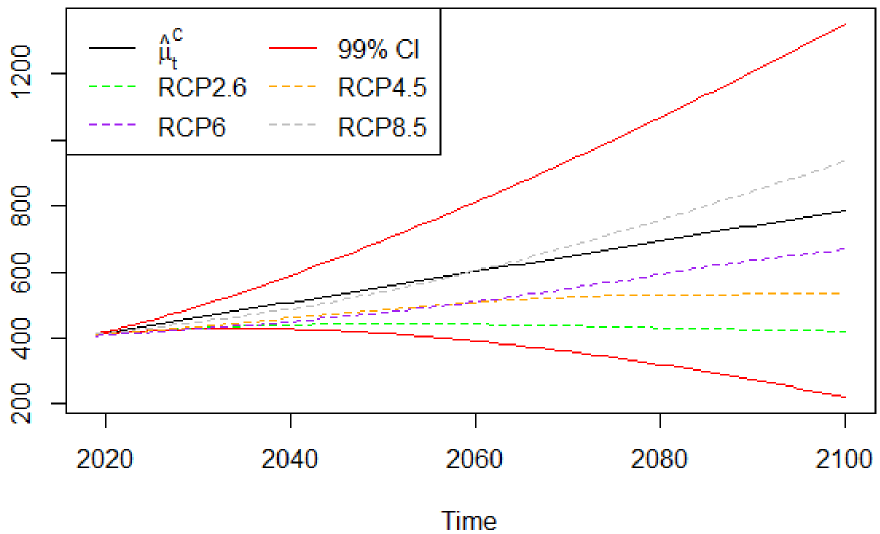


Figure 21: Out of sample forecasts of μ_t^C of C_t in ppmv, together with their 99% confidence intervals, compared to the RCP forecasted scenarios, assuming that E_t is equal to 20 GtC/yr in the out-of sample period.