A Multivariate Dynamic Statistical Model of the Global Carbon Budget 1959 – 2020*

Mikkel Bennedsen[a,b] *Eric Hillebrand*[a,b] *Siem Jan Koopman*[b,c]

[a]Department of Economics and Business Economics, Aarhus University, Denmark

[b]Center for Research in Econometric Analysis of Time Series (CREATES), Denmark

[c]Department of Econometrics, Vrije Universiteit Amsterdam, The Netherlands

September 26, 2022

Abstract

We propose a multivariate dynamic statistical model of the Global Carbon Budget as represented in the annual data set made available by the Global Carbon Project, covering the sample period 1959–2020. The model connects four main objects of interest: atmospheric carbon dioxide (CO_2) concentrations, anthropogenic CO_2 emissions, the absorption of CO_2 by the terrestrial biosphere (land sink), and by the ocean and marine biosphere (ocean sink). The model captures the global carbon budget equation, which states that emissions not absorbed by either land or ocean sinks must remain in the atmosphere and constitute a flow to the stock of atmospheric concentrations. Emissions depend on global economic activity as measured by World Gross Domestic Product while sink activities depend on the level of atmospheric concentrations and the Southern Oscillation Index. We derive the time-series properties of atmospheric concentrations from the model, showing that they contain one unit root and a near-second unit root. The statistical system allows for the estimation of key parameters of the global carbon cycle and for the assessment of estimation uncertainty. It also allows for the estimation and the uncertainty assessment of related variables such as the airborne fraction and the sink rate. We provide short-term forecasts of the components of the global carbon budget.

Keywords: Global Carbon Budget, World Gross Domestic Product, CO_2 emissions, CO_2 concentrations, El Niño Southern Oscillation, airborne fraction, sink rate, climate economics, climate econometrics.

*We are grateful for comments from Jurgen Doornik, David Hendry, Søren Johansen, Andrew Martinez, and Felix Pretis, as well as from participants at the fourth conference on Econometric Models of Climate Change in Milan in 2019, the Climate Econometrics online seminar 2020, the CREST Paris virtual seminar 2021, the NBER-NSF virtual conference 2021, and the 12th Workshop in Time Series Econometrics 2022 at the University of Zaragoza, Spain. EH posed this modeling problem as Case A of the Econometric Game in 2019 at the University of Amsterdam, and we are indebted to the organizers and participants of the Game. Special thanks go to Peter Boswijk and Joep Keuzenkamp. MB and EH acknowledge support from the Independent Research Fund Denmark. Addresses: Mikkel Bennedsen, Center for Research in Econometric Analysis of Time Series (CREATES), Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark, `mbennedsen@econ.au.dk`; Eric Hillebrand, CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark, `ehillebrand@creates.au.dk`; Siem Jan Koopman, Department of Econometrics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, `s.j.koopman@vu.nl`

1 Introduction

We propose a multivariate dynamic statistical model for the main time series variables contained in the most recent vintage of the Global Carbon Budget (GCB) data set from Friedlingstein et al. (2022). The model allows for the statistical estimation of key parameters of the global carbon cycle from the data and the assessment of the corresponding estimation uncertainty. It implies near-second unit root time series dynamics for atmospheric CO_2 concentrations. It allows for the statistical assessment of the airborne fraction and the sink rate and for short-term forecasts of the components of the global carbon budget.

The model connects atmospheric carbon dioxide (CO_2) concentrations, anthropogenic emissions, and uptake by the terrestrial biosphere (land sink) and by the ocean and the marine biosphere (ocean sink). The model has the global carbon budget equation as its cornerstone. It specifies both sinks as depending on atmospheric CO_2 concentrations and the El-Niño/Southern Oscillation (ENSO) cycle, and it specifies emissions as a random walk with drift determined by economic growth. The dynamics of atmospheric CO_2 concentrations are determined by the global carbon budget equation. Since sinks activity depends on concentrations in turn, the model captures simultaneity in the determination of the GCB variables. This allows for the data-driven study of the global carbon cycle employing a relatively small model for the GCB data set that consists of both observational data and output from several large-scale general circulation models (GCMs). Parameter estimates are obtained from the historical GCB data by maximum likelihood, and parameter uncertainty can be evaluated by means of statistical standard errors. In contrast, this uncertainty cannot be measured from GCMs or small-scale emulators.

The Global Carbon Project[1] curates and maintains a large database of time series variables that describe the dynamics of the carbon cycle in order to provide, for example, insights into how anthropogenically emitted CO_2 is transferred to the atmosphere, the oceans, and the terrestrial biosphere. The data are updated and published annually in a series of reports entitled "The Global Carbon Budget". Understanding the dynamics of the carbon cycle is vital for understanding the climate system in general and climate change in particular (e.g. Canadell et al., 2021).

The GCB data have previously been employed for various statistical analyses. For instance, a strand of literature uses these data to investigate whether the rate at which the carbon sinks absorb CO_2, as measured through either the so-called airborne fraction or sink rate, is decreasing. This literature is represented by, among others, Raupach et al. (2008), Knorr (2009), Le Quéré et al.

[1] `https://www.globalcarbonproject.org`.

(2009), Gloor et al. (2010), Raupach et al. (2014), and Bennedsen et al. (2019). Other studies propose to use the residual of the GCB data, referred to as the budget imbalance, to assess whether CO_2 emissions are reported truthfully by individual nations; see, e.g., Peters et al. (2017) and Bennedsen (2021). A common feature of previous statistical analyses of the GCB data is that model dimensions are limited to univariate or bivariate settings. These earlier studies do generally not consider all GCB variables simultaneously. There are a few early studies that assess parameter uncertainty in models of the global carbon cycle with methods inspired by statistics (Enting and Lassey, 1993; Parkinson and Young, 1998).

The key focus of this study is to incorporate all the main GCB time series variables into a single multivariate dynamic statistical model and estimate it using standard maximum likelihood methods. A crucial implication of a simultaneous approach is that it allows us to exploit the *carbon budget equation* to connect the various GCB variables. The carbon budget equation is an accounting identity expressing the fact that, since the Earth system is closed, anthropogenically released CO_2 must necessarily end up in either the atmosphere, the oceanic biosphere, or the terrestrial biosphere; see Friedlingstein et al. (2022). Our simultaneous approach thus provides a dynamic statistical model that is coherent with the physics underlying the GCB.

We first develop a nonlinear statistical model of the GCB, where the nonlinearity originates in the relationship between the CO_2 uptake of the sinks and the level of atmospheric CO_2 concentrations. Although such a nonlinear relation is expected on theoretical grounds (e.g., Bacastow and Keeling, 1973; Gifford, 1993; Joos et al., 1996), we find that a linear specification is adequate for the historical data for the period 1959–2020. Approximate linearity of the sink-concentrations relationship is also documented in previous work (e.g., Knorr, 2009; Raupach, 2013), and it is likely to be a consequence of the relatively low levels of atmospheric CO_2 concentrations over the period 1959–2020. Nonlinear effects in the sink-concentrations relationship could soon become important if atmospheric concentrations continue to rise (Canadell et al., 2007b; Raupach, 2013; Raupach et al., 2014; Bennedsen et al., 2019).

From the linear system equations of this model, we derive the time series dynamics of atmospheric CO_2 concentrations. We find that CO_2 concentrations follow single unit root dynamics that are, however, numerically quite close to a second unit root, and that they approach a second unit root as atmospheric concentrations increase. We show that the result is due to the dependence of the sinks on concentrations.

Motivated by our findings of linearity of the sink-concentration relationship over the period 1959–

2020, we consider a multivariate linear dynamic statistical model of the GCB. The model features a measure of global economic activity (World Gross Domestic Product, GDP) as a driver of emissions and the Southern Oscillation Index (SOI) variable as a proxy for the ENSO in the sinks dynamics. In addition, the model includes a number of dummy variables for specific unusual events in atmospheric concentrations and in the relation of emissions and World GDP. We present the estimation results that include parameter estimation uncertainty measures. Using this model, we present a comprehensive statistical analysis of the Global Carbon Budget data. The statistical features of the historical in-sample estimates of all GCB variables are documented in detail, including those of related variables such as the budget imbalance, airborne fraction, and sink rate. All presented parameter estimates are accompanied with statistical standard errors while the in-sample estimates are accompanied with confidence intervals. We show how the modeling framework can be used for the simultaneous and coherent forecasting of all GCB variables.

The remainder of the paper is organized as follows. Section 2 introduces the model, discusses its key assumptions, provides the details of the state space representation of the dynamic model, and investigates the model-implied dynamic properties of atmospheric concentrations. Section 3 describes the data set that we use and discusses time series properties of the data series. Section 4 presents and discusses the estimation results and the residual diagnostics for a model that includes World GDP and SOI variables. Section 5 discusses the statistical forecasting of the GCB variables based on our model, the in-sample estimation of key variables such as airborne fraction and sink rate, and the diagnostic analysis of the GCB imbalance. Section 6 concludes. Mathematical derivations, details on stationarity tests, an extensive Monte Carlo simulation study, a model validation exercise, and a forecast model for the SOI are provided in supplementary material.

2 A dynamic statistical model for the global carbon budget

We develop a statistical state space model for the global carbon budget (GCB) comprising the following four variables: atmospheric CO_2 concentrations (C_t^*), anthropogenic CO_2 emissions (E_t^*), CO_2 uptake by the terrestrial biosphere (land sink, $S\_LND_t^*$), and CO_2 uptake by the ocean and marine biosphere (ocean sink, $S\_OCN_t^*$). We denote the unobserved states in the state-space model with asterisks; later in this section, we connect the unobserved states with the data series, for which we use the same variable names without asterisk.

The flow series E_t^*, $S\_LND_t^*$, and $S\_OCN_t^*$ are measured in gigatons of carbon (GtC) per year; the stock series C_t^* is measured in GtC. The foundation of our statistical model is the global carbon budget equation as given by (Friedlingstein et al., 2022)

$$G\_ATM^*_{t+1} := C^*_{t+1} - C^*_t = E^*_{t+1} - S\_LND^*_{t+1} - S\_OCN^*_{t+1},$$

where G_ATM^* represents the change in atmospheric concentrations in GtC per year. The budget equation expresses the fact that emissions not absorbed by land or ocean sinks constitute a flow to the stock of atmospheric concentrations. It also implies a dynamic process for concentrations,

$$C^*_{t+1} = C^*_t + E^*_{t+1} - S\_LND^*_{t+1} - S\_OCN^*_{t+1}. \tag{1}$$

The updating equation (1) for C^*_t serves as the cornerstone for our statistical model. We complete the model by specifying dynamic equations for the variables $S\_LND^*_t$, $S\_OCN^*_t$, and E^*_t.

2.1 Land and ocean sinks

The sink variables $S\_LND^*_t$ and $S\_OCN^*_t$ represent the CO_2 fluxes from the atmosphere to the land biosphere and the ocean, respectively, in year t. The magnitude of these uptakes depends primarily on the level of CO_2 concentrations in the atmosphere C^*_t. In the case of S_LND^*, this relation is due to the *fertilization effect*, where increased levels of CO_2 in the atmosphere cause increased CO_2 uptake by the terrestrial biosphere (e.g., Bacastow and Keeling, 1973; Gifford, 1993). In the case of S_OCN^*, the relation is due to the fact that increased levels of CO_2 in the atmosphere will, other things being equal, cause the partial pressure differential for CO_2 between the atmosphere and the surface layer of the ocean to increase, which in turn implies an increased CO_2 uptake by the ocean (Joos et al., 2001). For these reasons, we expect that S_LND^* and S_OCN^* will be positively related to C^*_t. Further, for moderate levels of atmospheric CO_2 concentrations it has been found that these relations are approximately linear (Raupach, 2013). However, due to a *saturation effect* it is expected that this approximate linearity will likely break down as the level of C^*_t increases (Canadell et al., 2007b; Le Quéré et al., 2007). Whether the current level of atmospheric CO_2 concentrations ($C_{2020} \approx 880$ GtC) implies that we have already entered a regime where nonlinear effects in the sink-concentrations relationship become important has been the subject of much recent work (e.g. Knorr, 2009; Le Quéré et al., 2009; Gloor et al., 2010; Raupach et al., 2014; Bennedsen et al., 2019). Nonlinear functional forms of the sinks have been developed and proposed by, among others, Bacastow and Keeling (1973), Gifford (1993), and Joos et al. (1996).

Although the level of atmospheric CO_2 concentrations is by far the most important factor in determining the long-term behavior of the sinks, the short-term can also be influenced by factors other than the global carbon budget as defined here, the El-Niño/Southern Oscillation (ENSO) cycle explaining most of the variation. These considerations lead us to specify the following dynamic equations, which are possibly nonlinear in atmospheric CO_2 concentrations, C^*,

$$S\_LND^*_{t+1} = c_1 + f_{SL}(C^*_{t+1}) + b_1 \mathrm{ENSO}_{t+1}, \qquad S\_OCN^*_{t+1} = c_2 + f_{SO}(C^*_{t+1}) + b_2 \mathrm{ENSO}_{t+1}, \tag{2}$$

where $c_1, c_2, b_1, b_2 \in \mathbb{R}$ are constants, $f_{SL}(\cdot)$ and $f_{SO}(\cdot)$ are particular sink functions for land and ocean, respectively, and ENSO_{t+1} denotes a measure of the ENSO activity in year $t+1$.

In this paper, we accommodate the nonlinear aspects by including time-varying parameters in the model. To motivate this approach, consider the function $f_{SL}(\cdot)$ in the land sink equation in (2), which we can rewrite as $f_{SL}(C^*_{t+1}) = f^*_{SL}(C^*_{t+1})\, C^*_{t+1}$ where $f^*_{SL}(C^*_{t+1}) = f_{SL}(C^*_{t+1})/C^*_{t+1}$. By treating $f^*_{SL}(C^*_{t+1})$ as a time-varying parameter $\widetilde{\beta}_{1,t+1}$ and ignoring its dependence on C^*_{t+1}, we can consider, for example, a random walk specification for $\widetilde{\beta}_{1,t+1}$, that is, $\widetilde{\beta}_{1,t+1} = \widetilde{\beta}_{1,t} + \omega_{1,t}$, where $\omega_{1,t}$ is an independently and identically distributed random innovation variable with mean zero and variance $\sigma^2_{\omega,1} > 0$, and it is mutually independent from all other innovations in the model. The resulting approximation $f_{SL}(C^*_{t+1}) \approx \widetilde{\beta}_{1,t+1}\, C^*_{t+1}$ delivers a specification that is nonlinear because it features the product of two state variables. The same arguments can be used for the ocean sink equation in (2) to obtain a time-varying parameter specification $f_{SO}(C^*_{t+1}) \approx \widetilde{\beta}_{2,t+1}\, C^*_{t+1}$. Simulation experiments, presented in the supplementary material S1, show high levels of accuracy of this approximation for various nonlinear sink specifications, including those from Bacastow and Keeling (1973), Gifford (1993), and Joos et al. (1996).

2.2 Emissions

The dynamic evolution of emissions E^* is assumed to follow a random walk process with a drift governed by global economic activity (ECON), capturing the strong dependence that has historically existed between economic activity and anthropogenic CO_2 emissions. Bennedsen et al. (2021) show that U.S. CO_2 emissions can be modeled effectively by industrial production indices, leading to accurate forecasts. Friedlingstein et al. (2022) and earlier vintages of the GCB (Friedlingstein et al., 2020, 2019; Le Quéré et al., 2018, 2017) model and forecast emissions by measures of economic activity, following Raupach et al. (2007). The energy economics literature has discussed the relation of energy consumption and macroeconomic activity at length (e.g., Stern, 1993, 2000; Oh and Lee, 2004; Lee, 2005; Zhang and Cheng, 2009; Ozturk, 2010). Our dynamic model for E^* is therefore given by

$$E^*_{t+1} = E^*_t + \beta_{5,t+1}\Delta\mathrm{ECON}_{t+1} + X^E_t, \tag{3}$$

where $\Delta\mathrm{ECON}_{t+1} = \mathrm{ECON}_{t+1} - \mathrm{ECON}_t$ is the change in economic activity from year t to year $t+1$, $\beta_{5,t+1}$ is a time-varying coefficient of economic activity, and X^E_t is a stationary random innovation. Analogously to the time-varying coefficients in the sink equations, we model β_5 as a random walk: $\beta_{5,t+1} = \beta_{5,t} + \omega_{5,t}$. Conditional on the exogenous variables ENSO and ECON, X^E_t is the only source of randomness driving the budget variables (1)–(3), which implies that the randomness in the state variables of the statistical model of the GCB is due only to X^E_t. In our study, we consider a stationary first-order autoregressive process for X^E_t, which we specify as

$$X^E_{t+1} = \phi_E\, X^E_t + \kappa_t, \tag{4}$$

where $|\phi_E| < 1$ is the autoregressive coefficient and κ_t is a sequence of mean-zero independent random variables.

2.3 Outlier events

We include a number of dummy variables in our model to capture outliers and structural breaks. The selection of the dummies was conducted using a variety of methods and criteria. The point of departure was a purely data-driven search with AutoMetrics (Doornik, 2009; Pretis et al., 2018). The set of selected dummies was modified by a search with the following criteria: (i) The number of dummies should be minimal. (ii) The dummies should have an identifying event or narrative. (iii) The dummies should be statistically significant at least at the 10% level. (iv) The numerical maximum likelihood should be reasonably close to its highest value obtained from models with a larger set of dummies. (v) All considered residual diagnostics should have values within an appropriate range. (Details of the estimation and residual diagnostics are discussed below in Section 4.)

The final set of dummy variables contains: (1) 1991 in the state equation for G_ATM^* captures the Pinatubo minimum (Bousquet et al., 2000; Angert et al., 2004). (2) 1991 in the state equation for E^*: The collapse of the Soviet Union, the 1990 oil price shock, and the first Gulf War in 1991 are associated with a decrease in emissions. The relation between oil price crises, energy consumption, and macroeconomic activity has been discussed at length in the econometrics and energy economics literature, see, e.g., Hamilton (1983); Perron (1989); Hamilton (1996, 2003); Barsky and Kilian (2004);

Kilian (2008, 2009); Stern and Kander (2012). (3) 1997 in the measurement equation for E: There is a strong spike in levels of E in 1997 due to burning of South East Asian peatlands (Houghton and Nassikas, 2017). (4) 1996 in variance: Panels [j], [k], and [l] of Figure 1, presented below, show that first differences of E inherit an increase in variance from first differences of land-use change (E_LUC) in 1996. See the supplementary material S2 for a discussion of possible reasons for this increase in variance.

2.4 State equations for the GCB variables

In summary, the dynamic model for the GCB variables is specified as

$$\begin{aligned} C^*_{t+1} &= C^*_t + E^*_{t+1} - S\_LND^*_{t+1} - S\_OCN^*_{t+1} + \beta_7 I1991, && (5) \\ S\_LND^*_{t+1} &= c_1 + \frac{\beta_{1,t+1}}{C_{1750}} C^*_{t+1} + \beta_3 \mathrm{ENSO}_{t+1}, && (6) \\ S\_OCN^*_{t+1} &= c_2 + \frac{\beta_{2,t+1}}{C_{1750}} C^*_{t+1} + \beta_4 \mathrm{ENSO}_{t+1}, && (7) \\ E^*_{t+1} &= E^*_t + \beta_{5,t+1} \Delta\mathrm{ECON}_{t+1} + \beta_8 I1991 + X^E_t, && (8) \end{aligned}$$

where we set $C_{1750} = 593.43$GtC as the level of atmospheric CO_2 concentrations in the pre-industrial era, here taken as the level in the year 1750, and use it as a scaling of the time-varying coefficients in the sinks equations. As explained in the subsection above, the variance of the E^* innovation X^E_t is subject to a break. We have $X^E_{t+1} = \phi_E X^E_t + \kappa_t$ with $\kappa_t \sim N(0, \sigma^2_{\eta_4} \times (s^2_E)^{I_{t\geq 1996}})$, defining a variance change from 1996 onwards. The variables IYEAR denote indicator (dummy) variables for the year stated.

Equations (5) to (8) specify the state equations for the variables of interest. In addition, the state vector in the state space model contains the stationary process X^E_t, the coefficient processes $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{5,t}$, and further stationary processes in the measurement equations to be defined in the next section. Since the coefficient processes $\beta_{1,t}$ and $\beta_{2,t}$ are state processes in the model, the products $\beta_{1,t} C^*_t$ and $\beta_{2,t} C^*_t$ render the model nonlinear.

2.5 Observation equations for the GCB variables

Annual observations of the four variables at a global level are provided by the Global Carbon Project (Friedlingstein et al., 2022) for the years from 1959 up to 2020. In this study, they are denoted with the same variable names but without the asterisks. Atmospheric CO_2 concentrations are instrumental measurements. Emissions are computed from the use of fossil energy carriers as reported by countries' authorities. To compute the variable "anthropogenic emissions", we take fossil fuel emissions plus land use change emissions minus the cement carbonation sink, see also Section 3 below. These observations are subject to measurement errors and other irregularities due to data collection. The observations of the land and ocean sinks, on the other hand, are averages over the outputs of several GCM/Earth system models selected by the Global Carbon Project. We will treat these observations statistically as data in our model. For the sinks processes, the model should be understood as providing an approximation to the more complex climate models, and it only captures parts of the more detailed interrelations in the climatologically and mathematically more involved large scale climate models. We show, however, that our model approximations are sufficiently accurate for a statistical analysis of the historical data, in the sense that they render the residuals statistically indistinguishable from white noise. The deviations of the observed variables (without asterisk) from the unobserved model variables (with asterisk) are therefore a mix of measurement errors (in particular, for the concentrations and emission variables) and approximation errors (in particular, for the land and ocean sink variables).

Given the dynamic specifications of the four model variables, we complete our statistical model for the observed variables with the measurement or observation equations, which are given by

$$C_t = C_t^* + X_{1,t}, \tag{9}$$

$$S\_LND_t = S\_LND_t^* + X_{2,t}, \tag{10}$$

$$S\_OCN_t = S\_OCN_t^* + X_{3,t}, \tag{11}$$

$$E_t = E_t^* + \beta_6 I1997 + X_{4,t}, \tag{12}$$

where $X_{1,t}$ and $X_{4,t}$, which are associated with C_t and E_t, respectively, can be mainly regarded as measurement errors, whereas $X_{2,t}$ and $X_{3,t}$, which are associated with S_LND_t and S_OCN_t, respectively, mainly represent processes and features that are not captured by our statistical model. The time index $t = 1, \ldots, T$ counts years, where T is the number of available yearly observations for the four variables. The observed variables are collected in the observation vector $y_t = (C_t, S\_LND_t, S\_OCN_t, E_t)'$.

In our study, we consider stationary first-order autoregressive processes for $X_{i,t}$, for $i = 1, \ldots, 4$, which we specify as

$$X_{i,t+1} = \phi_i \, X_{i,t} + \eta_{i,t}, \tag{13}$$

where $|\phi_i| < 1$ is the autoregressive coefficient and $\eta_{i,t}$ is an independently and identically distributed random innovation variable with mean zero and variance $\sigma_i^2 > 0$, and it is mutually independent from innovation κ_t in equation (4), for $i = 1, \ldots, 4$. We allow for correlations between the innovations $\eta_{i,t}$ and $\eta_{j,t}$, as denoted by r_{ij}, for a selection of pairs $i, j = 1, \ldots, 4$.

The dummy variable for 1997 in emissions captures the peat burning events in equatorial Asia (Houghton and Nassikas, 2017). It is included in the measurement equation because it is an outlier in levels of emissions. Treating it in equation (8) for differences in emissions instead would require two dummies.

2.6 State space representation of the GCB model

When the sink variable specifications in (2) are nonlinear, the state space representation is nonlinear as well. The nonlinear state space representation of the GCB model (5)–(12) is obtained by defining the state vector $\alpha_t = (C_t^*, S\_LND_t^*, S\_OCN_t^*, E_t^*, \beta_{1,t}, \beta_{2,t}, \beta_{5,t}, \xi_t')'$ which contains all unobserved variables in the GCB model, the time-varying coefficients, and the measurement error and innovation vector $\xi_t = (X_{1,t}, \ldots, X_{4,t}, X_t^E)'$.

The nonlinear sink equations do not only affect the state equations for the sink variables in (2), but also the budget equation in (5). Hence, we have a fully nonlinear GCB model that can be expressed in state space form as

$$\alpha_{t+1} = \widetilde{T}(\alpha_t, \xi_t), \qquad y_t = \widetilde{Z}(\alpha_t, \xi_t), \tag{14}$$

where $\widetilde{T}()$ and $\widetilde{Z}()$ are nonlinear vector equations that capture the specifications implied by the equations (5)–(8).

In the empirical section 4 we analyze the nonlinear model by means of the Extended Kalman filter. We find that the time-varying parameters $\beta_{1,t}$ and $\beta_{2,t}$ are estimated to be constant for the historical period 1959–2020, indicating that a linear specification of the sink-concentrations relationship is adequate over this time period. Hence the GCB model reduces to a linear dynamic statistical model which we discuss in more detail below.

The GCB modeling framework above reduces to a set of linear equations when we consider linear approximations for the sink variables in (2), which obtain when the time-varying coefficients of the sinks equations are assumed to be constant: $\beta_{1,t} = \beta_1 > 0$ and $\beta_{2,t} = \beta_2 > 0$, for all t. (The time-varying emissions intensity $\beta_{5,t}$ does not introduce nonlinearity in the model, since $ECON_t$ is an exogenous variable and not an element of the state vector. It is also estimated to be constant on the sample.) In the linear case, the dynamic equations for the sink variables can be expressed as

$$f_{SL}(C^*_{t+1}) = c_1 + \frac{\beta_1}{C_{1750}} C^*_{t+1}, \tag{15}$$

$$f_{SO}(C^*_{t+1}) = c_2 + \frac{\beta_2}{C_{1750}} C^*_{t+1}, \tag{16}$$

where the constants c_i, for $i = 1, 2$, are intercepts, the slopes $\beta_i / C_{1750} > 0$, for $i = 1, 2$, are fractions of concentrations that are absorbed annually by the two sink variables, and $C_{1750} = 593.43$GtC is set equal to the concentration levels in 1750. In the supplementary material S3, we show how the sink equations of Bacastow and Keeling (1973), Gifford (1993), and Joos et al. (1996), are approximated by these linear specifications and the scaling of the parameters by pre-industrial concentrations C_{1750}.

The linear system is then given in standard state space form as

$$\begin{aligned} y_t &= Z_t \alpha_t, \\ \alpha_{t+1} &= T_t \alpha_t + \eta_t, \quad \eta_t \sim \mathsf{N}(0, Q_t), \end{aligned} \tag{17}$$

where y_t is the vector of observations defined in Section 2.5 and α_t is the state vector as defined in this section. The GCB model does not feature observation disturbances in the measurement equation for y_t. The deviations of the unobserved state processes (with asterisks) from the observations (without asterisks) are given by covariance-stationary processes $X_{i,t}$, for $i = 1, \ldots, 4$, which are themselves unobserved processes in the state equation. The key assumption for the linear GCB model in state space form is that the two equations in (17) are linear in the state vector α_t and that the disturbance vector η_t is from a normal distribution. It requires the system matrices Z_t, T_t and Q_t to be fixed (non-stochastic) at time t. In Section S4 of the supplementary material, we define the system matrices Z_t, T_t, and Q_t for the linear model, and we confirm that they are fixed. Given the linear Gaussian system, the Kalman filter provides the minimum mean-squares estimator (MMSE) of the state vector α_t given the past observation vectors $y_1, \ldots, y_{t-1}$ and its mean-squared error (MSE) variance matrix, for $t = 1, 2, \ldots$, recursively. The Kalman filter further provides the observation prediction error vectors and their variance matrices, which provide the input for computing the Gaussian likelihood function via the prediction error decomposition. The fixed system matrices contain some unknown coefficients which are estimated by numerically maximizing the log-likelihood function using a quasi-Newton method. The standard maximum likelihood (ML) theory applies to this estimation process. For example, we obtain the standard errors of the ML estimates from the Fisher information matrix. Given the system matrices, with the unknown coefficients replaced by their ML estimates, the observation prediction errors from the Kalman filter are used to validate the adequacy of the distribution assumption by examining diagnostic statistics based on standardized prediction errors; see Section 4 and Table 1. Finally, we obtain the smoothed MMSE of α_t, given all observed data, with its MSE variance matrix using a Kalman filter smoothing method. Further details of this estimation methodology are provided in Durbin and Koopman (2012, Ch. 4 and 7).

2.7 Model-implied dynamics of atmospheric concentrations

We study the dynamic properties of atmospheric concentrations in the linear model. To this end, we ignore for a moment the dependence of the sinks on ENSO, of emissions on ECON, and all dummy effects. This simplifies the derivations without changing the main conclusions drawn here. For emissions, we introduce a constant drift $d > 0$, which can be thought of as the average annual increase in emissions due to economic activity over the period 1959–2020. Thus, we let $E^*_{t+1} = E^*_t + d + X^E_t$, which implies that E^*_t is the sum of the linear trend function $E^*_0 + d\,t$ and the stochastic process $X^S_t = \sum_{j=0}^{t-1} X^E_j$, for $t = 1, \ldots, T$. In the terminology adopted in the dynamic econometrics literature, we have $X^S_t \sim I(1)$, i.e. the process X^S_t is non-stationary and integrated of order one. It implies that the first difference $X^S_{t+1} - X^S_t = X^E_t$ is a stationary process; see Hamilton (1994, p. 437) for a textbook treatment. By inserting the equation for E^*_t, (15), and (16) into the budget equation (1), we can represent C^*_t as a first-order autoregressive process,

$$
\begin{aligned}
C^*_{t+1} &= C^*_t + d + E^*_t + X^E_t - \beta C^*_{t+1} - c_1 - c_2 \\
&= C^*_t + c + d\,t + X^S_t + X^E_t - \beta C^*_{t+1},
\end{aligned}
$$

where $c = d + E^*_0 - c_1 - c_2$ and $\beta = (\beta_1 + \beta_2)/C_{1750}$. Note that while these derivations rely on the linearity of the sinks, we can actually gain insights into their nonlinear behavior by assessing the quality of the approximation, as we discuss below.

Re-arranging terms, we obtain $(1+\beta)C^*_{t+1} = C^*_t + c + d\,t + X^S_{t+1}$ and hence the first-order difference equation

$$
C^*_{t+1} = \delta C^*_t + c^* + d^* t + \delta X^S_{t+1},
$$

where $\delta = (1+\beta)^{-1}$, $c^* = \delta\, c$, and $d^* = \delta\, d$. Solving the difference equation yields

$$C_t^* = \delta^t \left[C_0 - \frac{c^*}{1-\delta} + \frac{d^*}{(1-\delta)^2} \right] + \left[\frac{c^*}{1-\delta} - \frac{d^*}{(1-\delta)^2} \right] + \frac{d^*}{1-\delta} t + \sum_{j=0}^{t-1} \delta^{j+1} X_{t-j}^S \tag{18}$$
$$\sim o(1) + O(1) + O(t) + I(1) \sim O(t) + I(1),$$

where for sequences a_t, b_t, we write $a_t \sim o(b_t)$ if $a_t/b_t \to 0$ as $t \to \infty$, and $a_t \sim O(b_t)$ if $a_t/b_t \to const$ as $t \to \infty$, where *const* is a finite constant. This solution of the difference equation for C_t^* implies that, in a model where sinks are assumed to depend linearly on concentrations and where emissions are a random walk plus drift, the resulting level of atmospheric concentrations will be a linear time trend plus an $I(1)$ term. Since we model emissions E_t^* as a random walk around a linear trend, and atmospheric concentrations are accumulating the emissions, one might expect concentrations to be I(2). However, there is no second unit root because the sinks in turn take up a fraction of the emissions, which leads to atmospheric concentrations being $O(t) + I(1)$.

This conclusion relies on the fact that the autoregressive coefficient δ satisfies $|\delta| < 1$. In fact, if $\delta = 1$, the term $\sum_{j=0}^{t-1} \delta^{j+1} X_{t-j}^S = \sum_{j=0}^{t-1} X_{t-j}^S$ would be integrated of order two, that is, $I(2)$ instead of $I(1)$. In our empirical study below, we find that $\delta \approx 0.98$. As atmospheric concentrations increase and quadratic terms become important in a Taylor approximation of the nonlinear sink functions, δ grows even closer to one, pushing the process even closer to a second unit root, and the "Keeling curve" (Fig. 1 (a) below) will display more positive curvature. The discussion of the sink rate in Section 5 below is related to this issue, since it shows that the capacity of the sinks to increase uptake in response to an increase in concentrations is diminishing. In the supplementary material S5 we discuss an impulse response function of atmospheric concentrations with respect to emissions based on solution (18).

3 The data set

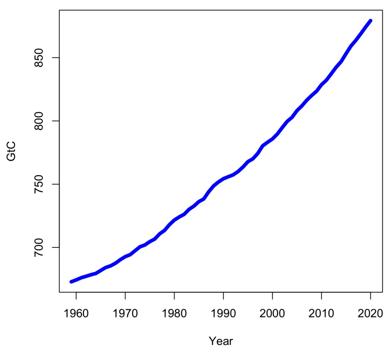
3.1 Global Carbon Budget data 1959–2020

Figure 1 displays the time series data set from the Global Carbon Project that we employ in our study, both in levels and in first differences. The GCB time series are annual, observed from 1959 to 2020, measured in GtC per year, and obtained from the global file of Friedlingstein et al. (2022), available at `https://www.icos-cp.eu/science-and-impact/global-carbon-budget/2021`.

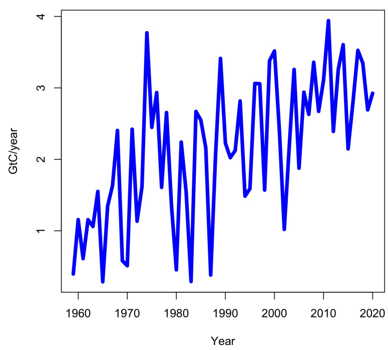
Panel [a] in Figure 1 presents atmospheric concentrations C (Dlugokencky and Tans, 2020), while

Figure 1: **GCB annual time series** 1959 – 2020. C: Atmospheric CO_2 concentrations, E_FF: fossil fuel emissions, E_LUC: land-use change emissions, E: anthropogenic emissions (FF+LUC), S_LND: land sink, S_OCN: ocean sink.

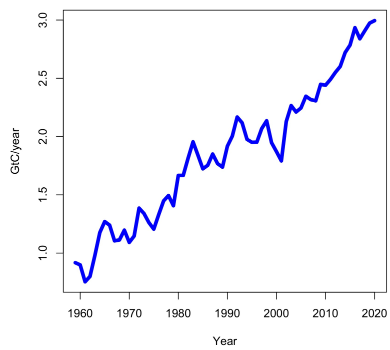
(a) C



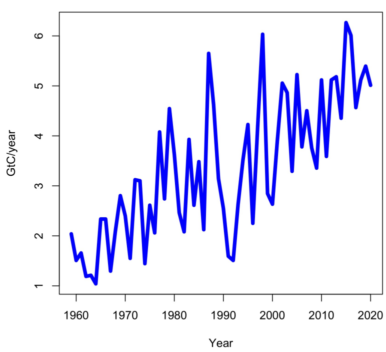
(b) S_LND



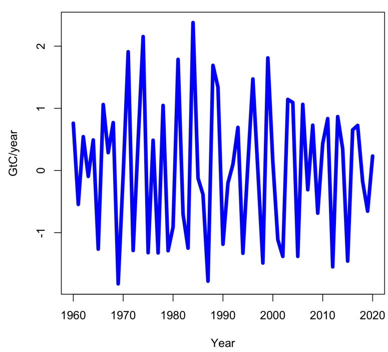
(c) S_OCN



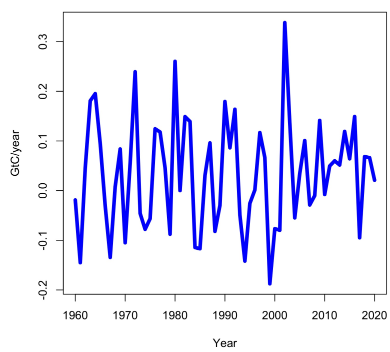
(d) First differences C



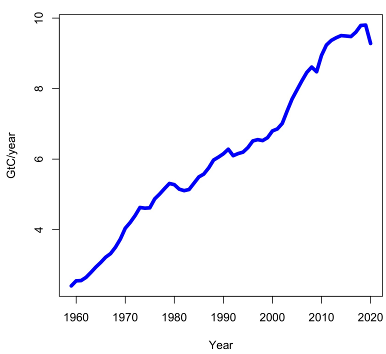
(e) First differences S_LND



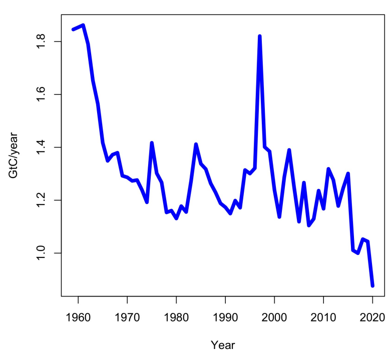
(f) First differences S_OCN



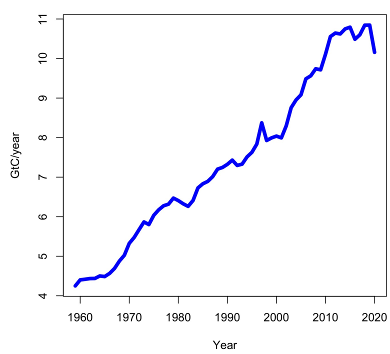
(g) E_FF



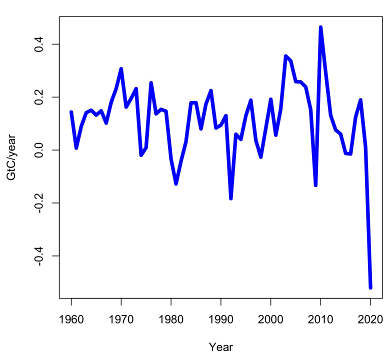
(h) E_LUC



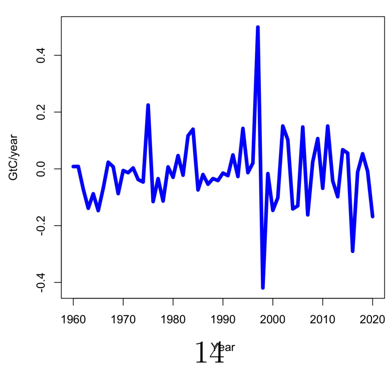
(i) E



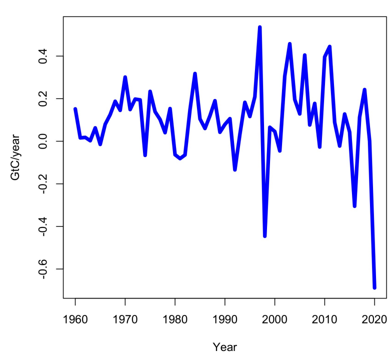
(j) First differences E_FF



(k) First differences E_LUC



(l) First differences E



Panels [b] and [c] show the sinks. The time series of the land sink (S_LND) is the mean of the outputs of 17 different models (Haverd et al., 2018; Melton et al., 2020; Yuan et al., 2014; Lawrence et al., 2019; Tian et al., 2015; Meiyappan et al., 2015; Delire et al., 2020; Mauritsen et al., 2019; Sellar et al., 2019; Smith et al., 2014; Poulter et al., 2011; Lienert and Joos, 2018; Zaehle and Friend, 2010; Vuichard et al., 2019; Walker et al., 2017; Kato et al., 2013; Yue and Unger, 2015), and that of the ocean sink (S_OCN) is the mean of the outputs of 8 different models (Schwinger et al., 2016; Berthet et al., 2019; Hauck et al., 2020; Liao et al., 2020; Doney et al., 2009; Aumont et al., 2015; Wright et al., 2021; Lacroix et al., 2021), which were constrained to surface partial pressure observations following Landschützer et al. (2016); Rödenbeck et al. (2014); Denvil-Sommer et al. (2019); Gregor et al. (2019); Watson et al. (2020); Zeng et al. (2014); Iida et al. (2021); Gregor and Gruber (2021). Panels [d] through [f] contain the first differences of the series. The first difference of C, i.e. $C_t - C_{t-1}$, represents changes in atmospheric concentrations (G_ATM). Panel [g] shows fossil fuel emissions E_FF, which are calculated including cement carbonation from the global file (see Friedlingstein et al., 2020, p. 3277, for a discussion on how to include cement carbonation into the fossil fuel emissions time series). Panel [h] shows land-use change emissions E_LUC, which are the average of three series prepared by Houghton and Nassikas (2017), Hansis et al. (2015), and Gasser et al. (2020). Panel [i] shows the sum of E_FF and E_LUC, labeled anthropogenic emissions E. Panels [j] through [l] show the first differences of the variables immediately above.

As is clear from Figure 1, the most conspicuous dynamic property of the GCB data series over the period 1959–2020 is that they are trending upwards. An exception is E_LUC, which has been hovering around a constant value for most of the period, with a downward trend since 2016. To shed some light on the statistical properties of the data series, we have run a battery of tests for stationarity and unit roots of these variables; the results are discussed in the supplementary material S6.

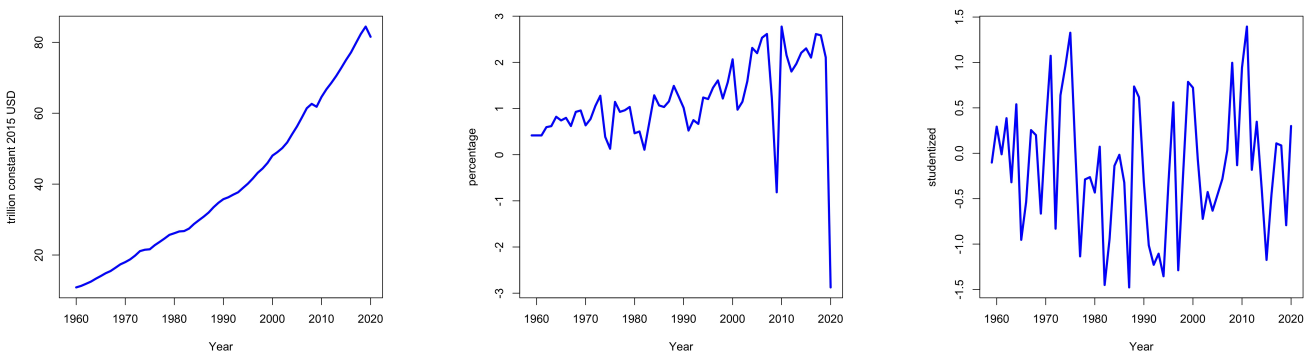
3.2 Explanatory variables

Figure 2 displays the time series that act as explanatory variables. The ECON variable is taken as the logarithm of gross domestic product world-wide (World GDP) in constant 2015 Dollars, which is obtained from World Bank (2021) (Series ID NY.GDP.MKTP.KD). The year-to-year log-differences of World GDP are also displayed, as we will adopt the growth rate in World GDP (ΔECON) as an explanatory variable for changes in emissions. The ENSO variable is taken as the Southern Oscillation Index (SOI), and it is obtained from Climatic Research Unit (2021); Ropelewski and

Jones (1987). The SOI is defined as the studentized measure of differences in atmospheric pressure at sea level between Tahiti and Darwin, Australia. Positive (negative) values correspond to La Niña (El Niño) phases. We employ SOI as the ENSO explanatory variable for the sink processes. We have also experimented with versions of our model with Niño 3.4 and Oceanic Niño Indices: the resulting estimation results have been very similar. The likely reason for the similar estimation outcomes is that our analysis is based on annual time series and these indices are very similar at the yearly sampling frequency.

Figure 2: **Explanatory variables** 1959 – 2020. World Gross Domestic Product (GDP, constant 2015 USD), in levels and in first log-differences, and the Southern Oscillation Index (SOI)

(a) World GDP (b) World GDP log-differences (c) Southern Oscillation Index



4 Maximum likelihood estimation and residual diagnostics

In this section, we present the maximum likelihood estimates of the parameters in our GCB model as proposed in Section 2.6, using the GCB data set as presented in Section 3. First, we discuss the estimation results for the nonlinear model of Section 2.6. Although we specify the sink coefficients $\beta_{1,t}$ and $\beta_{2,t}$ as time-varying, we find that they are estimated to be constant over time on the sample. This provides justification of the linear model of Section 2.6. We estimate this model and present and discuss the goodness-of-fit and the residual diagnostics.

A simulation study examining the performance of our estimation procedure when applied to both the nonlinear and the linear specifications can be found in the supplementary material S1 and S7, respectively. There we find that the time-varying parameter approach to accommodate the nonlinearity of the general GCB model can capture various nonlinear dependencies in the sink-concentrations relationship, if they are present. We further find good finite sample properties of the estimation procedure as applied to the linear model.

4.1 Estimation results for nonlinear model

We specify the time-varying parameters in equations (6), (7), and (8) as independent random walk processes. Thus, we have $\beta_{i,t+1} = \beta_{i,t} + \omega_{i,t}$, for $i = 1, 2, 5$, and $t = 1, \ldots, T$, where $\omega_{i,t}$ is an independently and identically distributed random innovation variable with mean zero and variance $\sigma^2_{\beta_i} > 0$, and it is mutually independent from all other innovations in the model, for $i = 1, 2, 5$. We note that when one (or several) of these variances $\sigma^2_{\beta_i} > 0$, $i = 1, 2, 5$, is estimated as zero, the corresponding time-varying coefficient reduces to a constant. In the cases of $\beta_{1,t}$ and $\beta_{2,t}$ in the sinks equations, constancy or time-variation determines the linearity or nonlinearity of the model.

The nonlinearity of this setup obtains because the time-varying variables $\beta_{1,t}$ and $\beta_{2,t}$ are multiplied by the dynamic stochastic state variable C^*_t. Hence, we cannot treat this system directly via the Kalman filter and smoother methods, which are valid only for linear state space systems. Instead, we address this multiplicativity by means of the Extended Kalman filter and smoother (EKF); see (Durbin and Koopman, 2012, pp. 226–237). Maximum likelihood estimation and residual diagnostic checking are not affected, as long as the EKF is used for filtering and smoothing. In contrast, in equation (8), $\beta_{5,t}$ is multiplied by $ECON_t$, which is not a state process and therefore does not entail nonlinearity.

We estimate the parameters of the resulting model using maximum likelihood, with the log-likelihood function evaluated by the EKF. The variances $\sigma^2_{\beta_1}$, $\sigma^2_{\beta_2}$, and $\sigma^2_{\beta_5}$ are estimated to be effectively zero, at 1e-10 (2.4e-10), 7e-10 (3e-7), and 2e-6 (5e-4), respectively, where the numbers in parentheses are standard errors. On the GCB data 1959-2020, the linearity assumption is therefore still accurate enough for statistical purposes.

The estimates of the remaining parameters are very close to those from the linear specification presented in the next section. The residual diagnostics and the smoothed estimates of the state vector are also very similar to those of the linear model. The full set of results for the nonlinear specification, together with plots of the smoothed processes for the time-varying coefficients, is reported in the supplementary material S8.

These results show that the linear model specification is appropriate for the sample period studied here. It should be stressed that the relatively short sample of yearly observations from 1959 to 2020 (62 observations) and the range of atmospheric concentrations observed during the sample period can make it challenging to empirically establish nonlinear sinks effects.

4.2 Parameter estimation for the linear multivariate dynamic model

The state space representation of the linear dynamic model is provided in detail in Section 2.6. This representation facilitates the estimation of the parameters in the model by maximum likelihood using linear Gaussian state space methods as discussed in, for example, Durbin and Koopman (2012). The state vector contains the dynamic (stationary and non-stationary) features of the model and the linear regression effects. The stationary elements of the state vector (X_t^E and $X_{i,t}$ for $i = 1, \ldots, 4$) are initialized based on corresponding unconditional moments while the non-stationary elements are subject to diffuse initial conditions. The state equation disturbance vector is given by $\eta_t = (\eta_{1,t}, \ldots, \eta_{4,t}, \kappa_t)'$ where $\eta_{i,t}$ corresponds to the disturbance in the autoregressive processes $X_{i,t}$, for $i = 1, \ldots, 4$, while κ_t is the disturbance in the autoregressive process X_t^E. Letting $r_{ij} = Corr(\eta_{i,t}, \eta_{j,t})$, we assume that $r_{ij} = 0$ for all $i \neq j$, except for r_{12} (correlation between residual innovations in C and S_LND) and r_{13} (correlation between residual innovations in C and S_OCN). We experimented with non-zero values for other correlations r_{ij}, but this resulted in estimates that were close to zero and insignificant. Initial analyses show that we can set the measurement error for E in (12) to zero, that is $X_{4,t} = 0$ and hence $\phi_4 = \sigma^2_{\eta_4} = 0$. Further, the measurement error for the land sink $X_{2,t}$ does not show any evidence of serial correlation, and hence we set $\phi_2 = 0$.

The resulting 12×1 parameter vector ψ is given by

$$\psi = (\beta_1, \beta_2, \phi_1, \phi_3, \phi_E, \sigma^2_{\eta_1}, \sigma^2_{\eta_2}, \sigma^2_{\eta_3}, \sigma^2_{\kappa}, r_{12}, r_{13}, s_E)'.$$

The estimation of ψ is based on maximizing the log-likelihood function that is evaluated by an augmented Kalman filter in order to account for the diffuse initial conditions in the state vector (Durbin and Koopman, 2012, p. 173). The remaining coefficients c_1, c_2 and β_j, $j = 3, \ldots, 8$, are added to the state vector. These coefficients are incorporated in the state space framework by treating them as states with transition equation, e.g., $c_{1,t+1} = c_{1,t}$ without error, rendering them constant states. In effect, this approach concentrates these coefficients out of the likelihood.

Table 1 displays the estimated parameter values and their standard errors. Figure 3 presents the smoothed states, together with the time series data of C, G_ATM, E, S_LND and S_OCN. The coefficients pertaining to SOI in the sinks processes are highly significant. They are of opposite sign: La Niña phases (positive SOI) correspond to higher land uptake whereas they correspond to lower ocean uptake. This aligns with expectations from physical considerations, see, for example, Feely et al. (1999) and Haverd et al. (2018). The increase in variance of κ_t in 1996, captured by $(\hat{s}_E)^2 =$

Figure 3: **Smoothed estimates** for state vector elements: C_t^* atmospheric concentrations, $X_{1,t}$ error process for C_t, $G\_ATM_t^*$ first differences of C_t^*, $S\_LNS_t^*$ land sink, $X_{2,t}$ deviations process for S_LND_t, $S\_OCN_t^*$ ocean sink, $X_{3,t}$ deviations process for S_OCN_t, E_t^* emissions, X_t^E innovations process that is cumulated in E_t^*.

(a) C_t^* (b) $X_{1,t}$ (c) $G\_ATM_t^* = C_{t+1}^* - C_t^*$

(d) $S\_LND_t^*$ (e) $X_{2,t}$ (f) $S\_OCN_t^*$

(g) $X_{3,t}$ (h) E_t^* (i) $X_{E,t}$

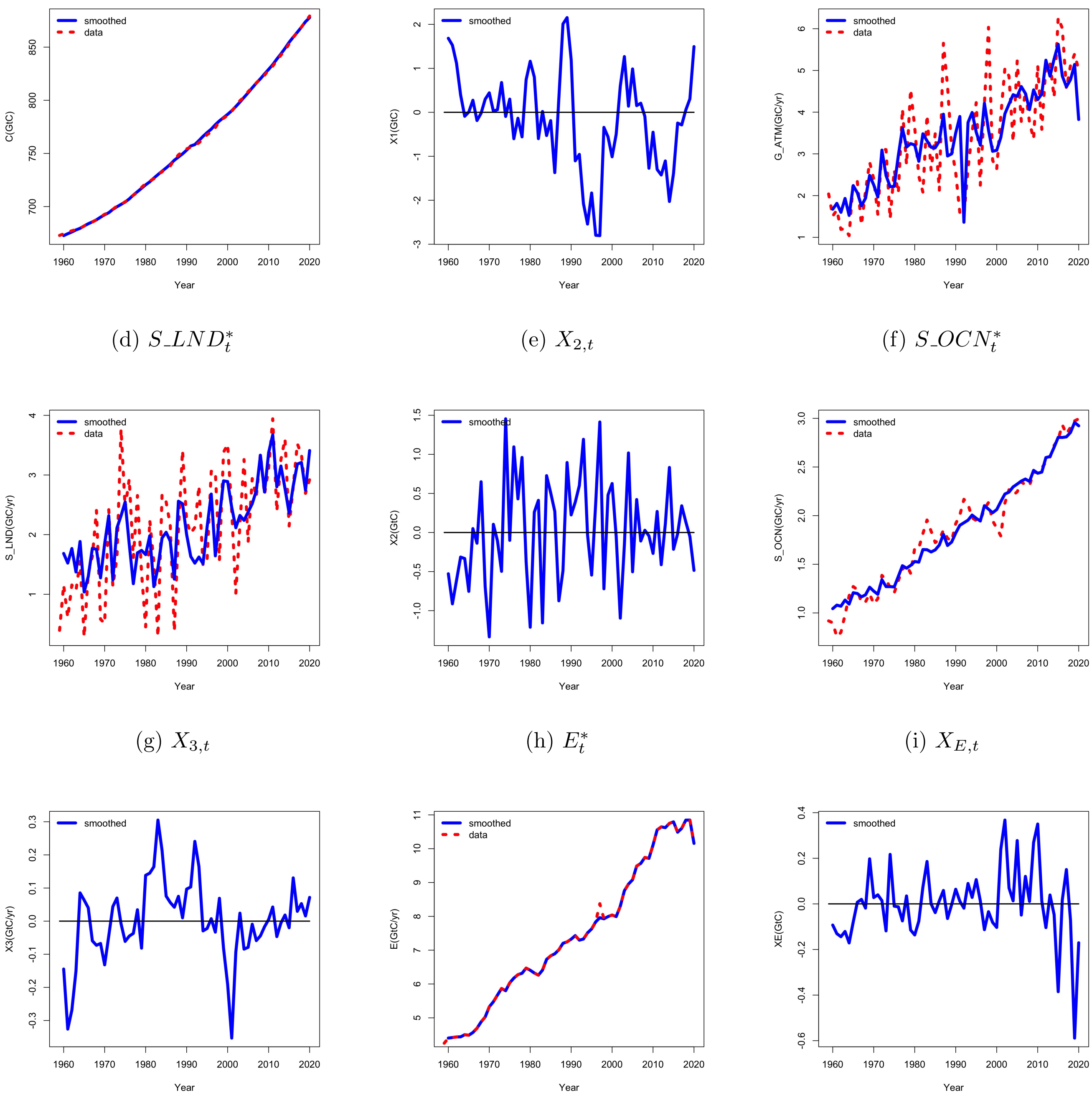


Table 1: **Parameter estimation results**

The statistical GCB model is for $y_t = (C, S\_LND, S\_OCN, E)'_t$ and is discussed in Section 2. The parameter estimates are provided with their asymptotic standard errors in parantheses below. The "linear" parameters are placed in the state vector, and they are effectively concentrated out from the likelihood function: their estimates and the corresponding standard errors (in brackets, below the estimates) are computed by the Kalman filter recursions. The "other" and "variance" parameters are placed in the parameter vector; its estimate and the corresponding asymptotic standard errors are obtained from numerically maximizing the log-likelihood function.The residual diagnostics are for the standardized prediction residuals for the four variables in y_t, obtained from the Kalman filter. We report sample statistics together with the Jarque-Bera test for normality, the Ljung-Box (1 lag), and the Durbin-Watson statistics for first-order autocorrelation.

Parameter estimates

	Linear parameters		Other parameters		Variance parameters
c_1	-4.13	β_1	4.98	$\sigma^2_{\eta_1}$	0.62
	(0.04)		(0.45)		(0.12)
c_2	-5.11	β_2	5.44	$\sigma^2_{\eta_2}$	0.42
	(0.03)		(0.30)		(0.08)
β_3	0.58	ϕ_1	0.75	$\sigma^2_{\eta_3}$	0.008
	(0.10)		(0.10)		(0.001)
β_4	-0.06	ϕ_3	0.68	σ^2_{κ}	0.009
	(0.02)		(0.10)		(0.002)
β_5	2.89	ϕ_E	0.29	r_{12}	-0.58
	(0.50)		(0.14)		(0.09)
β_6	0.41			r_{13}	0.03
	(0.08)				(0.11)
β_7	-2.49			s_E	2.24
	(0.66)				(0.44)
β_8	-0.21				
	(0.09)				

Residual Diagnostics

	C	E	S_LND	S_OCN
mean	-0.014	0.040	0.071	0.099
standard deviation	0.995	0.979	0.991	0.944
skewness	0.070	0.150	0.030	0.070
kurtosis	2.731	3.558	2.607	3.321
Ljung-Box	0.000	0.099	0.198	0.015
Jarque-Bera	0.237	1.035	0.409	0.316
Durbin-Watson	1.976	1.914	2.095	2.025

2.24^2, is highly significant and more than quadruples the pre-1996 variance (see supplementary material S2). The dummy variables are all of the expected sign. Figure 3 indicates that the inclusion of SOI in the sinks makes them dynamic, tracing the data better than a simple linear function of C^* would be able to.

The coefficient pertaining to $\Delta ECON_t$ is highly significant, $\hat{\beta}_5 = 2.89$ (0.50), indicating that World GDP growth plays an important role in determining changes in emissions. To further investigate the effect of including/excluding logarithmic GDP in the model, we also considered a version of the model where we add a constant drift d to the equation for E^*, that is

$$E^*_{t+1} \;=\; E^*_t + d + \beta_5 \Delta \log GDP_{t+1} + \beta_8 I1991 + X^E_t.$$

From this specification, we obtain the insignificant estimate $\hat{d} = -0.027$ (0.043) for the constant d and a significant estimate $\hat{\beta}_5 = 3.480$ (1.038) for the growth rate of World GDP. We can conclude that the data clearly prefer the time-varying drift $\beta_5 \Delta \log GDP_{t+1}$ over the constant drift d, in the random walk for E.

The residual diagnostics presented in Table 1 do not reveal evidence of non-normality remaining in the standardized prediction residuals, nor evidence of serial correlation. Hence, we can conclude that the model under consideration provides a good statistical description of the data. If the standardized prediction residuals are statistically indistinguishable from white noise, as the residual diagnostics show, additional terms in the model will not be able to substantially improve on the fit of the model to the data. Even though we know some of the residual variation to be physically meaningful, for example in the sinks data, since they were generated by (and averaged over) different GCMs, from a statistical point of view, and on the data set we consider, it is sufficient to describe the sinks as linear functions of atmospheric concentrations and SOI. Similarly, it is sufficient to describe changes in emissions as a linear function of World GDP growth, and it is sufficient to describe changes in atmospheric concentrations by the carbon budget equation (plus the various dummies that we include in the model).

Canadell et al. (2021) report on p. 691 that the ocean sink evaluated from global ocean biogeochemical models grew from 1.0 ± 0.3 GtC per year in 1960-1969 to 2.5 ± 0.3 GtC per year in 2010-2019. Average atmospheric concentrations in the 1960s were 681 GtC and in the 2010s 850 GtC. This implies a rough estimate of β_2 at $(2.5 \pm 0.3 - 1.0 \pm 0.3)/(850 - 681) = 5.27$ with an uncertainty range of [3.05, 7.48], if we ignore uncertainty in the measurement of atmospheric concentrations. On p. 694, they report that over the same period, the land sink increased from 0.3 ± 0.6 GtC to 1.8 ± 0.8 GtC. In the same fashion, we arrive at a rough estimate of β_1 of 5.27 with an uncertainty range of [0.35, 10.18]. Gloor et al. (2010) report estimates of the reciprocals of these coefficients, implying slightly higher values at $\beta_1 = 7.29$ and $\beta_1 = 8.54$. The difference in the estimates presented in Gloor et al. (2010) to those of Table 1 and the ones implied in Canadell et al. (2021) are likely due to the longer sample of data available now, as well as improvements in the data series themselves, stemming from recent developments in the construction of historical sinks time series using GCMs and in reconstructing historical estimates of CO_2 emissions (Friedlingstein et al., 2022).

The estimate of $\hat{\phi}_3 = 0.68(0.10)$ reflects a predictability in the ocean sink that has been discussed in the large-scale model literature (Sabine et al., 2004; McKinley et al., 2017). It likely reflects internal variability driven by the Southern Annular Mode, the Pacific Decadal Oscillation, and the North Atlantic Oscillation, see (McKinley et al., 2017, p. 127).

The supplementary material S9 presents a validation exercise, where we estimate the model on the subsample 1959–2010. Given observations on SOI and on World GDP for the validation sample 2011–2020, we compute the imputed values for the four variables of the model on this validation sample and compare them to the actual outcome. This exercise also illustrates that the model describes the data well.

5 A statistical analysis of the Global Carbon Budget

In this section we provide various illustrations of a statistical analysis of the Global Carbon Budget based on the multivariate dynamic statistical framework. We start with the statistical forecasting of the variables in the GCB. Furthermore, we show that other key measures, such as the budget imbalance, the airborne fraction, and the sink rate, can be extracted and analyzed using our statistical framework.

5.1 Forecasting

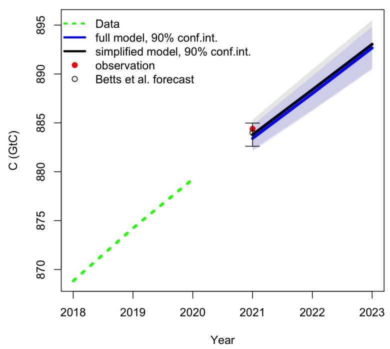
Recently, there has been a focus on forecasting the global carbon cycle at short horizons of a few years (Li and Ilyina, 2018; Séférian et al., 2018; Lovenduski et al., 2019b,a; Spring and Ilyina, 2020). These efforts have been based on Earth System modeling. Li et al. (2022) connect this approach with the Global Carbon Budget data set in order to arrive at predictions of observational data. Betts et al. (2016) and Betts et al. (2018) propose a statistical forecast model for ΔC as a function of emissions and ENSO3.4 sea-surface temperature anomalies.

We approach the forecasting problem with the statistical GCB model, which allows us to forecast all four components (C, E, S_LND, S_OCN) simultaneously subject to the global carbon budget equation. First differences $G\_ATM = \Delta C$ are automatically included. Given the data sample from 1959 to 2020, we forecast the years 2021 to 2023.

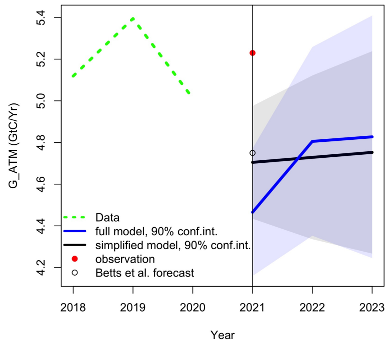
Forecasting the model necessitates forecasts for World GDP growth and for the SOI for these years. We employ the GDP growth rate of 6.1% for 2021 and projections 3.2% and 2.9% for 2022 and 2023, respectively, from the IMF World Economic Outlook July 2022 (IMF, 2022).

Figure 4: **Forecasts for C, G_ATM, E, S_LND, and S_OCN** with 90% pointwise confidence intervals. The model employs forecasts of SOI generated from a separate model (see supplementary material S10) and of World GDP growth from the IMF. The full model (with forecasts in blue and pointwise confidence intervals in light blue shade) is specified in Section 2. The simplified model (with forecasts in dark grey and pointwise confidence intervals in light gray shade) does not feature SOI in the sinks nor World GDP in emissions. For 2021 and for C and G_ATM, a forecast of a model inspired by Betts et al. (2018) is shown for comparison (with confidence interval, black vertical lines in panels (a) for C and (b) for G_ATM); for E, S_LND, and S_OCN, the forecast of the Global Carbon Project is shown for comparison (without confidence interval).

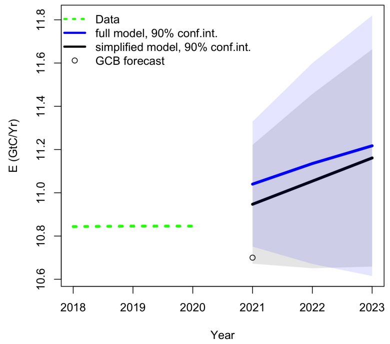
(a) C



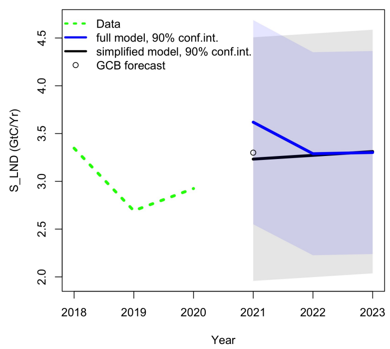
(b) G_ATM



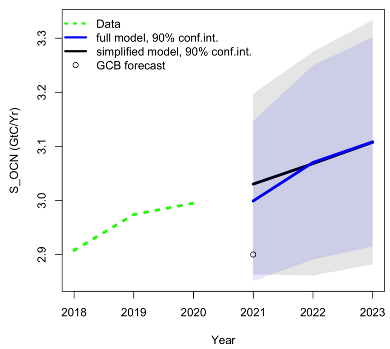
(c) E



(d) S_LND



(e) S_OCN



The forecasts of monthly SOI are based on a historical data set obtained from the Climatic Research Unit (2021) and from Ropelewski and Jones (1987). The model-based monthly forecasts from September 2021 to December 2023 are obtained from a structural time series model with a level component, a monthly seasonal component, a second-order stochastic cycle (with the cycle-period being estimated close to 4 years), and a first-order autoregressive component; see Harvey (1989) for a textbook treatment. A similar model is adopted in Petrova et al. (2017). Further details of the data, the time series model, and the estimation results are presented in the supplementary material S10. The SOI is predicted to switch from a La-Niña period (positive numbers) in 2021 and 2022 to an El-Niño period (negative numbers) in 2023.

Due to the COVID pandemic, we are in a forecasting situation where there is an outlier at the end of the sample. In future updates of the model, a dummy in the emissions equation will likely have to be added for this event. At this point in time, such a dummy makes no substantial difference for the parameter estimates, but the forecasts are affected. In particular, there is little value in a forecast that uses the unusually low 2020 emissions as the best forecast for 2021. Therefore, only for the purposes of this forecast exercise, we do not use the 2020 observation for emissions but repeat the 2019 observation instead, so that the 2021 forecast (and subsequent ones) are based on pre-pandemic emission levels. This effectively dummies the COVID year 2020 out for the forecast.

Figure 4 presents the forecasts of the GCB variables together with estimates of the forecast uncertainty by means of the 90% point-wise confidence bands (blue, and light blue shades for pointwise confidence intervals). As a benchmark, we also show the forecasts from a simplified version of the model that does not contain SOI in the sinks, World-GDP in emissions, or any dummy variables (black, and gray shades for pointwise confidence intervals).

The Global Carbon Project provides projections of the GCB variables for one year following the current vintage. These are not forecasts proper for all variables, more recent information about the variables enters. For increases in atmospheric concentrations, for example, actual observations are used (Friedlingstein et al., 2022, p. 1941). For the sinks, these are proper forecasts from a neural network prediction. The projections for 2021 are: $E_{2021} = 10.7$ GtC (our forecast: 11.04 ± 0.35), $G_ATM_{2021} = 5.3$GtC (our forecast: 4.47 ± 0.37 GtC), $S\_OCN_{2021} = 2.9$ GtC (our forecast: 3.0 ± 0.18 GtC), $S\_LND_{2021} = 3.3$ GtC (our forecast: 3.62 ± 1.30 GtC). Note that the Global Carbon Project thus projects a budget imbalance of -0.80 GtC, whereas in our model, the forecast of the budget imbalance is -0.04. For C and G_ATM, we compare with the available observation for 2021 and with the interval forecast from the model of Betts et al. (2016, 2018), equipped with our own forecast for E_{2021}. The Betts et al. forecast can only be performed for one year ahead, unless one produces forecasts of E and the ENSO SST anomaly for several years ahead. For the sinks forecasts, the current La-Niña phase (positive SOI index numbers) in the full model leads to higher predicted land sink activity in 2021 ($\beta_3 > 0$) and to lower ocean sink activity ($\beta_4 < 0$) in 2021 compared to the base model. This results in a forecast for G_ATM from the full model that is lower than that from the base model, despite the high world-GDP growth rate in 2021.

5.2 Budget Imbalance

Friedlingstein et al. (2022) define and discuss the budget imbalance (BIM), which is the residual of the GCB equation. The budget imbalance can be used to assess whether the different GCB data sources are internally consistent. It has also been suggested recently that it may be used to detect potential misreportings of future global CO_2 emissions (Peters et al., 2017; Bennedsen, 2021). Hence, the accurate modeling and prediction of the budget imbalance is an important endeavour. In our model, the BIM variable is defined as $BIM^* = E^* - G\_ATM^* - S\_LND^* - S\_OCN^* = 0$ with corresponding measurement

$$\begin{aligned} BIM_t &= E_t - \Delta C_t - S\_LND_t - S\_OCN_t \\ &= E_t^* - G\_ATM_t^* - S\_LND_t^* - S\_OCN_t^* + \beta_6 I1997 - \beta_7 I1991 - (X_{1,t} - X_{1,t-1}) - X_{2,t} - X_{3,t} \\ &= \beta_6 I1997 - \beta_7 I1991 - \Delta X_{1,t} - X_{2,t} - X_{3,t}. \end{aligned}$$

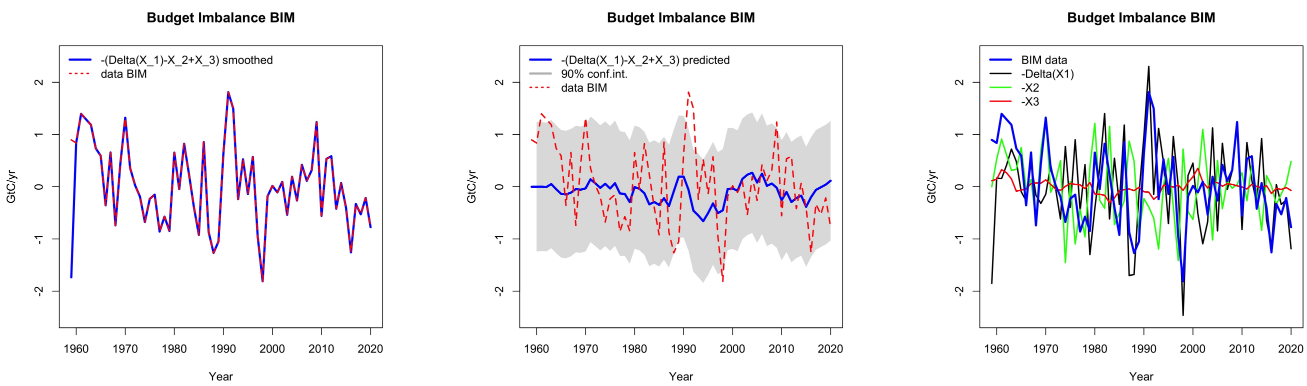
Hence, the BIM observation variable implied by the model is a stationary process and two outlier dummies. Panel (a) in Figure 5 presents the smoothed estimates of BIM_t together with its data counterpart. The perfect overlap shows that the system of model equations is internally consistent.

Figure 5: **Budget imbalance** BIM, smoothed values, one-year ahead predictions, and components

(a) Smoothed $-(\Delta X_1 + X_2 + X_3) + \beta_6 I1997 - \beta_7 I1991$

(b) One-year ahead predictions

(c) Components $-\Delta X_1$, $-X_2$, $-X_3$



Panel (b) in Figure 5 presents the one-year ahead predictions for BIM obtained from the in-sample one-year ahead predictions of $\Delta X_{1,t} + X_{2,t} + X_{3,t}$, together with 90% pointwise confidence intervals. The individual predictions for $-\Delta X_{1,t}$, $-X_{2,t}$, and $-X_{3,t}$ are presented in Panel (c) of Figure 5. We learn that most of the variation originates from $-\Delta X_{1,t}$, the measurements error process in atmospheric concentrations, and $-X_{2,t}$, the approximation error process in the land sink. A variance decomposition reveals that they jointly contribute 97% (75% and 22%, respectively).

5.3 Airborne fraction and sink rate

The airborne fraction is defined as

$$AF = \frac{G\_ATM}{E},$$

and is the part of emissions that remains in the atmosphere. On the sample, this fraction is on average 0.44, but it shows substantial variation in the data. The sink rate

$$SR = \frac{S\_LND + S\_OCN}{C},$$

is a measure of the capacity of the sinks to absorb atmospheric CO_2. Figure 6 presents the airborne fraction in panel (a) and the sink rate in panel (b). Whether or not the airborne fraction is increasing and/or the sink rate is decreasing has been the subject of much debate in the last fifteen years (Canadell et al., 2007a; Raupach et al., 2008; Knorr, 2009; Le Quéré et al., 2009; Gloor et al., 2010; Raupach et al., 2014; Rayner et al., 2015; Ballantyne et al., 2015; Bennedsen et al., 2019).

Figure 6: **Airborne fraction and sink rate**, smoothed estimates from the model and a simplified version that excludes SOI from the sinks and World GDP from emissions, with the corresponding observations for comparison

(a) Airborne fraction (b) Sink rate

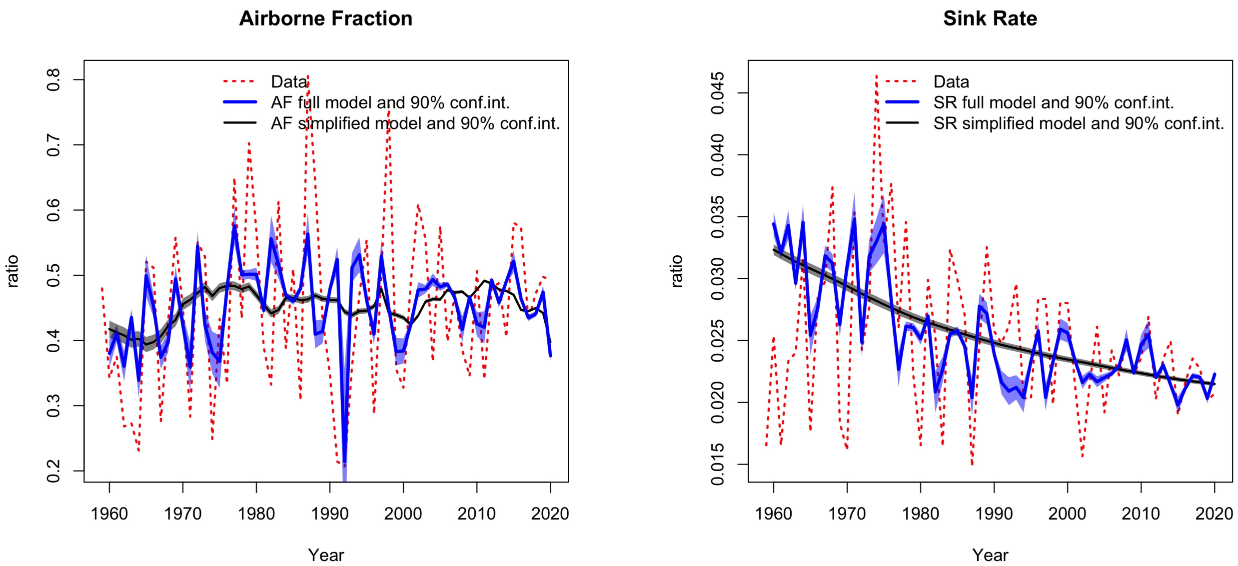


Figure 6 shows, however, that by calculating AF^* and SR^*, the versions of the ratios with state processes in the numerators and denominators, we obtain new estimates of these variables that focus on a few key relations of the sinks, and thus they show much less variation. Raupach et al. (2014) and Bennedsen et al. (2019) show that there is no evidence of an increase in the airborne fraction, but that there is evidence of a decline in the sink rate. In Figure 6 this can be seen with the naked eye, in particular for the results from the simplified benchmark model without SOI in the sinks, which specifies the sink processes just as affine transformations of atmospheric concentrations. The figure also shows confidence intervals obtained from simulation smoothing (Durbin and Koopman, 2002). Note that these confidence intervals are not relative to the data but should be interpreted as follows. If we were given a large number of trajectories sampled from the models, and if we were to extract the smoothed state variables and compute the airborne fraction and the sink rate from these trajectories, the confidence bands will cover these imputed variables point-wise 90% of the time.

6 Conclusions and directions for further research

We have proposed a multivariate dynamic statistical model for the global carbon budget, consisting of the time series variables: atmospheric CO_2 concentrations, anthropogenic CO_2 emissions, and land and ocean CO_2 uptake (sinks). The cornerstone of the model is the budget equation, which ensures that the fraction of emissions that is not absorbed by the terrestrial biosphere or the ocean constitutes an annual flow to the stock of atmospheric concentrations. We discussed the central assumptions of random walk with drift dynamics dependent on economic activity for anthropogenic emissions and nonlinear and linear dependence of sinks on atmospheric concentrations. The model equations allow for a closed-form solution for atmospheric concentrations; they reveal stochastic integration of order one with a second near-unit root for concentrations. The dynamics of atmospheric concentrations approach a second unit root as sinks uptake degrades.

We presented a comprehensive statistical analysis of the global carbon budget data set, as provided by the Global Carbon Project, including model parameter estimates, residual diagnostics, and smoothed estimates of the model variables. Based on our multivariate dynamic statistical model, we produced forecasts for all GCB variables in the years 2021, 2022, and 2023, we discussed the airborne fraction and sink rate estimates, and we decomposed the variation in the budget imbalance into contributions from concentrations, land sink, and ocean sink.

We plan several directions for extending the work on our multivariate dynamic statistical model. For example, it is conceptually straightforward to include individual ensemble members rather than the averages of them for S_LND and S_OCN. The model structure is also conducive for increasing the resolution on the macroeconomic sphere. For example, we can replace World GDP in emissions by factors obtained from large macroeconomic data sets. The mechanics of the global carbon cycle can be extended to include elements of widely used small-scale climate models. Finally, an interesting extension is to incorporate energy balance modules in our multivariate dynamic model in order to provide a modeling connection to global temperatures.

References

Angert, A., Biraud, S., Bonfils, C., Buermann, W., and Fung, I. (2004). CO2 seasonality indicates origins of post-Pinatubo sink. *Geophysical Research Letters*, 31(11).

Aumont, O., Éthé, C., Tagliabue, A., Bopp, L., and Gehlen, M. (2015). PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development Discussions*, 8(2).

Bacastow, R. and Keeling, C. D. (1973). Atmospheric carbon dioxide and radiocarbon in the natural cycle: II. changes from A. D. 1700 to 2070 as deduced from a geochemical model. In *Carbon and the biosphere conference proceedings; Upton, New York, USA*, pages 86–135. Brookhaven Symposia in Biology.

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.

Barsky, R. and Kilian, L. (2004). Oil and the macroeconomy since the 1970s. *Journal of Economic Perspectives*, 18(4):115–134.

Bennedsen, M. (2021). Designing a statistical procedure for monitoring global carbon dioxide emissions. *Climatic Change*, 166(3):32.

Bennedsen, M., Hillebrand, E., and Koopman, S. (2019). Trend analysis of the airborne fraction and sink rate of anthropogenically released CO2. *Biogeosciences*, 16(18):3651–3663.

Bennedsen, M., Hillebrand, E., and Koopman, S. (2021). Modeling, Forecasting, and Nowcasting U.S. CO2 Emissions Using Many Macroeconomic Predictors. *Energy Economics*, 96:105118.

Berthet, S., Séférian, R., Bricaud, C., Chevallier, M., Voldoire, A., and Ethé, C. (2019). Evaluation of an online grid-coarsening algorithm in a global eddy-admitting ocean biogeochemical model. *Journal of Advances in Modeling Earth Systems*, 11(6):1759–1783.

Betts, R. A., Jones, C. D., Knight, J. R., Keeling, R. F., and Kennedy, J. J. (2016). El niño and a record co2 rise. *Nature Climate Change*, 6(9):806–810.

Betts, R. A., Jones, C. D., Knight, J. R., Keeling, R. F., Kennedy, J. J., Wiltshire, A. J., Andrew, R. M., and AragÃ£o, L. E. O. C. (2018). A successful prediction of the record co¡sub¿2¡/sub¿ rise associated with the 2015/2016 el niño. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1760):20170301.

Bousquet, P., Peylin, P., Ciais, P., Le Quéré, C., Friedlingstein, P., and Tans, P. P. (2000). Regional changes in carbon dioxide fluxes of land and oceans since 1980. *Science*, 290(5495):1342–1346.

Canadell, J.G., Monteiro, P., Costa, M., da Cunha, L. C., Cox, P., Eliseev, A., Henson, S., Ishii, M., Jaccard, S., Koven, C., Lohila, A., Patra, P., Piao, S., Rogelj, J., Syampungani, S., Zaehle, S., and Zickfeld, K. (2021). Global carbon and other biogeochemical cycles and feedbacks. In Masson-Delmotte, V., e. a., editor, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Canadell, J., Le Quéré, C., Raupach, M., Field, C., Buitenhuis, E., Ciais, P., Conway, T., Gillett, N., Houghton, R. A., and Marland, G. (2007a). Contributions to accelerating atmospheric CO2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the National Academy of Sciences*, 104(47):18866–18870.

Canadell, J., Pataki, D., Gifford, R., Houghton, R., Luo, Y., Raupach, M., Smith, P., and Steffen, W. (2007b). Saturation of the terrestrial carbon sink. *Terrestrial Ecosystems in a Changing World*, pages 59–78.

Climatic Research Unit, U. o. E. A. (2021). Southern Oscillation Index. `https://crudata.uea.ac.uk/cru/data/soi/`. Accessed: 2020-02-15.

Delire, C., Séférian, R., Decharme, B., Alkama, R., Calvet, J.-C., Carrer, D., Gibelin, A.-L., Joetzjer, E., Morel, X., Rocher, M., et al. (2020). The global land carbon cycle simulated with isba-ctrip: Improvements over the last decade. *Journal of Advances in Modeling Earth Systems*, 12(9):e2019MS001886.

Denvil-Sommer, A., Gehlen, M., Vrac, M., and Mejia, C. (2019). LSCE-FFNN-v1: the reconstruction of surface ocean pCO2. *Geoscientific Model Development*, 12:2091–2105.

Dlugokencky, E. and Tans, P. (2020). Trends in atmospheric carbon dioxide. Technical report, National Oceanic and Atmospheric Administration, Earth System Research Laboratory (NOAA/ESRL).

Doney, S., Lima, I., Feely, R., Glover, D., Lindsay, K., Mahowald, N., Moore, J., and Wanninkhof, R. (2009). Mechanisms governing interannual variability in upper-ocean inorganic carbon system and air–sea CO2 fluxes: Physical climate and atmospheric dust. *Deep Sea Research Part II: Topical Studies in Oceanography*, 56(8-10):640–655.

Doornik, J. (2009). Autometrics. In Castle, J. and Shephard, N., editors, *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, pages 88–121. Oxford University Press.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods.* Oxford University Press, Oxford, 2 edition.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616.

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

Enting, I. and Lassey, K. (1993). Projections of Future CO_2. CSIRO Division of Atmospheric Research Technical Paper no. 27, http://www.cmar.csiro.au/e-print/open/enting_2000e.pdf.

Feely, R. A., Wanninkhof, R., Takahashi, T., and Tans, P. (1999). Influence of El Niño on the equatorial Pacific contribution to atmospheric CO_2 accumulation. *Nature*, 398(6728):597–601.

Friedlingstein, P., Jones, M., and et al. (2019). Global Carbon Budget 2019. *Earth System Science Data*, 11(4):1783–1838.

Friedlingstein, P., Jones, M. W., and et al. (2022). Global Carbon Budget 2021. *Earth System Science Data*, 14:1917–2005.

Friedlingstein, P., O'Sullivan, and et al. (2020). Global Carbon Budget 2020. *Earth System Science Data*, 12(4):3269–3340.

Gasser, T., Crepin, L., Quilcaille, Y., Houghton, R. A., Ciais, P., and Obersteiner, M. (2020). Historical CO2 emissions from land use and land cover change and their uncertainty. *Biogeosciences*, 17(15):4075–4101.

Gifford, R. (1993). Implications of CO2 effects on vegetation for the global carbon budget. In Heimann, M., editor, *The global carbon cycle*, pages 159–199. Springer.

Gloor, M., Sarmienti, J. L., and Gruber, N. (2010). What can be learned about carbon cycle climate feedbacks from the CO2 airborne fraction? *Atmospheric Chemistry and Physics*, 10:7739 – 7751.

Gregor, L. and Gruber, N. (2021). OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth System Science Data*, 13(2):777–808.

Gregor, L., Lebehot, A. D., Kok, S., and Scheel Monteiro, P. M. (2019). A comparative assessment of the uncertainties of global surface ocean CO2 estimates using a machine-learning ensemble (CSIR-ML6 version 2019a)–have we hit the wall? *Geoscientific Model Development*, 12(12):5113–5136.

Hamilton, J. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy*, 91(2):228–248.

Hamilton, J. (1996). This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics*, 38(2):215–220.

Hamilton, J. (2003). What is an oil shock? *Journal of Econometrics*, 113(2):363–398.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press.

Hansis, E., Davis, S., and Pongratz, J. (2015). Relevance of methodological choices for accounting of land use change carbon fluxes. *Global Biogeochemical Cycles*, 29(8):1230–1246.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge.

Hauck, J., Zeising, M., Le Quéré, C., Gruber, N., Bakker, D. C., Bopp, L., Chau, T. T. T., Gürses, Ö., Ilyina, T., Landschützer, P., et al. (2020). Consistency and challenges in the ocean carbon sink estimate for the global carbon budget. *Frontiers in Marine Science*, 7:852.

Haverd, V., Smith, B., Nieradzik, L., Briggs, P., Woodgate, W., Trudinger, C., Canadell, J., and Cuntz, M. (2018). A new version of the CABLE land surface model (Subversion revision r4601) incorporating land use and land cover change, woody vegetation demography, and a novel optimisation-based approach to plant coordination of photosynthesis. *Geoscientific Model Development*, 11:2995–3026.

Houghton, R. and Nassikas, A. (2017). Global and regional fluxes of carbon from land use and land cover change 1850–2015. *Global Biogeochemical Cycles*, 31(3):456–472.

Iida, Y., Takatani, Y., Kojima, A., and Ishii, M. (2021). Global trends of ocean CO2 sink and ocean acidification: an observation-based reconstruction of surface ocean inorganic carbon variables. *Journal of Oceanography*, 77(2):323–358.

IMF (2022). World Economic Outlook Update, July 2022.

Joos, F., Bruno, M., Fink, R., Siegenthaler, U., Stocker, T., Le Quere, C., and Sarmiento, J. (1996). An efficient and accurate representation of complex oceanic and biospheric models of anthropogenic carbon uptake. *Tellus B*, 48(3):397–417.

Joos, F., Prentice, I., Sitch, S., Meyer, R., Hooss, G., Plattner, G.-K., Gerber, S., and Hasselmann, K. (2001). Global warming feedbacks on terrestrial carbon uptake under the intergovernmental panel on climate change (IPCC) emission scenarios. *Global Biogeochemical Cycles*, 15(4):891–907.

Kato, E., Kinoshita, T., Ito, A., Kawamiya, M., and Yamagata, Y. (2013). Evaluation of spatially explicit emission scenario of land-use change and biomass burning using a process-based biogeochemical model. *Journal of Land Use Science*, 8(1):104–122.

Kilian, L. (2008). The economic effects of energy price shocks. *Journal of Economic Literature*, 46(4):871–909.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–69.

Knorr, W. (2009). Is the airborne fraction of anthropogenic CO2 emissions increasing? *Geophysical Research Letters*, 36.

Lacroix, F., Ilyina, T., Mathis, M., Laruelle, G. G., and Regnier, P. (2021). Historical increases in land-derived nutrient inputs may alleviate effects of a changing physical climate on the oceanic carbon cycle. *Global Change Biology*, 27(21):5491–5513.

Landschützer, P., Gruber, N., and Bakker, D. (2016). Decadal variations and trends of the global ocean carbon sink. *Global Biogeochemical Cycles*, 30(10):1396–1417.

Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., et al. (2019). The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, 11(12):4245–4287.

Le Quéré, C., Andrew, R., and et al. (2017). Global Carbon Budget 2017. *Earth System Science Data Discussions*, pages 1–79.

Le Quéré, C., Andrew, R., and et al. (2018). Global Carbon Budget 2018. *Earth System Science Data*, 10:2141–2194.

Le Quéré, C., Raupach, M., Canadell, J., Marland, G., Bopp, L., Ciais, P., Conway, T., Doney, S., Feely, R., Foster, P., Friedlingstein, P., Gurney, K., Houghton, R., House, J., Huntingford, C., Levy, P., Lomas, M., Majkut, J., Metzl, N., Ometto, J., Peters, G., Prentice, I., Randerson, J., Running, S., Sarmiento, J., Schuster, U., Sitch, S., Takahashi, T., Viovy, N., van der Werf, G., and Woodward, F. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, 2:831 – 836.

Le Quéré, C., Rödenbeck, C., Buitenhuis, E. T., Conway, T. J., Langenfelds, R., Gomez, A., Labuschagne, C., Ramonet, M., Nakazawa, T., Metzl, N., Gillett, N., and Heimann, M. (2007). Saturation of the southern ocean CO2 sink due to recent climate change. *Science*, 316(5832):1735–1738.

Lee, C.-C. (2005). Energy consumption and GDP in developing countries: A cointegrated panel analysis. *Energy Economics*, 27(3):415 – 427.

Li, H. and Ilyina, T. (2018). Current and future decadal trends in the oceanic carbon uptake are dominated by internal variability. *Geophysical Research Letters*, 45(2):916–925.

Li, H., Ilyina, T., Loughran, T., Spring, A., and Pongratz, J. (2022). Reconstructions and predictions of the global carbon budget with an emission-driven earth system model. *Earth System Dynamics Discussions*, 2022:1–26.

Liao, E., Resplandy, L., Liu, J., and Bowman, K. W. (2020). Amplification of the ocean carbon sink during El Niños: role of poleward Ekman transport and influence on atmospheric CO2. *Global Biogeochemical Cycles*, 34(9):e2020GB006574.

Lienert, S. and Joos, F. (2018). A Bayesian ensemble data assimilation to constrain model parameters and land-use carbon emissions. *Biogeosciences*, 15(9):2909–2930.

Lovenduski, N. S., Bonan, G. B., Yeager, S. G., Lindsay, K., and Lombardozzi, D. L. (2019a). High predictability of terrestrial carbon fluxes from an initialized decadal prediction system. *Environmental Research Letters*, 14(12):124074.

Lovenduski, N. S., Yeager, S. G., Lindsay, K., and Long, M. C. (2019b). Predicting near-term variability in ocean carbon uptake. *Earth System Dynamics*, 10(1):45–57.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1. 2) and its response to increasing CO2. *Journal of Advances in Modeling Earth Systems*, 11(4):998–1038.

McKinley, G. A., Fay, A. R., Lovenduski, N. S., and Pilcher, D. J. (2017). Natural variability and anthropogenic trends in the ocean carbon sink. *Annual Review of Marine Science*, 9(1):125–150. PMID: 27620831.

Meiyappan, P., Jain, A., and House, J. (2015). Increased influence of nitrogen limitation on CO2 emissions from future land use and land use change. *Global Biogeochemical Cycles*, 29(9):1524–1548.

Melton, J. R., Arora, V. K., Wisernig-Cojoc, E., Seiler, C., Fortier, M., Chan, E., and Teckentrup, L. (2020). CLASSIC v1. 0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM) – Part 1: Model framework and site-level performance. *Geoscientific Model Development*, 13(6):2825–2850.

Oh, W. and Lee, K. (2004). Causal relationship between energy consumption and GDP revisited: the case of Korea 1970–1999. *Energy Economics*, 26(1):51 – 59.

Ozturk, I. (2010). A literature survey on energy–growth nexus. *Energy Policy*, 38(1):340 – 349.

Parkinson, S. and Young, P. (1998). Uncertainty and sensitivity in global carbon cycle modeling. *Climate Research*, (9):157–174.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, pages 1361–1401.

Peters, G. P., Le Quéré, C., Andrew, R. M., Canadell, J. G., Friedlingstein, P., Ilyina, T., Jackson, R. B., Joos, F., Korsbakken, J. I., McKinley, G. A., Sitch, S., and Tans, P. (2017). Towards real-time verification of CO2 emissions. *Nature Climate Change*, 7(12):848–850.

Petrova, D., Koopman, S., Ballester, J., and Rodo, X. (2017). Improving the Long-Lead Predictability of El Nino Using a Novel Forecasting Scheme Based on a Dynamic Components Model. *Climate Dynamics*, 48(3):1249–1276.

Poulter, B., Frank, D., Hodson, E., and Zimmermann, N. (2011). Impacts of land cover and climate data selection on understanding terrestrial carbon dynamics and the CO2 airborne fraction. *Biogeosciences*, 8:2027–2026.

Pretis, F., Reade, J., and Sucarrat, G. (2018). Automated general-to-specific (gets) regression modeling and indicator saturation methods for the detection of outliers and structural breaks. *Journal of Statistical Software*, 86(3).

Raupach, M., Marland, G., Ciais, P., Le Quéré, C., Canadell, J., Klepper, G., and Field, C. (2007). Global and regional drivers of accelerating CO2 emissions. *Proceedings of the National Academy of Sciences*, 104(24):10288–10293.

Raupach, M. R. (2013). The exponential eigenmodes of the carbon-climate sytem, and their implications for ratios of responses to forcings. *Earth System Dynamics*, 4:31 – 49.

Raupach, M. R., Canadell, J. G., and Quéré, C. L. (2008). Anthropogenic and biophysical contributions to increasing atmospheric CO2 growth rate and airborne fraction. *Biogeosciences*, 5:1601 – 1613.

Raupach, M. R., Gloor, M., Sarmiento, J. L., Canadell, J. G., Frölicher, T. L., Gasser, T., Houghton, R. A., Le Quéré, C., and Trudinger, C. M. (2014). The declining uptake rate of atmospheric CO_2 by land and ocean sinks. *Biogeosciences*, 11(13):3453–3475.

Rayner, P. J., Stavert, A., Scholze, M., Ahlström, A., Allison, C. E., and Law, R. M. (2015). Recent changes in the global and regional carbon cycle: analysis of first-order diagnostics. *Biogeosciences*, 12(3):835–844.

Rödenbeck, C., Bakker, D., Metzl, N., Olsen, A., Sabine, C., Cassar, N., Reum, F., Keeling, R., and Heimann, M. (2014). Interannual sea–air CO2 flux variability from an observation-driven ocean mixed-layer scheme. *Biogeosciences*, 11:4599–4613.

Ropelewski, C. and Jones, P. (1987). An extension of the Tahiti-Darwin Southern Oscillation Index. *Monthly Weather Review*, 115:2161–2165.

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C. S., Wallace, D. W. R., Tilbrook, B., Millero, F. J., Peng, T.-H., Kozyr, A., Ono, T., and Rios, A. F. (2004). The Oceanic Sink for Anthropogenic CO2. *Science*, 305(5682):367–371.

Schwinger, J., Goris, N., Tjiputra, J., Kriest, I., Bentsen, M., Bethke, I., Ilicak, M., Assmann, K., and Heinze, C. (2016). Evaluation of NorESM-OC (versions 1 and 1.2), the ocean carbon-cycle stand-alone configuration of the Norwegian Earth System Model (NorESM1). *Geoscientific Model Development*, 9:2589–2622.

Séférian, R., Berthet, S., and Chevallier, M. (2018). Assessing the decadal predictability of land and ocean carbon uptake. *Geophysical Research Letters*, 45(5):2455–2466.

Sellar, A., Jones, C., Mulcahy, J., Tang, Y., Yool, A., Wiltshire, A., O'connor, F., Stringer, M., Hill, R., Palmieri, J., et al. (2019). UKESM1: Description and evaluation of the UK Earth System Model. *Journal of Advances in Modeling Earth Systems*, 11(12):4513–4558.

Smith, B., Warlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S. (2014). Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences*, 11:2027–2054.

Spring, A. and Ilyina, T. (2020). Predictability horizons in the global carbon cycle inferred from a perfect-model framework. *Geophysical Research Letters*, 47(9):e2019GL085311. e2019GL085311 10.1029/2019GL085311.

Stern, D. (1993). Energy and economic growth in the USA: A multivariate approach. *Energy Economics*, 15(2):137 – 150.

Stern, D. (2000). A multivariate cointegration analysis of the role of energy in the US macroeconomy. *Energy Economics*, 22(2):267 – 283.

Stern, D. and Kander, A. (2012). The role of energy in the industrial revolution and modern economic growth. *The Energy Journal*, 33(3).

Tian, H., Chen, G., Lu, C., Xu, X., Hayes, D., Ren, W., Pan, S., Huntzinger, D., and Wofsy, S. (2015). North American terrestrial CO2 uptake largely offset by CH4 and N2O emissions: toward a full accounting of the greenhouse gas budget. *Climatic Change*, 129(3-4):413–426.

Vuichard, N., Messina, P., Luyssaert, S., Guenet, B., Zaehle, S., Ghattas, J., Bastrikov, V., and Peylin, P. (2019). Accounting for carbon and nitrogen interactions in the global terrestrial ecosystem model ORCHIDEE (trunk version, rev 4999): multi-scale evaluation of gross primary production. *Geoscientific Model Development*, 12(11):4751–4779.

Walker, A., Quaife, T., van Bodegom, P., De Kauwe, M., Keenan, T., Joiner, J., Lomas, M., MacBean, N., Xu, C., Yang, X., et al. (2017). The impact of alternative trait-scaling hypotheses for the maximum photosynthetic carboxylation rate (Vcmax) on global gross primary production. *New Phytologist*, 215(4):1370–1386.

Watson, A. J., Schuster, U., Shutler, J. D., Holding, T., Ashton, I. G., Landschützer, P., Woolf, D. K., and Goddijn-Murphy, L. (2020). Revised estimates of ocean-atmosphere CO2 flux are consistent with ocean carbon inventory. *Nature communications*, 11(1):1–6.

World Bank (2021). World GDP (constant 2010 US$). `https://data.worldbank.org/indicator/NY.GDP.MKTP.KD`. Accessed: 2021-02-15.

Wright, R. M., Le Quéré, C., Buitenhuis, E., Pitois, S., and Gibbons, M. J. (2021). Role of jellyfish in the plankton ecosystem revealed using a global ocean biogeochemical model. *Biogeosciences*, 18(4):1291–1320.

Yuan, W., Liu, D., Dong, W., Liu, S., Zhou, G., Yu, G., Zhao, T., Feng, J., Ma, Z., Chen, J., Chen, Y., Chen, S., Han, S., Huang, J., Li, L., Liu, H., Liu, S., Ma, M., Wang, Y., Xia, J., Xu, W., Zhang, Q., Zhao, X., and Zhao, L. (2014). Multiyear precipitation reduction strongly decreases carbon uptake over northern China. *Journal of Geophysical Research: Biogeosciences*, 119(5):881–896.

Yue, X. and Unger, N. (2015). The Yale Interactive terrestrial Biosphere model version 1.0: description, evaluation and implementation into NASA GISS ModelE2. *Geoscientific Model Development*, 8(8):2399–2417.

Zaehle, S. and Friend, A. (2010). Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates. *Global Biogeochemical Cycles*, 24(1).

Zeng, J., Nojiri, Y., Landschützer, P., Telszewski, M., and Nakaoka, S. (2014). A global surface ocean fCO2 climatology based on a feed-forward neural network. *Journal of Atmospheric and Oceanic Technology*, 31(8):1838–1849.

Zhang, X.-P. and Cheng, X.-M. (2009). Energy consumption, carbon emissions, and economic growth in China. *Ecological Economics*, 68(10):2706 – 2712.