nature climate change

Article

https://doi.org/10.1038/s41558-023-01761-x

Surprising stability of recent global carbon cycling enables improved fossil fuel emission verification

Received: 14 August 2022

Accepted: 10 July 2023

Published online: 17 August 2023

Check for updates

Benjamin Birner [1], Christian Rödenbeck [2], Julia L. Dohner [1], Armin Schwartzman [3] & Ralph F. Keeling [1]

The interannual to decadal variability in natural carbon sinks limits the explanation of recent changes in atmospheric CO_2 concentration. Here we account for interannual and decadal variability using a simple quasi-mechanistic model of the net land carbon exchange with terms scaling with atmospheric CO_2 and a weighted spatial average of temperature anomalies. This approach reduces the unexplained residual in Earth's carbon cycle budget from ±0.76 GtC per year obtained using process models to ±0.50 GtC per year, with the largest improvements on decadal timescales despite assuming constant dynamics. Our findings reveal remarkable stability of the carbon cycle and allow verification of reported global emissions to within 4.4% (95% confidence level) over the five-year stocktake cycle of the Paris Agreement—half the uncertainty reported previously.

Independent verification of reported fossil fuel emissions from atmospheric observations is critical to transparently tracking progress towards the reduction targets formalized in the Paris Agreement[1–4]. Uncertainty (1σ) of reported annual fossil fuel emissions is estimated to be around the 5% level[5], allowing for methodological uncertainties in deriving CO_2 fluxes from energy statistics and incomplete data, with larger uncertainty on decadal timescales due to potential systematic errors in reporting[6]. At the global scale, fossil fuel emissions can be verified in principle by quantifying all other terms in the global carbon budget[5,7] as follows:

$$F = \mathrm{AGR} - \mathrm{LU} + B + O + \delta \quad (1)$$

where LU is land use and land cover change, O is the ocean sink, B is the natural terrestrial carbon exchange and AGR is the atmospheric CO_2 growth rate, and where F includes small contributions from the cement manufacture and the carbonation sink as well as other industrial processes. This approach has been hampered, however, by unexplained discrepancies in the budget on interannual and decadal timescales[1–5], as reflected by δ, sometimes called the budget imbalance. In the 2021 assessment by the Global Carbon Project (GCP)[5], δ has a near-zero mean but varies strongly from year to year between −0.73 and 0.8 GtC yr^{-1} (1σ range), with average imbalances as large as 0.6 GtC yr^{-1} during some decades, corresponding to roughly 6–8% of current fossil fuel emissions.

On interannual timescales, the main contributor to δ is unexplained exchanges of atmospheric CO_2 with the land biosphere ($B_{net} = B - \mathrm{LU}$)[5]. Interannual variability in this exchange is mostly controlled by natural climate variability in the tropics and semi-arid extratropical regions[8,9]. Much of this variability is tied to the El Niño/Southern Oscillation (ENSO), but the complex nature of climate control has so far hampered our ability to fully account for these flux fluctuations. ENSO modulates terrestrial CO_2 release via temperature- and moisture-driven changes in net primary productivity and, to a lesser extent, heterotrophic respiration and biomass burning[10–14]. Soil moisture plays a critical role in net biome exchange through land–atmosphere interactions that amplify temperature extremes[9,15–17]. However, local compensation of water-driven anomalies in photosynthesis and respiration and spatial anticorrelation of precipitation patterns limit the influence of water availability relative to temperature on global carbon exchange[13,18]. Interannual variability in ocean carbon sequestration is roughly 80% smaller than on land and follows different temporal patterns[5,11,19].

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA. [2]Max Planck Institute for Biogeochemistry, Jena, Germany. [3]Division of Biostatistics and Halıcıoğlu Data Science Institute, University of California San Diego, La Jolla, CA, USA. e-mail: bbirner@ucsd.edu

https://doi.org/10.1038/s41558-023-01761-x

There is also substantial unexplained variability in CO_2 on decadal timescales[5,20]. Bastos et al.[21] reviewed the unexplained low AGR in the 1940s and suggested a connection to increased land abandonment caused by World War II. Rafelski et al.[22] evaluated the same period and identified coincident global cooling as an important contribution to a low AGR. Keeling et al.[23] showed that an anomalous decrease in the natural carbon sinks in the mid-1980s was necessary to reconcile the stagnating AGR with a concurrent deceleration in fossil emissions. Keenan et al.[24] and Ballantyne et al.[25] discussed a temporary slowdown in the AGR after 2002. Both explain the decade-long slowdown by increased carbon sequestration into the terrestrial biosphere linked to the 'warming hiatus' and ongoing CO_2 fertilization.

Two general approaches have been taken to model land carbon exchanges (LU and B), one using dynamic global vegetation models[5,10,11] and one using linear regression of B against observational climate indices, such as average tropical land temperature or ENSO state[11,13,14,19,22,26,27]. However, both methods leave a substantial fraction of variability in the observations unexplained. While global regression models have the advantage of being less complex and less computationally expensive than dynamic global vegetation models, they do not fully account for (1) spatial heterogeneity in the temperature sensitivity of carbon exchange or (2) variations with season (for example, spring versus summer warming can have opposite effects on plant growth).

Information on spatial and seasonal details in the temperature sensitivity of land carbon exchange is available from the extensive set of atmospheric CO_2 records observed across Earth's surface. On the basis of 196 measurement stations, Rödenbeck et al.[19,28,29] derived annually repeating, season-specific gridded temperature sensitivities of land carbon fluxes within an atmospheric inversion framework ('NEE-T inversion'). The estimated sensitivities are ecologically meaningful and possess predictive power as they are (1) robust against cross-validation and (2) compatible within uncertainties with analogous regression coefficients from fully independent eddy covariance CO_2 flux measurements. Rödenbeck et al.[29] used these temperature sensitivities to study interannual variability of land carbon fluxes and showed that previously observed changes in temperature sensitivity[12] were partially explained by the pattern of warming rather than ecosystem or physiological changes.

Here we make use of these spatially and seasonally resolved temperature sensitivities to improve on previous global regressions of the net land flux ($B_{net,mod}$). We develop a predictive model of global carbon fluxes that can be applied on interannual to decadal timescales to study the stability of carbon cycle dynamics and to verify fossil fuel emissions.

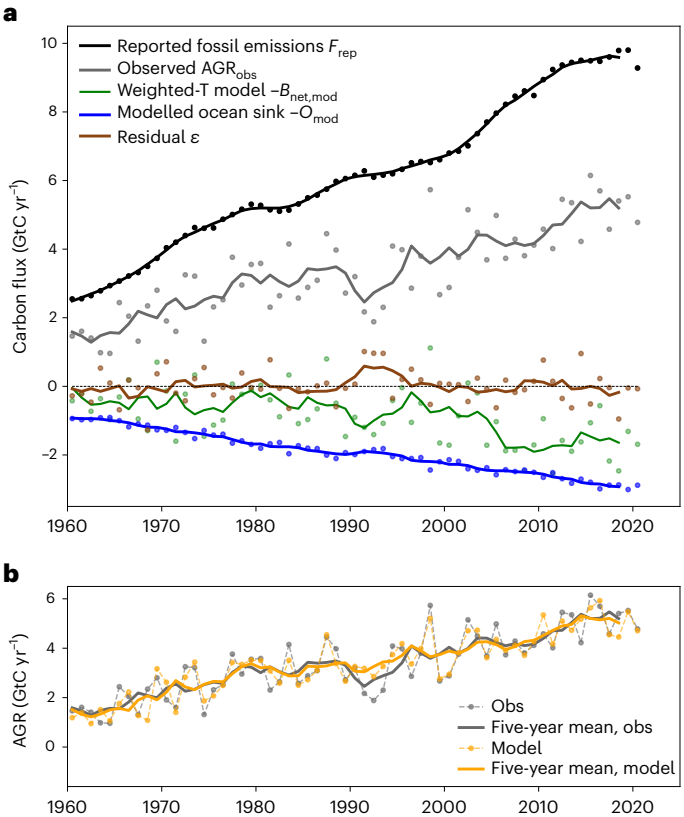


Fig. 1 | Re-evaluation of Earth's carbon budget. a,**b**, Earth's carbon budget recalculated using our 'Weighted-T' regression model of terrestrial carbon exchange and a simple ocean model (**a**) and a comparison between the observed and our modelled atmospheric growth rate of CO_2 (**b**). Reported annual fossil fuel (F_{rep}, including cement effects) emissions are from the Global Carbon Project 2021[5]. The atmospheric CO_2 growth rate (AGR_{obs}) is calculated from the average of observed CO_2 concentrations at Mauna Loa and the South Pole observatories[39]. The ocean sink (O_{mod}) is determined using a pulse-response model driven with observed CO_2 concentrations[33]. The net land carbon exchange ($B_{net,mod}$) is derived from the 'Weighted-T' model (see text). Annual average fluxes (dots) are shown at the midpoint of each year and five-year running means (solid lines) are applied to all data to guide the eye. **b**, Comparison of the observed AGR to the growth rate calculated as $AGR_{mod} = F_{rep} - O_{mod} - B_{net,mod}$.

Simple regression model of the net land sink

We estimate the annual net land carbon exchange B_{net} (Fig. 1), using the model

$$B_{net,mod} = a \times \Gamma + b \times CO_2 + c \quad (2)$$

where the term $a \times \Gamma$ represents temperature-driven carbon fluxes, with $\Gamma = \sum_{i=1}^{n} \gamma_i T_i$ being a weighted spatial average of local land temperature anomalies T_i with weights for each grid cell i of n being proportional to the local seasonally resolved temperature sensitivity γ_i taken from the CarboScope inversion run 'sEXTocNEET_v2021'[19]. γ_i values repeat annually and have been shown to possess predictive power outside this study by Rödenbeck et al.[19,29]. Temperature anomalies T_i are taken from Berkeley Earth[30] and have been decadally detrended by subtracting an exponential moving average with a 2.5-year time constant (Methods). This effectively forces climate-driven fluxes to behave as exchanges with a fast-turnover pool as in Rafelski et al.[22]. The term $b \times CO_2$ accounts for a long-term increasing sink, including but not limited to atmospheric CO_2 fertilization. Average atmospheric CO_2 concentrations are calculated as the mean of CO_2 concentrations at Mauna Loa and the South Pole, which closely track global trends in the global NOAA CO_2 record from sampling stations in the well-mixed marine boundary layer in recent years[31] but are continuous from 1958 to 2021. Following work by Dohner et al.[32], the global mean land use flux is assumed constant over the regression period and contained in the regression coefficient c, but c also adjusts for the non-zero pre-industrial values of $a \times \Gamma$ and $b \times CO_2$.

The coefficients a, b and c (Table 1) are derived from a linear regression of $B_{net,mod}$ against the residual land sink ($B_{net,res}$):

$$B_{net,res} = F_{rep} - O_{mod} - AGR_{obs} = B_{net,mod} + \epsilon \quad (3)$$

where AGR_{obs} is also calculated from the average of CO_2 concentrations at Mauna Loa and the South Pole. F_{rep} is GCP-reported fossil emissions including the cement carbonation sink[5], O_{mod} is the ocean carbon uptake calculated using the well-established global pulse-response model by Joos et al.[33] and ϵ is the residual of the optimized model. The parameters are tuned to match the overall growth in CO_2. The tuning also partly compensates for imperfections in the representation of the ocean sink O, but replacing O_{mod} with the central ocean sink estimate from GCP has little impact (Methods and Extended Data Table 2). In the following, $B_{net,mod}$ according to equation (2) (here in combination with O_{mod})

https://doi.org/10.1038/s41558-023-01761-x

Table 1 | Regression results and performance metrics for various model configurations using different inputs for predictor *Γ* and equivalent calculations for the Global Carbon Project between 1960 and 2020

Parameter	Weighted-T model	Tropical-T model	ENSO model	Global Carbon Project
Model configuration: input used for predictor *Γ*	Globally integrated, *γ*-weighted and detrended land temperature anomalies	Tropical average detrended land temperature anomalies	Niño 3.4 index	Not a regression result
a (GtC yr^{-1} K^{-1})	−4.384±0.408	−3.880±0.467	−0.810±0.241	
b (yr^{-1})	0.00839±0.00131	0.00895±0.00132	0.01020±0.00142	
c (GtC yr^{-1})	−5.293±0.974	−5.731±0.971	−6.613±1.074	
r^2	0.749±0.059	0.638±0.074	0.472±0.207	
RMSE (GtC yr^{-1})	0.50±0.09	0.60±0.10	0.74±0.21	0.76±0.11
Decadal RMSE (GtC yr^{-1})	0.16±0.04	0.20±0.03	0.24±0.07	0.36±0.14

Uncertainties (1*σ*) coefficients *a*, *b*, and *c* from the regression model in equation (2) as well as the model's coefficient of determination r^2 are obtained by jackknifing with five-year non-overlapping blocks. RMSE values represent the true generalization error and are derived using cross-validation on the same blocks. Decadal RMSE is calculated analogously after smoothing the residual model errors *δ* (or the GCP budget imbalance *δ*) with a ten-year moving average. Uncertainty of the generalization error (that is, uncertainty on RMSE) is approximated following the method of Nadeau and Bengio[38]. See Methods for details.

is called the 'Weighted-T' model. Values and uncertainties of all coefficients as well as the generalization error (that is, the root mean square error (RMSE) of the model when presented with new data) are derived using jackknifing and cross-validation, respectively, by repeatedly declaring non-overlapping five-year windows as validation data and the remaining 55 years as training data for the regression (Methods). This method prevents overfitting and ensures that the regression model has predictive power over interannual to decadal timescales.

The Weighted-T model has a cross-validated RMSE of 0.50 ± 0.09 GtC yr^{-1} (1*σ*) and can explain approximately 75 ± 6% of the variance in the residual land sink ($B_{net,res}$). RMSE values reduce to 0.16 ± 0.04 GtC yr^{-1} after filtering the optimized model residual *ϵ* to decadal timescales. A corresponding RMSE calculation of the budget imbalance *δ* from process models reported by the GCP[5] yields substantially higher values of 0.76 ± 0.11 GtC yr^{-1} and 0.36 ± 0.14 GtC yr^{-1} on interannual and decadal timescales, respectively (Extended Data Figs. 1 and 2). For a fair comparison, given that these models are not explicitly tuned, we also detrend *δ* by subtracting a best fit $\alpha \times CO_2 + \beta$. However, this fit only accounts for 6% of the variance and does not reduce the cross-validated RMSE. We hence consider *δ* and *ϵ* to be fundamentally comparable quantities. Performance gains of the Weighted-T model compared to GCP process models are greatest on decadal timescales as demonstrated by the smaller ratio of decadal to interannual RMSE (GCP/Weighted-T interannual: 0.76/0.50 = 1.52; versus decadal: 0.36/0.16 = 2.25). We also repeated the cross-validation with 10, 15, 20 and 30-year windows, and observed a substantial increase in the RMSE only for windows longer than 15 years, supporting that the model has predictive power on decadal timescales.

We also tested two simpler model configurations, replacing predictor Γ with either unweighted tropical mean land temperature (|lat|≤25, 'Tropical-T model') or the four-month lagged Niño 3.4 index ('ENSO model'). The ENSO and Tropical-T models both yield higher RMSE values on interannual and decadal timescales than the Weighted-T model (Table 1). Allowing different lag times for Γ of up to one year for all model configurations does not improve the model fits. Notably, *ϵ* of the Weighted-T model is not temporally autocorrelated, in contrast to the residuals of the Tropical-T and ENSO models or the GCP budget imbalance[3], leading to slower error accumulation and therefore higher predictive power on decadal timescales. Including global mean stratospheric aerosol optical thickness at 550 nm[34] as an additional predictor for volcanic influences in the regression did not improve the performance of the Weighted-T model and only modestly improved the performance of the ENSO model, yet the Weighted-T model still outperforms all other model configurations (Extended Data Figs. 3, 4 and Extended Data Table 1). Similarly, excluding the three years affected by the Mount Pinatubo eruption (1991–1993) from our regression analysis improves the performance of all models, but the Weighted-T model continues to have the lowest RMSE among models (Extended Data Table 3).

Surprising stability of the carbon cycle

The atmospheric growth rate of CO_2 calculated using the Weighted-T land model (Fig. 1) provides a unified understanding of a wide range of interannual and decadal features previously only considered in isolation. Accounting for spatial and seasonal heterogeneity in the sensitivity of land carbon fluxes to temperature and other correlated climate variables substantially improves our ability to reproduce annual CO_2 growth compared to alternative models. The Weighted-T model successfully captures the large ENSO-related impacts on AGR, including the strong El Niño event in 1997–1998 and the subsequent La Niña, while also accounting for the impacts of temperature on the carbon cycle on other timescales. Except for the three years immediately following the eruption of Mount Pinatubo in 1991, the modelled AGR reproduces most decadal features of the observations, including the stagnating AGR in the 1980s following two decades of steady increase, as well as slow AGR growth in the 1990s and 2000s when the land sink and fossil emissions underwent large decadal changes. The poorer performance following Pinatubo indicates that influences neglected by the model, such as impacts of diffuse light on plant growth[35], played an important role during this period.

The good performance of the Weighted-T model suggests a perhaps surprising stability of the sensitivity of the net carbon exchange to climate and CO_2 forcing at the global level since 1958. The Weighted-T model achieves good agreement with observations despite assuming a constant local land response to temperature and atmospheric CO_2 throughout 1958–2021. Although all regression coefficients in the land sink model are constants and the values of the local temperature sensitivity γ_i from the 'NEE-T inversion'[19] repeat annually, the residual *ϵ* of the Weighted-T model shows no trend and is homoscedastic (that is, the magnitude of the residual is stable over the time series). Our results therefore contrast with previous studies that highlight substantial changes in the physiological response of the biosphere to temperature[12,25,36]. Instead, the good agreement with observations obtained with the Weighted-T model supports previous findings that the apparent changes in land carbon sensitivity at the global scale are partly explained by shifts in the pattern of warming[29]. Although the $a \times \Gamma$ term of the Weighted-T model accounts only for carbon exchange with fast-responding carbon pools (approximately 2.5-year turnover) which attenuate most but not all decadal variability in temperature, this term nevertheless accounts for much of the decadal behaviour of the model. Mechanistically, the success of the

https://doi.org/10.1038/s41558-023-01761-x

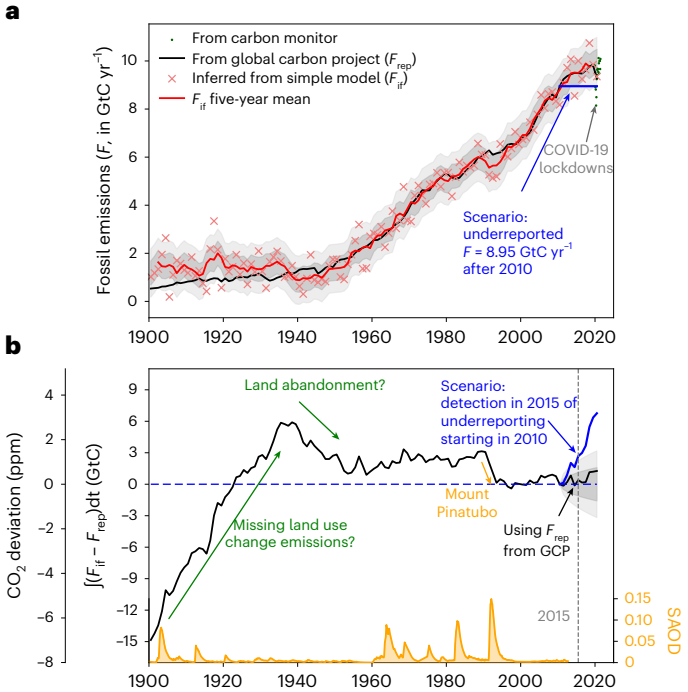


Fig. 2 | Fossil fuel emissions and time-integrated difference between inferred and reported fossil fuel emissions. a, Comparison between reported emissions (black line) and emissions inferred based on observed atmospheric CO_2 (red crosses and red line). Uncertainty in annual inferred emissions is given by the RMSE and shown as 1σ and 2σ envelopes around the five-year running mean (dark and light grey, respectively). Annual reported emissions are obtained from the Global Carbon Project[5]. Beginning in 2020 we also show rescaled monthly emission estimates (green dots; Methods) from Carbon Monitor[37] to highlight impacts of the COVID-19 pandemic. The solid blue line illustrates the counterfactual scenario of falsely reported constant emissions after year 2010 (see text). **b**, Integrated difference between annual inferred emissions (F_{if}), GCP-reported emissions (F_{rep}, black line) and emissions reported in our counterfactual scenario (blue line). The *y*-axis can be directly mapped to a deviation between the observed atmospheric CO_2 concentrations (in ppm) and concentrations expected from the reported emissions (second *y*-axis on the left, scaling factor is 2.124 GtC ppm^{-1}, see text). The range of integrated emission differences that are plausible after 2010 if emissions were reported correctly are shown as grey bands (1σ and 2σ) around a dotted blue line. Note that the integration constant is arbitrarily chosen to anchor the time series to zero in 2010. Stratospheric aerosol optical depth at 550 nm (SAOD, orange line and area) is shown as a proxy for volcanic activity[34].

model indicates that short-lived pools dominate variability in land carbon fluxes over a broad range of timescales as suggested previously[22]. The term $b \times CO_2$ accounts for the multi-decadal response of the Weighted-T model to increasing CO_2, changing climate and other correlated factors, but has less of a mechanistic underpinning and is therefore not suitable for long-term forecasting.

Verifying global fossil fuel emissions

The ability of the Weighted-T regression model to tightly reproduce the interannual variations in the natural part of the global carbon budget also provides a means to verify reported emissions at improved accuracy with observations of atmospheric CO_2 concentrations[1,3,4]. The Paris Agreement has a five-year stocktake cycle in which ongoing emission reduction efforts are reviewed. Reported global emissions (F_{rep}) can be compared to inferred emissions (F_{if}) calculated by combining equations (1)–(3):

$$F_{if} = AGR_{obs} + O_{mod} + (a \times \Gamma + b \times CO_2 + c) \quad (4)$$

Consider the following counterfactual scenario (Fig. 2): global emissions were reported to be constant at a value of F_{rep} = 8.95 GtC yr^{-1} after 2010, while in reality emissions continued to increase. In 2015 an interested party could use our regression model to hindcast emissions since 2010 (F_{if}) using available observations of temperature and CO_2. The difference between the inferred F_{if} and reported F_{rep} is detectable in the cumulative CO_2 flux as an unexplained buildup of CO_2 in the atmosphere that is already statistically significant by 2015 and continues to grow thereafter (Fig. 2b). In contrast, it would take almost ten years to detect a significant disagreement if, instead of the cumulative emissions (that is, the impact on atmospheric CO_2 concentrations), we used the difference in F_{if} and F_{rep} directly (that is, the impact on AGR, Fig. 2a). This is because cumulative uncertainty increases as the square root of the time difference (that is, $\sigma_{cum} = \sqrt{\Delta t} \times \mathrm{RMSE}$) while the emission signal ($S = \Delta t \times F$) grows linearly, as it is simply integrated over time[3]. Our scenario illustrates that, if reported emissions stayed constant at current levels of around 10 GtC yr^{-1}, any misreporting of greater than $\mathrm{CI}_{95\%} \times \frac{\sqrt{\Delta t}\mathrm{RMSE}}{\Delta t \times S} = 1.96\frac{\sqrt{5}\times 0.5}{5\times 10} = 4.4\%$ could be detected at the 95% confidence level ($\mathrm{CI}_{95\%} = 1.96 \times \sigma_{cum}$) after Δt = 5 years and would manifest as a disagreement with observed CO_2 concentrations of >1 ppm. Following previous approaches that use the GCP budget terms for emission verification which have a larger and temporally autocorrelated uncertainty given by its budget imbalance[3,4], we could instead only confidently detect a two times larger misreporting of emissions (8.8%) over the same window (Methods). As we apply the verification only once (at the five-year window), there is no issue with repeat testing[3]. Nevertheless, even with the improved performance of the Weighted-T model, we are still unable to detect short-lived emission changes such as the 0.52 GtC reduction in 2020 global emissions compared to 2019 caused by the COVID-19 pandemic[5,37].

Outside the period affected by the Mount Pinatubo volcanic eruption, F_{rep} and F_{if} agree remarkably well since 1958, supporting the accuracy of reported emissions F_{rep}. Figure 2 confirms the reported plateauing of emissions in the last decade and our regression model finds no evidence of inaccurate reporting of emissions since the Paris Agreement was adopted in December 2015. The model also does not detect any significant underreporting of emissions in the 1990s and 2000s, as suggested previously using atmospheric data employing a different approach[1,2].

Our carbon cycle model can also be applied to study emissions before the start of direct atmospheric CO_2 measurements using ice core reconstructions and gridded temperature fields available since 1900 (Fig. 2 and Extended Data Fig. 5). From 1900 until around 1935, annual inferred emissions are 0.59 GtC yr^{-1} higher than reported emissions, followed by the 1940s where inferred emissions are 0.3 GtC yr^{-1} lower. Although we cannot exclude contributions from unrepresented multi-decadal variability in the ocean sink and inaccurate temperature and/or CO_2 input data, we suggest that these disagreements are most likely explained by neglected variations in land use emissions, which are assumed constant in our regression model but almost certainly varied to some degree over the twentieth century[5]. LU has a large uncertainty, so 0.59 GtC yr^{-1} higher LU emissions at the beginning of the twentieth century compared to the average after 1960 are plausible. Bastos et al.[21] suggested that LU decreased anomalously in the 1940s because of the socioeconomic turmoil of World War II and explored several hypothetical LU scenarios to account for increased land abandonment. These scenarios yielded changes in AGR of up to 0.39 GtC yr^{-1}, more than sufficient to explain the underestimation of emissions in our hindcast during the 1940s. Previous carbon cycle models, in contrast, produce discrepancies too large to be completely explained by the land abandonment scenarios in the analysis of Bastos et al.[21]. The improved performance of the Weighted-T model in the 1940s compared to models reviewed by Bastos et al.[21] is probably explained by its ability to at least partially capture an increase in land carbon uptake associated with declining surface temperatures between 1940 and 1960, similar to the box model by Rafelski et al.[22].

In summary, by combining a simple regression model of the net land sink and a pulse-response ocean model[33], we built a framework that explains many of the important interannual and decadal features of the atmospheric CO_2 rise since 1960 and also provides a new context for understanding variability from 1900 from 1960. The net land sink model accounts for the effect of CO_2 fertilization (and/or other correlated long-term trends) and climate-driven CO_2 flux variations which are calculated based on the spatially and seasonally resolved temperature sensitivity of biosphere carbon exchange determined previously from the CarboScope global atmospheric CO_2 inversion[19]. The model framework can be used to verify reported global emissions to within 4.4% at the 95% confidence level over the five-year stocktake window of the Paris Agreement. Model parameters are constant yet the model's performance over the last six decades is consistently strong, suggesting that there were no large surprises in the carbon cycle. Discrepancies in the early twentieth century point to larger temporal variability in the land use flux before 1960.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41558-023-01761-x.

References

1. Francey, R. J. et al. Atmospheric verification of anthropogenic CO_2 emission trends. *Nat. Clim. Change* **3**, 520–524 (2013).
2. Francey, R. J. et al. Differences between trends in atmospheric CO_2 and the reported trends in anthropogenic CO_2 emissions. *Tellus B* **62**, 316–328 (2010).
3. Schwartzman, A. & Keeling, R. F. Achieving atmospheric verification of CO_2 emissions. *Nat. Clim. Change* **10**, 416–417 (2020).
4. Peters, G. P. et al. Towards real-time verification of CO_2 emissions. *Nat. Clim. Change* **7**, 848–850 (2017).
5. Friedlingstein, P. et al. Global Carbon Budget 2021. *Earth Syst. Sci. Data* **14**, 1917–2005 (2022).
6. Ballantyne, A. P. et al. Audit of the global carbon budget: estimate errors and their impact on uptake uncertainty. *Biogeosciences* **12**, 2565–2584 (2015).
7. Ballantyne, A. P., Alden, C. B., Miller, J. B., Tans, P. P. & White, J. W. C. Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature* **488**, 70–72 (2012).
8. Ahlström, A. et al. The dominant role of semi-arid ecosystems in the trend and variability of the land CO_2 sink. *Science* **348**, 895–899 (2015).
9. Wang, K. et al. Regional and seasonal partitioning of water and temperature controls on global land carbon uptake variability. *Nat. Commun.* **13**, 3469 (2022).
10. Kim, J. S., Kug, J. S., Yoon, J. H. & Jeong, S. J. Increased atmospheric CO_2 growth rate during El Niño driven by reduced terrestrial productivity in the CMIP5 ESMs. *J. Clim.* **29**, 8783–8805 (2016).
11. Zeng, N., Mariotti, A. & Wetzel, P. Terrestrial mechanisms of interannual CO_2 variability. *Glob. Biogeochem. Cycles* **19**, GB1016 (2005).
12. Wang, X. et al. A two-fold increase of carbon cycle sensitivity to tropical temperature variations. *Nature* **506**, 212–215 (2014).
13. Wang, W. et al. Variations in atmospheric CO_2 growth rates coupled with tropical temperature. *Proc. Natl Acad. Sci. USA* **110**, 13061–13066 (2013).
14. Cox, P. M. et al. Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature* **494**, 341–345 (2013).
15. Humphrey, V. et al. Sensitivity of atmospheric CO_2 growth rate to observed changes in terrestrial water storage. *Nature* **560**, 628–631 (2018).
16. Humphrey, V. et al. Soil moisture–atmosphere feedback dominates land carbon uptake variability. *Nature* **592**, 65–69 (2021).
17. Green, J. K. et al. Large influence of soil moisture on long-term terrestrial carbon uptake. *Nature* **565**, 476–479 (2019).
18. Jung, M. et al. Compensatory water effects link yearly global land CO_2 sink changes to temperature. *Nature* **541**, 516–520 (2017).
19. Rödenbeck, C., Zaehle, S., Keeling, R. & Heimann, M. How does the terrestrial carbon exchange respond to inter-annual climatic variations? A quantification based on atmospheric CO_2 data. *Biogeosciences* **15**, 2481–2498 (2018).
20. Li, W. et al. Reducing uncertainties in decadal variability of the global carbon budget with multiple datasets. *Proc. Natl Acad. Sci. USA* **113**, 13104–13108 (2016).
21. Bastos, A. et al. Re-evaluating the 1940s CO_2 plateau. *Biogeosciences* **13**, 4877–4897 (2016).
22. Rafelski, L. E., Piper, S. C. & Keeling, R. F. Climate effects on atmospheric carbon dioxide over the last century. *Tellus B* **61**, 718–731 (2009).
23. Keeling, C. D., Whorf, T. P., Wahlen, M. & van der Pflicht, J. Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980. *Nature* **375**, 666–670 (1995).
24. Keenan, T. F. et al. Recent pause in the growth rate of atmospheric CO_2 due to enhanced terrestrial carbon uptake. *Nat. Commun.* **7**, 13428 (2016).
25. Ballantyne, A. et al. Accelerating net terrestrial carbon uptake during the warming hiatus due to reduced respiration. *Nat. Clim. Change* **7**, 148–152 (2017).
26. Betts, R. A., Jones, C. D., Knight, J. R., Keeling, R. F. & Kennedy, J. J. El Niño and a record CO_2 rise. *Nat. Clim. Change* **6**, 806–810 (2016).
27. Jones, C. D. & Cox, P. M. On the significance of atmospheric CO_2 growth rate anomalies in 2002–2003. *Geophys. Res. Lett.* **32**, L14816 (2005).
28. Rödenbeck, C., Houweling, S., Gloor, M. & Heimann, M. CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmos. Chem. Phys.* **3**, 1919–1964 (2003).
29. Rödenbeck, C., Zaehle, S., Keeling, R. & Heimann, M. History of El Niño impacts on the global carbon cycle 1957–2017: a quantification from atmospheric CO_2 data. *Phil. Trans. R. Soc. B* **373**, 20170303 (2018).
30. Rohde, R. A. & Hausfather, Z. The Berkeley Earth Land / Ocean temperature record. *Earth Syst. Sci. Data* **12**, 3469–3479 (2020).
31. Lan, X., Tans, P., Thoning, K. & NOAA Global Monitoring Laboratory. NOAA Greenhouse Gas Marine Boundary Layer Reference – CO_2. NOAA GML. https://doi.org/10.15138/DVNP-F961 (2023).
32. Dohner, J. L., Birner, B., Schwartzman, A., Pongratz, J. & Keeling, R. F. Using the atmospheric CO_2 growth rate to constrain the CO_2 flux from land use and land cover change since 1900. *Glob. Change Biol.* **28**, 7327–7339 (2022).
33. Joos, F. et al. An efficient and accurate representation of complex oceanic and biospheric models of anthropogenic carbon uptake. *Tellus B* **48**, 397–417 (1996).
34. Sato, M., Hansen, J. E., Mccormick, P. M. & Pollack, J. B. Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.* **98**, 22987–22994 (1993).
35. Farquhar, G. D. & Roderick, M. L. Pinatubo, diffuse light, and the carbon cycle. *Science.* **299**, 1997–1998 (2003).
36. Zhu, Z. et al. The accelerating land carbon sink of the 2000s may not be driven predominantly by the warming hiatus. *Geophys. Res. Lett.* **45**, 1402–1409 (2018).

37. Liu, Z., Ciais, P., Deng, Z. et al. Near-real-time monitoring of global CO_2 emissions reveals the effects of the COVID-19 pandemic. *Nat. Commun.* **11**, 5172 (2020).
38. Nadeau, C. & Bengio, Y. Inference for the generalization error. *Mach. Learn.* **52**, 239–281 (2003).
39. Keeling, C. D. et al. Exchanges of atmospheric CO_2 and $^{13}CO_2$ with the terrestrial biosphere and oceans from 1978 to 2000. *Global Aspects, SIO Reference Series* **1–6**, 1–88 (2001).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023

Methods

Regression model input

In the following we describe the data needed to set up the regression model of the net land sink. The local temperature sensitivity of land carbon fluxes and the ocean carbon sink model need to be evaluated at least at monthly resolution for numerical accuracy and to account for seasonality in plant growth. Therefore, the ocean and land inputs to equation (2) are first calculated monthly and then averaged to annual resolution for the regression analysis.

Rödenbeck et al.[19] derived the local temperature sensitivity (γ^*) of interannual carbon fluxes in a global CO_2 inversion. Extending the classical carbon flux inversion in which time-dependent land carbon exchange is optimized, the 'NEE-T inversion' imposes a linear relationship between detrended temperature anomalies and interannual carbon fluxes in each land grid cell. The mean seasonal cycle of the carbon fluxes, the trend in seasonal cycle and the long-term flux trend are accounted for separately in the inversion. γ^* is seasonally resolved (with a four-week decorrelation scale) but repeats every year, yielding 52/4 = 13 independent degrees of freedom in time. The resulting γ^* field from CarboScope inversion 'sEXTocNEET_v2021' is provided on a spatial grid of 2.5° × 2° and at daily resolution, which we average monthly. To obtain the final weighting factors $\gamma(i,\mathrm{mon})$ used in our regression modelling, we normalize the monthly averaged sensitivity $\hat{\gamma}$ in each grid cell i out of n total cells according to:

$$\gamma(i,\mathrm{mon}) = \frac{1}{12}\frac{\hat{\gamma}(i,\mathrm{mon})}{\sum_{k=1}^{12}\sum_{l=1}^{n}\hat{\gamma}(l,k)} \tag{5}$$

Monthly temperature fields were downloaded from Berkeley Earth[30] on a 1° × 1° latitude–longitude grid. The data are re-gridded to match the resolution of the Jena CarboScope inversion results and subsequently decadally detrended in each grid cell by subtracting an exponential moving average of past temperatures with a 2.5-year time constant. This avoids causality problems associated with other filters, such as a centred running mean, which effectively includes information about the future, and allows the model to be readily updated because it produces valid output on the leading edge of available data. Subtracting the exponential moving average is essentially the same as assuming that temperature modulates the carbon flux into or out of a carbon pool with a 2.5-year turnover time, paralleling Rafelski et al.[22].

For the main configuration of our net land sink model, the 'Weighted-T' model, Γ is calculated as $\Gamma = \sum_{i=1}^{n}\gamma_i T_i$, the sum of all n gridded temperature sensitivity weights γ_i and local temperature anomalies T_i over the globe. For alternative configurations of the model, we also calculate global mean land temperature and the Niño 3.4 index from the re-gridded temperature fields and use either for Γ.

De-seasonalized CO_2 observations from the South Pole (SPO) and Mauna Loa (MLO, Hawaii) are available at monthly resolution from the Scripps CO_2 Program after 1958[39]. Their monthly average has been merged with a spline fit to the Law Dome ice core CO_2 record available from the Scripps CO_2 website to extend the time series to 1800[39,40].

The atmospheric growth rate of CO_2 (AGR) is calculated as the 12-month-centred difference. For example, the AGR for May 2018 is calculated as the difference in CO_2 concentrations between November 2018 and November 2017.

Monthly ocean carbon uptake is calculated using a pulse-response ocean model published by Joos et al.[33], which only needs a history of atmospheric CO_2 concentrations as input. The model fully accounts for the nonlinearity of carbon dissolution chemistry through a high-order parametrization and represents carbon exchange with the deep ocean via circulation through a simple pulse-response function. The model has been shown to effectively emulate the behaviour of three-dimensional ocean models for the twenty-first century at a fraction of the computational cost. The model is spun up from an equilibrium state in 1800 and with parameter values for the HILDA (HS + LS) model configuration given in Table 2 and section A2.2 of Joos et al.[33]. Model output is rescaled to match the integrated ocean sink between 1800 and 2021 of 176.03 GtC reported in the Global Carbon Project (GCP)[5].

We use annual emissions from fossil fuel combustion and industrial processes reported by the GCP[5]. To reduce the number of terms in the budget, the influence of the cement carbonation sink is grouped into the F_{rep} estimate.

Carbon Monitor[37] began reporting monthly fossil emissions starting in 2019, which we use to highlight effects of the COVID-19 pandemic in Fig. 2. To remove effects of the seasonal cycle from the data and build a coherent time series with emissions reported by the GCP[5], we first divided Carbon Monitor emissions after 2020 by their monthly value in 2019 and then multiplied that anomaly by the GCP annual emissions for 2019.

Regression model set-up and uncertainties

First, a regression model is fitted to all input data between 1960 and 2020 to get a best estimate of the coefficients a, b and c and the r-squared value of the model. Then, input data are split into $M = 12$ non-overlapping blocks of five years to determine uncertainty on model parameters and find the generalization error which measures how well the model performs on data it was not trained on. The grouping into blocks addresses potential concerns about autocorrelation in the data with a timescale of <5 years. $M - 1$ blocks, that is, the 'training data', are used to fit the regression model and obtain parameter estimates, and the remaining block, that is the 'validation data', is used to evaluate model performance according to the mean squared error (MSE). The splitting procedure is repeated using each block once as validation data and can be thought of as a combination of the statistical methods 'jackknifing' and 'cross-validation'.

Jackknifing produces a suite of $M = 12$ subjugate models that were fitted to different choices of training data and used to determine uncertainties for different parameters of the regression model, including coefficients a, b and c and the R^2 value. To account for the overlap in training data for different subjugate models, we report uncertainties δ_x for parameter x based on equation (6):

$$\delta_x = \sqrt{\frac{M-1}{M}\sum_{i=1}^{M}(x_i - \bar{x})^2} \tag{6}$$

where $\bar{x}$ is the mean of parameter x over all subjugate models i.

Cross-validation produces M estimates of the generalization error, that is MSE_j, derived from different validation periods $j = 1, \ldots, M$ (Extended Data Fig. 1). We find the average $\overline{\mathrm{MSE}}$ and RMSE ($\mathrm{RMSE} = \sqrt{\overline{\mathrm{MSE}}}$) of all subjugate models as the best estimate of the generalization error. For decadal RMSE values, the regression residual ϵ is smoothed with a ten-year moving average, including at the edges where this introduces a slight bias, before the MSE_j is calculated in the validation period.

The $\overline{\mathrm{MSE}}$ and RMSE of the GCP[5] budget imbalance are calculated similarly. GCP data are split into matching five-year blocks and the MSE is determined for each independent five-year block.

To allow quantitative comparison of different models and a statistical comparison to the GCP, some confidence bounds must be placed on the generalization error. Cross-validation produces a distribution of MSE values (MSE_j) obtained for different choices of the five-year validation blocks (j). Although the validation blocks are independent, the training data for each subjugate model overlap strongly. Therefore, the variance of MSE_j values captures only some of the uncertainty of the generalization error. To estimate the true generalization error we

use a heuristic method which relies on rescaling the variance of MSE_j values obtained from cross-validation[38]:

$$\mathrm{var}\left(\overline{\mathrm{MSE}}\right) \approx \left(\frac{1}{M} + \frac{n_2}{n_1}\right)\mathrm{var}\left(\mathrm{MSE}_j\right) \quad (7)$$

where n_2 is the length of the validation period and n_1 is the length of the training period for the model. Using this formula, we obtain values of 0.25 ± 0.08, 0.36 ± 0.12, 0.55 ± 0.31 and 0.58 ± 0.17 (1σ) for the MSE of the Weighted-T model, Tropical-T model, ENSO model and the GCP, respectively. We use the delta method (which involves a mapping according to a first-order Taylor expansion of the derivative) to estimate the corresponding RMSE values.

We perform two sensitivity tests to further investigate the role of volcanic activity and the ocean model in our regression analysis. We repeat the regression analysis including stratospheric aerosol optical depth[34] as an additional predictor to better capture the influence of volcanic eruptions on land carbon exchange. Results and regression parameters are presented in Extended Data Table 1. Although the Weighted-T model itself does not benefit from the additional predictor and we observe slightly improved performance of the ENSO and Tropical-T model, the Weighted-T model continues to outperform both. When the period of the Mount Pinatubo eruption (1991–1993) is excluded from the regression analysis instead, the performance of all models increases (Extended Data Table 3) but the results remain qualitatively the same. We also repeat the regression analysis using the mean of GCP ocean sink estimates instead of the Joos et al. pulse-response model. This leads to a minor improvement of the RMSE of all model configurations but does not alter our conclusions (Extended Data Table 2).

To support our claim that the GCP budget imbalance and the residual of our regression model are comparable quantities, we correct the GCP budget imbalance with a fit of the imbalance to the atmospheric CO_2 concentration. After cross-validation, this yields an RMSE of 0.8 ± 0.11 which is higher than the value calculated without the additional fitting correction. This shows that the atmospheric CO_2 record has no predictive power for the GCP budget imbalance and that our results are not an artefact of additional tuning.

Finally, we repeat the regression analysis after changing the CO_2 term to a logarithmic relationship, that is, $b \times \ln(CO_2)$, which does not substantially change the performance of the models.

GCP emission verification

Following refs. 3,4, we benchmark our ability to verify reported fossil fuel and land use change emissions from the GCP global carbon budget[5] using its residual imbalance as a metric of uncertainty. The budget imbalance is temporally autocorrelated and best represented by an AR(1) process with an AR coefficient of 0.35 as determined by the ARMIA function of the Python statsmodel package. Monte Carlo simulations (n = 200,000) of the GCP budget imbalance represented as this AR(1)-processes find a cumulative uncertainty of 4.4 GtC (95% CI) after five years, or 8.8% assuming ongoing emissions of 10 GtC yr^{-1}.

Data availability

Jena CarboScope inversion results including gridded temperature sensitivities γ_i are available from: https://www.bgc-jena.mpg.de/CarboScope/?ID=sEXTocNEET_v2021. Source data are provided with this paper. All other data used in this study are publicly available as cited. We also archived copies of the datasets on Zenodo together with the Python code used for the analysis[41].

Code availability

Python code for the derivation and visualization of the regression model is available on Zenodo: https://doi.org/10.5281/zenodo.8019283 (ref. 41).

References

40. MacFarling Meure, C. et al. Law Dome CO_2, CH_4 and N_2O ice core records extended to 2000 years BP. *Geophys. Res. Lett.* **33**, 1–4 (2006).
41. Birner, B. Data and software: surprising stability of recent global carbon cycling enables improved fossil fuel emission verification. *Zenodo* https://doi.org/10.5281/zenodo.8019283 (2023).

Acknowledgements

We thank Y. Jin for insightful discussions of the global carbon cycle. This work was supported by the National Science Foundation under grant nos. AGS-1940361 and OPP-1922922 (both R.F.K.), E. and W. Schmidt via recommendation of the Schmidt Futures programme and Earth Networks, an AEM brand.

Author contributions

B.B., C.R. and R.F.K. conceived the research project. B.B. and J.L.D. were responsible for software. Formal analysis and development of the methodology was by B.B. and A.S. B.B. and R.F.K. wrote the original draft, which was reviewed and edited by all the authors. Supervision was by R.F.K.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41558-023-01761-x.

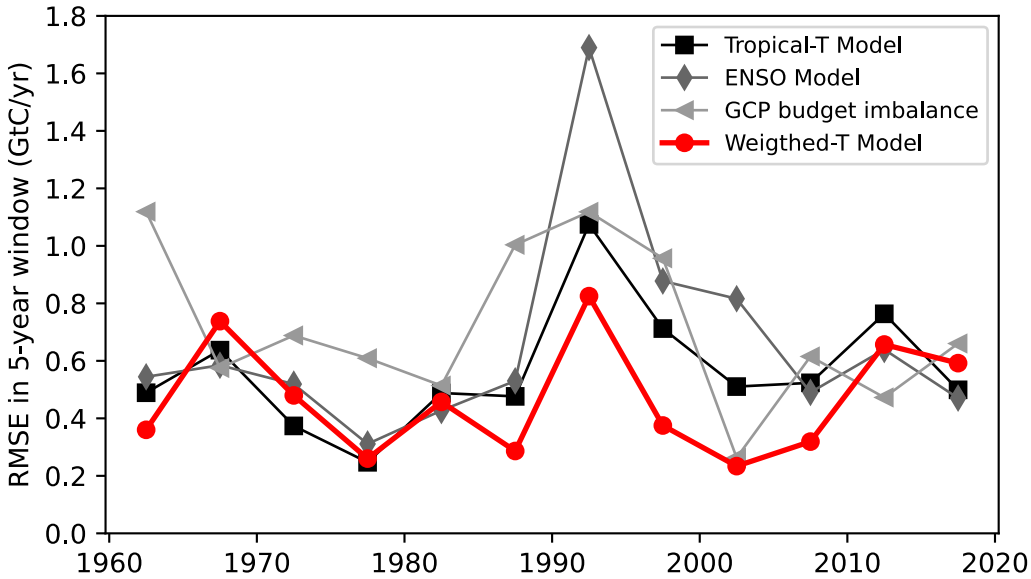
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41558-023-01761-x.

Correspondence and requests for materials should be addressed to Benjamin Birner.

Peer review information *Nature Climate Change* thanks Glen Peters and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

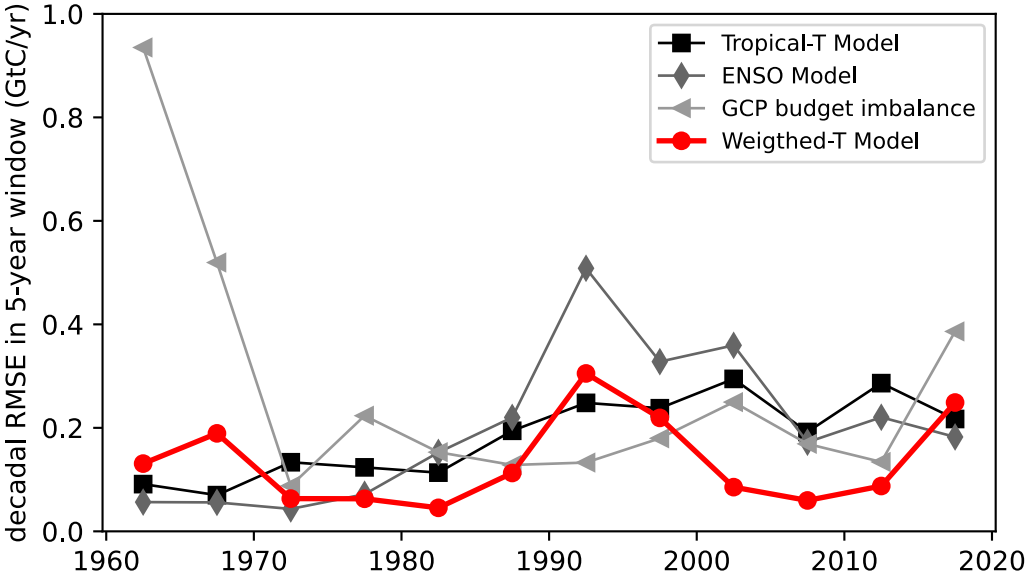
Reprints and permissions information is available at www.nature.com/reprints.

https://doi.org/10.1038/s41558-023-01761-x



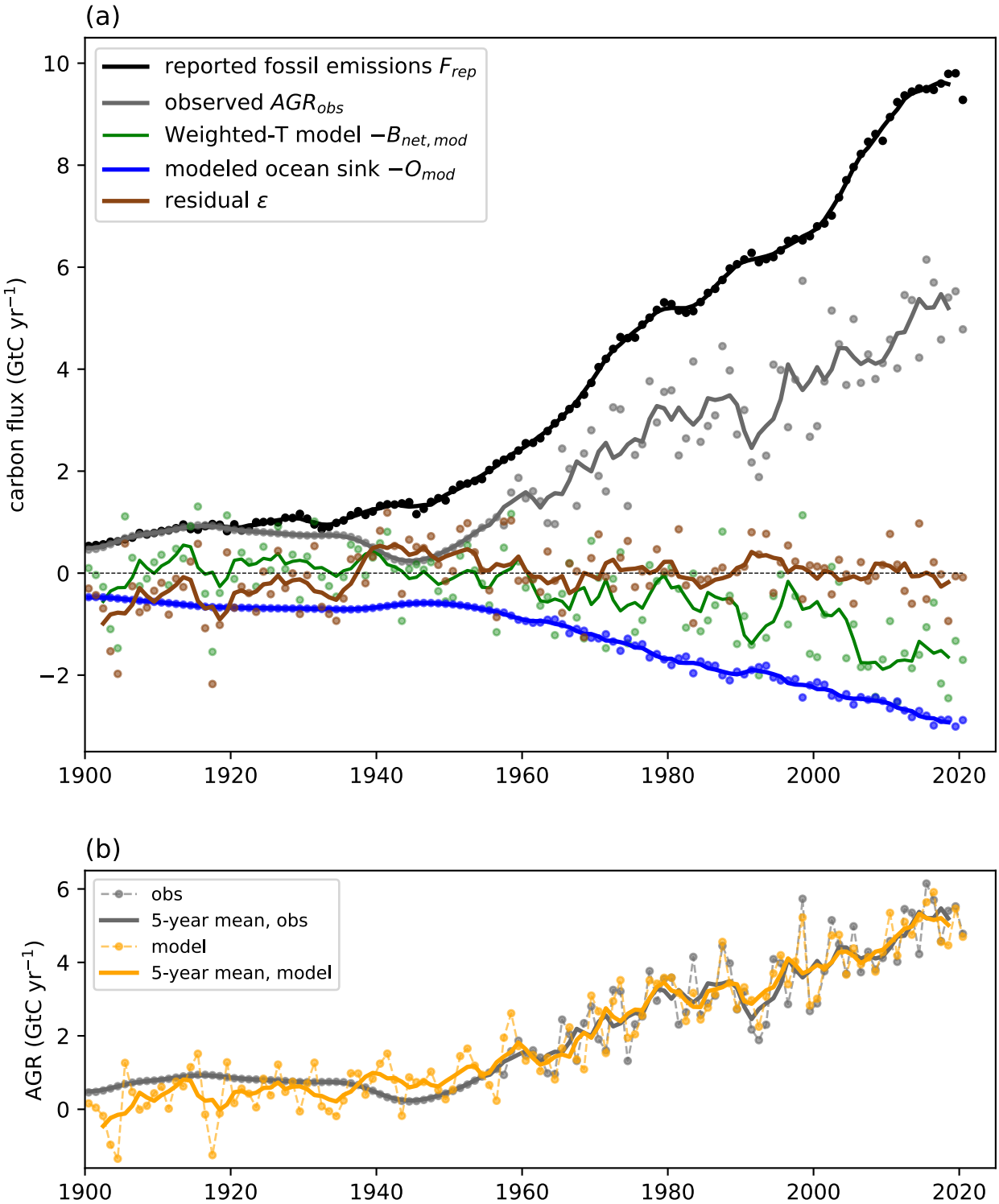
Extended Data Fig. 1 | Evolution of the RMSE of different regression model configurations and the budget imbalance of the Global Carbon Project 2021. RMSE values are calculated for a 5-year window of validation data and are plotted at the center of the window. The remaining 55 years are used as training data for the regression and are unused in the case of the GCP budget imbalance (see Methods).

https://doi.org/10.1038/s41558-023-01761-x



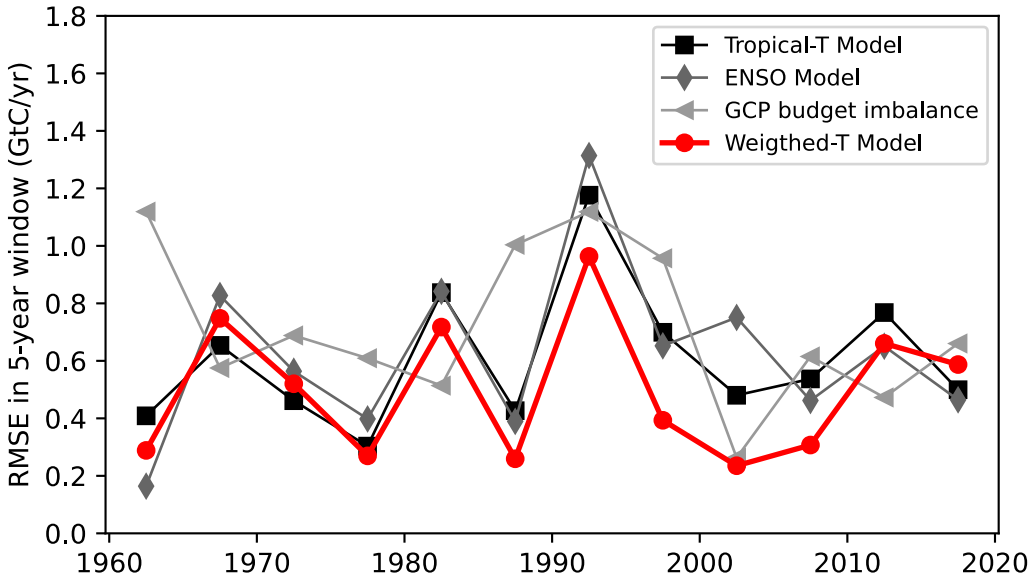
Extended Data Fig. 2 | Evolution of the RMSE in different model configurations on decadal time scales. This figure is similar to Extended Data Fig. 1 but with a 10-year running mean applied to the time series of residual model error ϵ before calculating the RMSE in each validation period.

https://doi.org/10.1038/s41558-023-01761-x



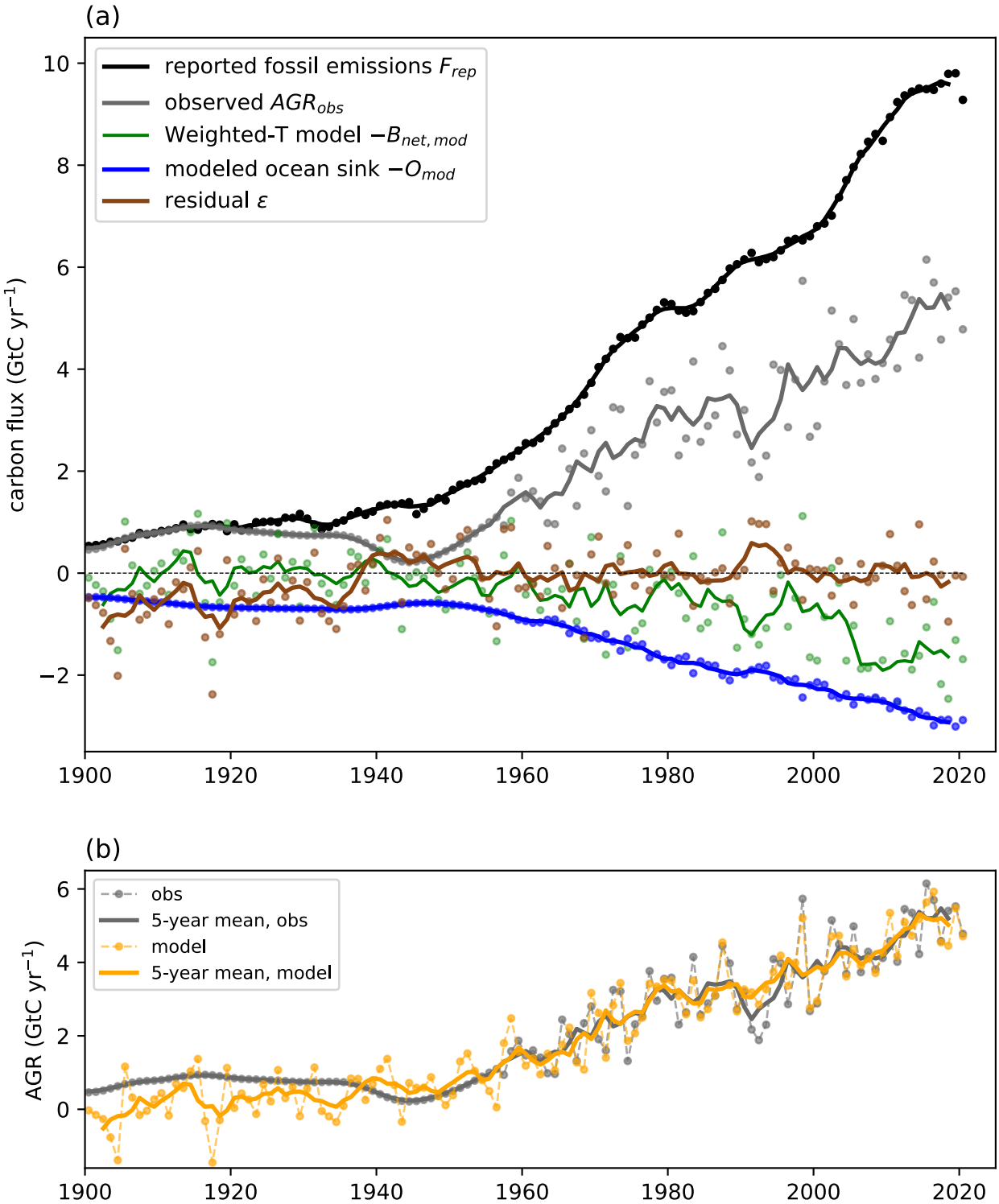
Extended Data Fig. 3 | Re-evaluated global carbon cycle since 1900 from a special model configuration including volcanic activity. Panels (a) and (b) of this figure are identical to Fig. 1a,b in the manuscript but extended to 1900 and with the land sink model including stratospheric optical thickness at 550 nm[34] as an additional predictor. Atmospheric CO_2 growth rate is calculated from a spline fit to Law Dome ice core data until 1958[40] and from the average of directly observed CO_2 concentrations at Mauna Loa and South Pole after[38].

https://doi.org/10.1038/s41558-023-01761-x



Extended Data Fig. 4 | Evolution of the RMSE in special model configurations including volcanic activity. This figure is similar to Extended Data Fig. 1 but for results including stratospheric optical thickness at 550 nm[34] as an additional predictor in net the land sink regressions.

https://doi.org/10.1038/s41558-023-01761-x



Extended Data Fig. 5 | Re-evaluation of the global carbon cycle since 1900. Panels (a) and (b) of this figure are identical to Fig. 1a,b in the manuscript but starting in 1900. Atmospheric CO_2 growth rate is calculated from a spline fit to Law Dome ice core data until 1958[40] and from the average of directly observed CO_2 concentrations at Mauna Loa and South Pole after[38].

https://doi.org/10.1038/s41558-023-01761-x

Extended Data Table 1 | Regression results and performance metrics for various model configurations including volcanic activity

parameter	Weighted-T model	Tropical-T model	ENSO model
a (GtCyr^{-1}K^{-1})	-4.302±0.395	-3.799±0.487	-1.046±0.140
b (yr^{-1})	0.00921±0.00191	0.00987±0.00195	0.01389±0.00198
c (GtCyr^{-1})	-6.009±1.494	-6.537±1.529	-9.634±1.583
d (GtCyr^{-1}SAOD^{-1})	5.770±9.139	6.576±10.008	19.707±12.100
r^2	0.765±0.056	0.658±0.058	0.637±0.077
RMSE (GtCyr^{-1})	0.55±0.10	0.65±0.11	0.68±0.14
decadal RMSE (GtCyr^{-1})	0.17±0.04	0.21±0.02	0.21±0.04

Same as Table 1 but the regression model configurations include stratospheric optical thickness at 550 nm (SAOD)[34] as an additional predictor, that is, $F_{rep} - O_{\mathrm{mod}} - AGR_{obs} = a \times \Gamma + b \times CO_2 + c + d \times SAOD + \epsilon$.

https://doi.org/10.1038/s41558-023-01761-x

Extended Data Table 2 | Regression results and performance metrics for various model configurations using a different ocean sink estimate

parameter	Weighted-T model	Tropical-T model	ENSO model
a ($GtCyr^{-1}K^{-1}$)	-4.112±0.366	-3.670±0.465	-0.839±0.207
b (yr^{-1})	0.00883±0.00179	0.00936±0.00178	0.0108±0.00164
c ($GtCyr^{-1}$)	-5.553±1.334	-5.971±1.313	-6.975±1.227
r^2	0.757±0.061	0.658±0.058	0.551±0.168
RMSE ($GtCyr^{-1}$)	0.49±0.09	0.58±0.07	0.67±0.16
decadal RMSE ($GtCyr^{-1}$)	0.19±0.05	0.21±0.04	0.22±0.05

Same as Table 1 but the GCP ocean sink estimate is used instead of the Joos et al. ocean model in the tuning and validation of the regression models.

https://doi.org/10.1038/s41558-023-01761-x

Extended Data Table 3 | Regression results and performance metrics for various model configurations with special treatment of the Pinatubo eruption years

parameter	Weighted-T model	Tropical-T model	ENSO model	Global Carbon Project
a ($GtCyr^{-1}K^{-1}$)	-4.384±0.368	-3.880±0.469	-0.810±0.109	
b (yr^{-1})	0.00839±0.00120	0.00895±0.00133	0.01020±0.00108	
c ($GtCyr^{-1}$)	-5.293±0.893	-5.731±0.977	-6.613±0.793	
r^2	0.749±0.048	0.638±0.034	0.472±0.060	
RMSE ($GtCyr^{-1}$)	0.44±0.07	0.56±0.07	0.54±0.07	0.71±0.10

Same as Table 1 but years 1991–1993 are excluded in the tuning and validation of the models.