When is a trend meaningful? Insights to carbon cycle variability from an initial-condition large ensemble

Gordon B. Bonan[1*], Clara Deser[1], William R. Wieder[1,2], Danica L. Lombardozzi[3,1], and Flavio Lehner[4,1,5]

[1] Climate and Global Dynamics Laboratory, NSF National Center for Atmospheric Research, Boulder, CO, USA

[2] Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, CO, USA

[3] Department of Ecosystem Science & Sustainability, Colorado State University, Fort Collins, CO, USA

[4] Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, USA

[5] Polar Bears International, Bozeman, MT, USA

* Corresponding author

Abstract

Internal variability creates a range of climate trajectories, which are superimposed upon the forced response. A single model realization may not represent forced climate change alone and may diverge from observations due to internal variability. We use an initial-condition large ensemble of simulations with the Community Earth System Model (CESM2) to show that internal variability produces a range of outcomes in the terrestrial carbon cycle. Trends in gross primary production (GPP) from 1991–2020 differ among ensemble members due to disparate temporal sequences in temperature, precipitation, and other physical drivers. We develop a method to quantify internal variability and apply it to the observational record. Observed changes in GPP at two long-running eddy covariance flux towers are consistent with internal variability, challenging the understanding of forced changes in the carbon cycle at these locations. A probabilistic framework that accounts for internal variability is needed to interpret carbon cycle trends.

Introduction

Climate change is evident in numerous disparate observations of the Earth system. The increase in atmospheric CO_2 measured at Mauna Loa, Hawaii, since 1959 is one of the iconic records of global change[1,2], as is the planetary warming seen in the time series of surface temperature measurements[3,4,5] and reflected in Arctic sea-ice loss[6]. Multidecadal changes in the biosphere, both greening and browning, are found in satellite-derived vegetation indices[7,8,9]. Further evidence for a changing biosphere is obtained from the worldwide network of eddy covariance flux towers, which measure energy, water, and CO_2 exchange between the biosphere and atmosphere[10,11,12]. Analyses of flux tower measurements find temporal increases in terrestrial productivity at many locations[13,14,15,16,17,18,19,20,21,22]. The time period over which eddy covariance flux towers have operated is comparatively short, however, and the datasets typically span 10–20 years of data. Carbon cycle trends observed over 24 years at Harvard Forest (1992–2015) and 25 years at Howland Forest (1996–2020) are the longest analyses to date[19,20]. Although the trends have been interpreted in terms of changes in climate, CO_2 concentration, and other forcings, the extent to which unforced variability in the climate system influences the trends is unknown.

The chaotic behavior of the atmosphere and its coupling with the ocean generates unforced variability at timescales from several days to decades[23,24,25,26]. Unforced variability (also known as internal variability) is evident in climate simulations over the historical era and projections of future climate change over the twenty-first century. Small perturbations of the initial atmosphere and ocean states produce different climate trajectories over the simulation period due to internal variability that is largely unpredictable. Large ensembles (typically 30–100 members) performed with a single model, in which the ensemble members differ only in their initial conditions, quantify the range of outcomes due to internal variability[23,25,27,28,29]. Each simulation is a unique expression of the random sequence of unforced variability, and each is equally plausible in its depiction of climate change.

Internal variability is seen in the range among ensemble members in multidecadal temperature and precipitation trends at regional scales[23–28,30,31,32]. Internal variability is a large source of uncertainty in climate projections at regional spatial scales and decadal timescales[26,28,33,34]. Internal variability can mask anthropogenic influences on climate, seen in the concept of time of emergence of the forced signal[35,36]. The observational record, itself, is just one of many possible trajectories by which temperature and precipitation trends could have unfolded as a result of internal variability. Indeed, one can construct an observational large ensemble based on the general statistics of the single observed record[32,37,38,39].

Internal variability is evident in other components of the Earth system including sea ice[40,41,42], snowmelt[43,44], sea level rise[45,46], and ocean biogeochemistry[47,48,49]. Less studied is the internal variability of the terrestrial carbon cycle. However, internal variability in temperature and precipitation can impart unforced variability in the terrestrial carbon cycle that can mask the forced signal[50,51].

Two analyses of annual gross primary production (GPP) at the Harvard Forest EMS eddy covariance flux tower (AmeriFlux US-Ha1; 42.5378°N, 72.1715°W) illustrate variability in GPP trends. Annual GPP increased over the period 1992–2004 at a rate of 36.3 g C m^{-2} yr^{-1} per year[13]. A longer time series that extends the observations to 2015 has a smaller annual trend equal to 23.3 g C m^{-2} yr^{-1} per year over the 24-year period[19]. One interpretation is that the post-2004 data evidence a change in the carbon cycle[11]. An alternative, but untested, interpretation is that the two trends differ as a result of internal variability.

We use a 50-member initial-condition large ensemble for the Community Earth System Model version 2 (CESM2) driven with historical forcing for 1850–2014 and SSP3-7.0 forcing for 2015–2100 (ref. 29) to examine how internal variability influences trends of annual GPP. We analyze the 30-year period

1991–2020. Thirty years is comparable to the longest observational records in the AmeriFlux network of eddy covariance flux towers[12]. We document the variability among ensemble members in GPP trends and show that the standard error of the linear regression trend obtained from the time series of a single ensemble member estimates the internal variability in trends across all 50 ensemble members. We apply this finding to estimate the internal variability of the AmeriFlux data for Harvard Forest over the period 1992–2020 (ref. 52) and calculate the probability of obtaining the different trends reported for 1992–2004 and 1992–2015. We demonstrate that internal variability generates sampling differences over the two time periods consistent with the observed trends. We supplement this observational analysis with additional AmeriFlux data for Morgan-Monroe State Forest (US-MMS; 39.3232°N, 86.4131°W) for 1999–2020 (ref. 53) and Howland Forest (US-Ho1; 45.2041°N, 68.7402°W) for 1996–2020 (ref. 20).

Results

Simulated GPP trends

Across much of North America, the CESM2 ensemble mean, which is indicative of the model's response to anthropogenic emissions, has a statistically significant increase in annual GPP from 1991 to 2020 (Fig. 1a). However, there is considerable variability among individual ensemble members. The ratio of the ensemble mean trend to the standard deviation of trends across ensemble members provides a measure of signal-to-noise (Fig. 1b). The forced signal (i.e., the ensemble mean) exceeds the noise (i.e., ensemble standard deviation) by a factor of four across portions of eastern Canada, Northeast US, and Southeast US. Elsewhere, the signal-to-noise ratio is less than two in Alaska and much of the contiguous US, and it is less than one in the Southwest extending into Mexico and in a broad region extending from the Canadian prairie to Midwest US. Two ensemble members with small and large continental mean trends illustrate the ensemble variability (Fig. 1c,d). Much of Canada has a statistically significant positive GPP trend in both members, though the magnitude varies between members. In contrast, GPP trends across Alaska are small and not statistically significant in ensemble member 1281.015 but are large and significant in 1231.018. GPP trends are negative in portions of Midwest US extending into the Canadian prairie in 1281.015, but not in 1231.018. Supplementary Fig. 1 further highlights ensemble variability in trends for 18 members chosen at random. Large ensemble variability is evident in Alaska, Northwest Canada, and the US west of the Mississippi River, and not all members have a statistically significant GPP trend in these regions.

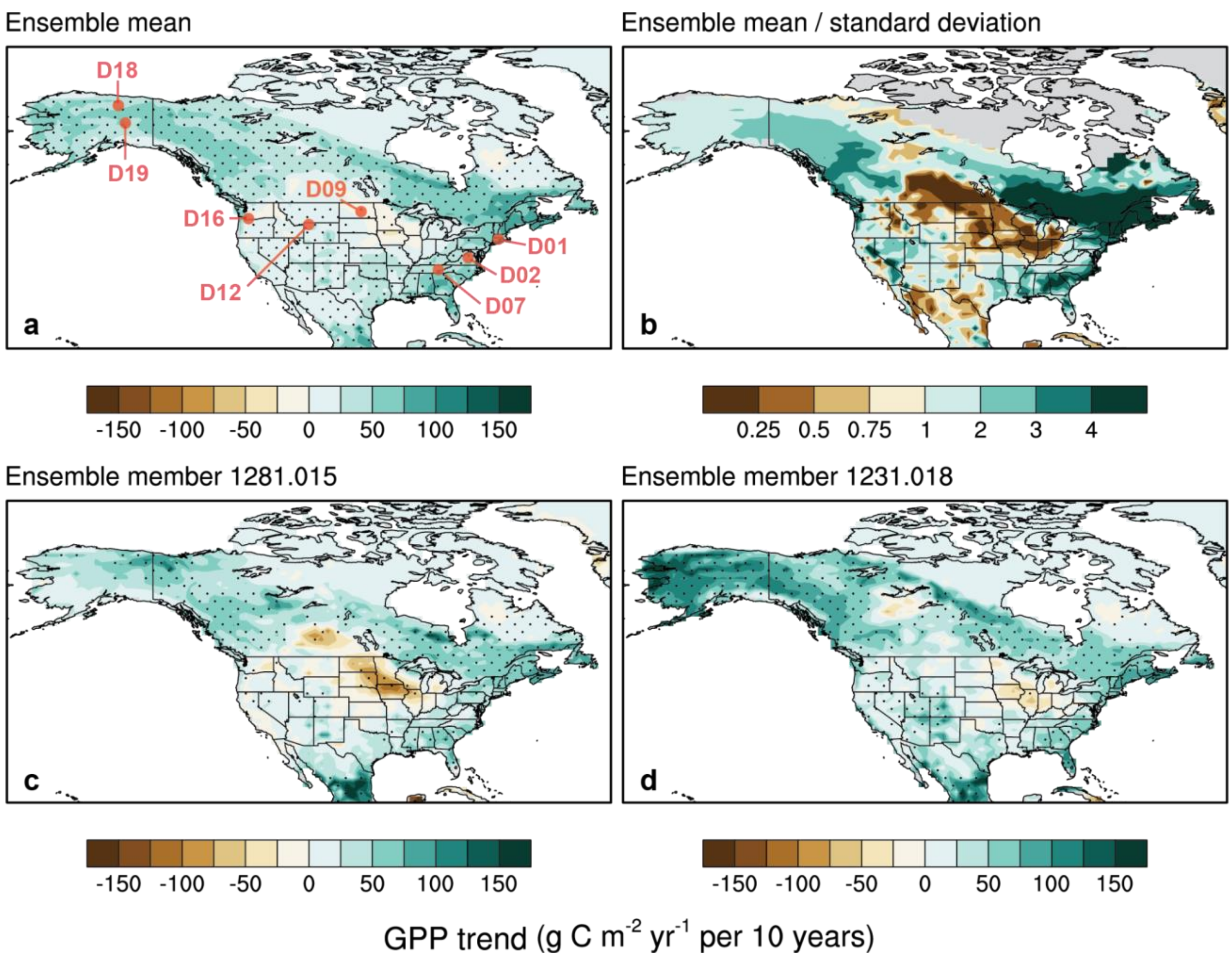

Fig. 1. Trends in annual GPP for 1991–2020. (**a**) Mean trend for the 50-member ensemble obtained from the ensemble mean time series. The colored circles indicate the location of 8 grid cells analyzed in

Fig. 2 and Supplementary Fig. 2. (**b**) Signal-to-noise ratio defined as the ensemble mean trend (absolute value) divided by the standard deviation of trends across the 50 members. Also shown are trends for two members with small (**c**) and large (**d**) continental mean trends. Ensemble members 1281.015 and 1231.018 are the members at the 3rd (6th percentile) and the 47th (94th percentile) ranks, respectively, based on continental mean trends. Stippling denotes statistical significance (n = 30 years; $p \leq 0.05$) using the ensemble mean time series or the individual ensemble member time series. Trends are multiplied by 10 to report the change in annual GPP per 10 years.

Histograms of trends across the 50 ensemble members for individual grid cells illustrate the ensemble variability (Fig. 2). In the Northeast, all of the members have a statistically significant trend, but the 95% confidence interval ranges from 45 to 100 g C m^{-2} yr^{-1} per 10 years. Ensemble variability is larger at the Taiga location, where only half of the ensemble members (52%) have a statistically significant trend, and the 95% confidence interval is 12–109 g C m^{-2} yr^{-1} per 10 years. Only one-quarter (26%) of the members have a statistically significant trend at the Mid-Atlantic grid cell, where the trend varies from negative in two members to greater than 100 g C m^{-2} yr^{-1} per 10 years in two members. Only 10 ensemble members (20%) have a statistically significant trend at the Northern Plains grid cell, and the trend ranges from negative (8 members) to positive (2 members). Similar ensemble variability is seen at other locations (Supplementary Fig. 2).

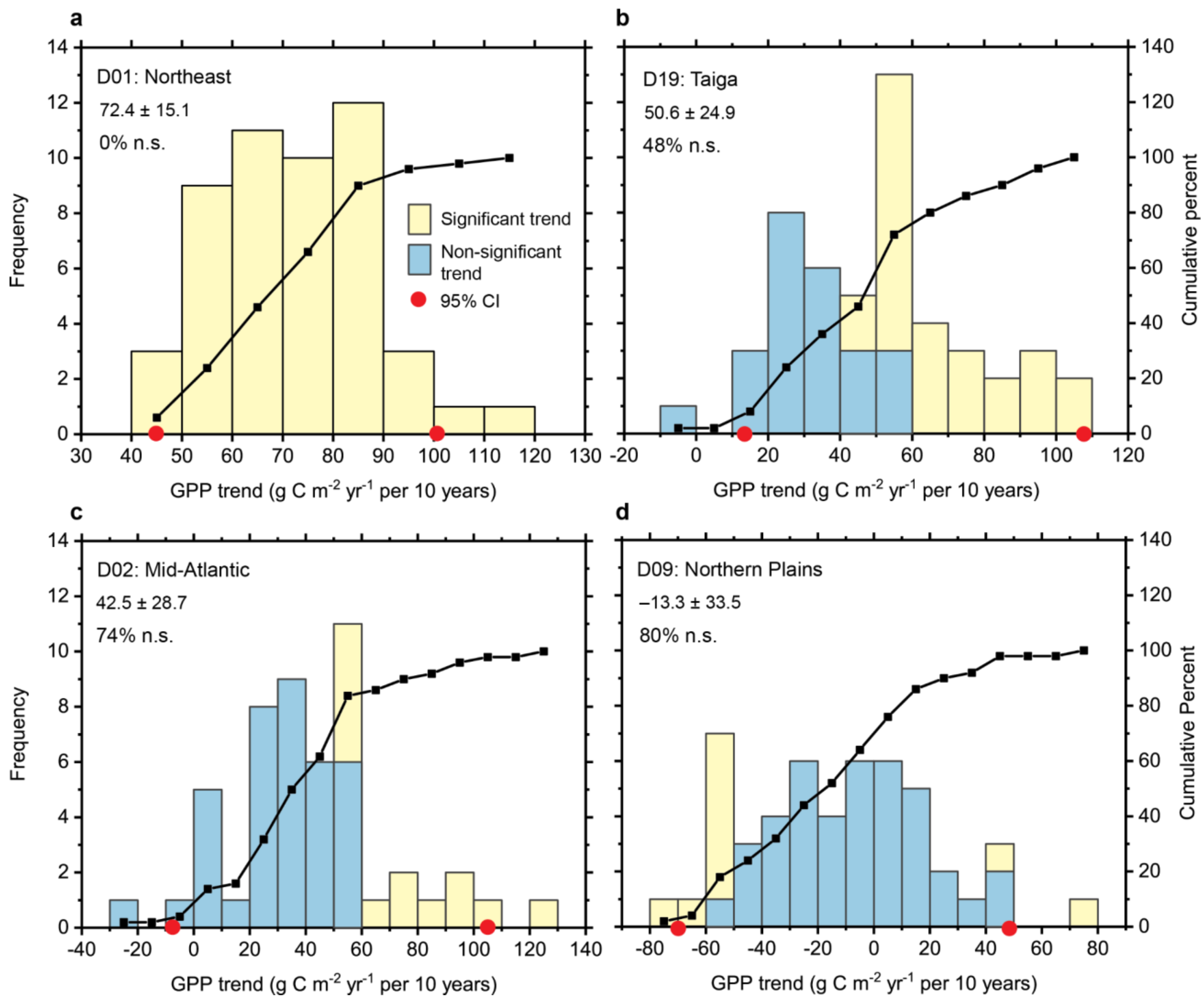

Fig. 2. Histogram of annual GPP trends for 1991–2020 at four grid cells. Grid cells correspond to the location of core terrestrial sites for four domains in the National Ecological Observatory Network (NEON). See Fig. 1 for the location of the sites. Panels show (**a**) D01: Northeast, (**b**) D19: Taiga, (**c**) D02: Mid-Atlantic, and (**d**) D09: Northern Plains sites. The left axis is the frequency distribution for the n = 50 ensemble members, and the black line is the cumulative distribution (right axis). Yellow shading shows members with a statistically significant trend (n = 30 years; $p \leq 0.05$), and light blue shading shows non-significant trends. The mean ± standard deviation and the percentage of members with a non-significant (n.s.) trend are provided in the upper left of each panel. Also shown is the 95% confidence interval (red circles) obtained as the range of trends (n = 48) after excluding the smallest and largest trends.

In British Columbia, eastern portions of Canada, Northeast US, and parts of Southeast US, more than 90% of the members have a statistically significant positive GPP trend (Fig. 3a). Other regions show large variability among ensemble members. In Alaska, the ensemble mean trend is statistically significant (Fig. 1a), but only about half of the members (40–60%) have a statistically significant trend across much of the region (Fig. 3a). A wide region of the interior continent has a significant positive trend in at least one but less than 10 (20%) of the members. The negative GPP trend in the Canadian prairie extending into Midwest US is statistically significant in only 10–30% of the members (Fig. 3b).

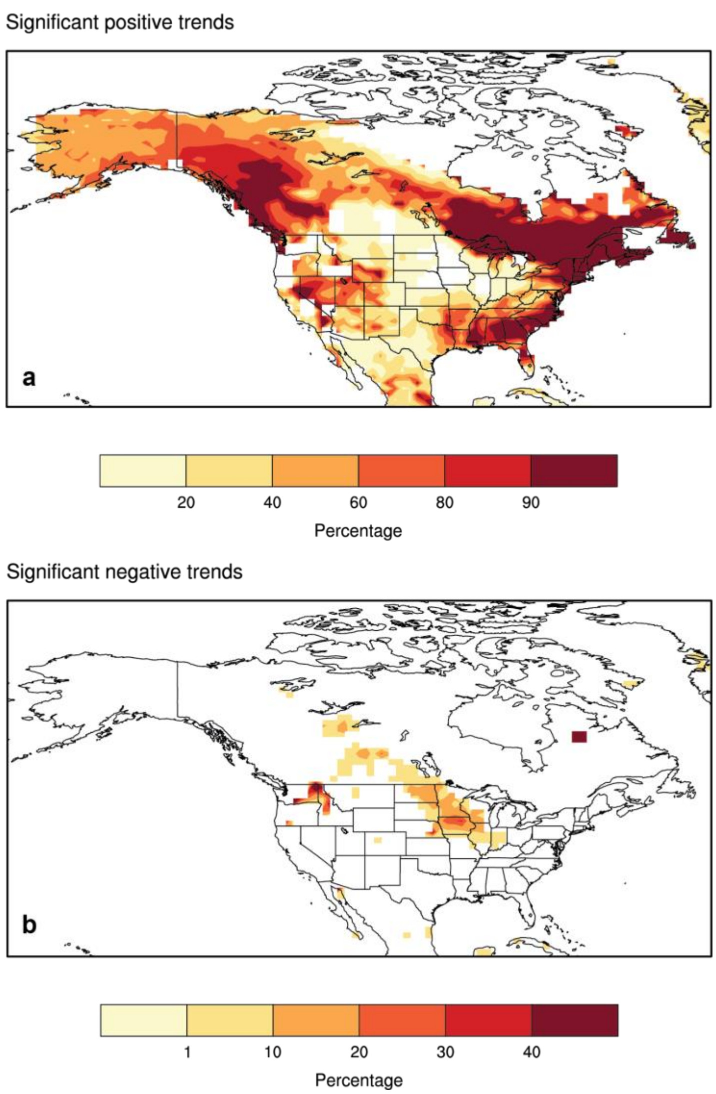

Fig. 3. Percentage of ensemble members with statistically significant trends in annual GPP for 1991–2020. Percentages are given for (**a**) positive and (**b**) negative trends. Non-significant trends (n = 30 years; $p > 0.05$) are masked.

The 95% confidence interval for annual GPP trends obtained directly from the 50-member ensemble shows a wide range of trends among members (Fig. 4a). The range exceeds 100 g C m^{-2} yr^{-1} per 10 years across portions of Alaska, northern Canada, the Canadian prairie extending into Midwest US, the Mid-Atlantic region, and the Central Plains extending into Mexico. GPP trends range from negative to positive in some regions, most prominently in the Canadian prairie extending into Midwest US (Supplementary Fig. 3).

The standard error of the linear regression trend (s_{b1}, equation 3), which quantifies the interannual variability about the linear trend within a single ensemble member, is also an estimate of the variability in trends among ensemble members. That s_{b1} samples internal variability has been shown previously for temperature and precipitation[54], and a similar result pertains to GPP. The 95% confidence interval obtained using s_{b1} for a single ensemble member (Fig. 4b) approximates the 95% confidence interval of the 50-member ensemble (Fig. 4a). This is also evident for other ensemble members (Supplementary Fig. 4). Differences between the two estimates are mostly within ± 25 g C m^{-2} yr^{-1} per 10 years (Supplementary Fig. 5). A prominent exception is a region of Canada extending from the Northwest Territories into Saskatchewan, where the difference is larger. The magnitude of s_{b1} varies among ensemble members. However, the statistical distribution of confidence intervals obtained from s_{b1} includes the 95% confidence interval of the 50-member ensemble. This is evident at the Northeast location, where all ensemble members have a statistically significant trend (Fig. 2a) and the variability among members in s_{b1} (and therefore 95% confidence intervals) is small (Fig. 4c). Ensemble variability is larger at the Taiga location and only half of the ensemble members have a statistically significant trend

(Fig. 2b), but the 95% confidence intervals obtained from s_{b1} still encompass that obtained from the 50-member ensemble (Fig. 4d). Similar results are found at the other locations (Supplementary Fig. 6).

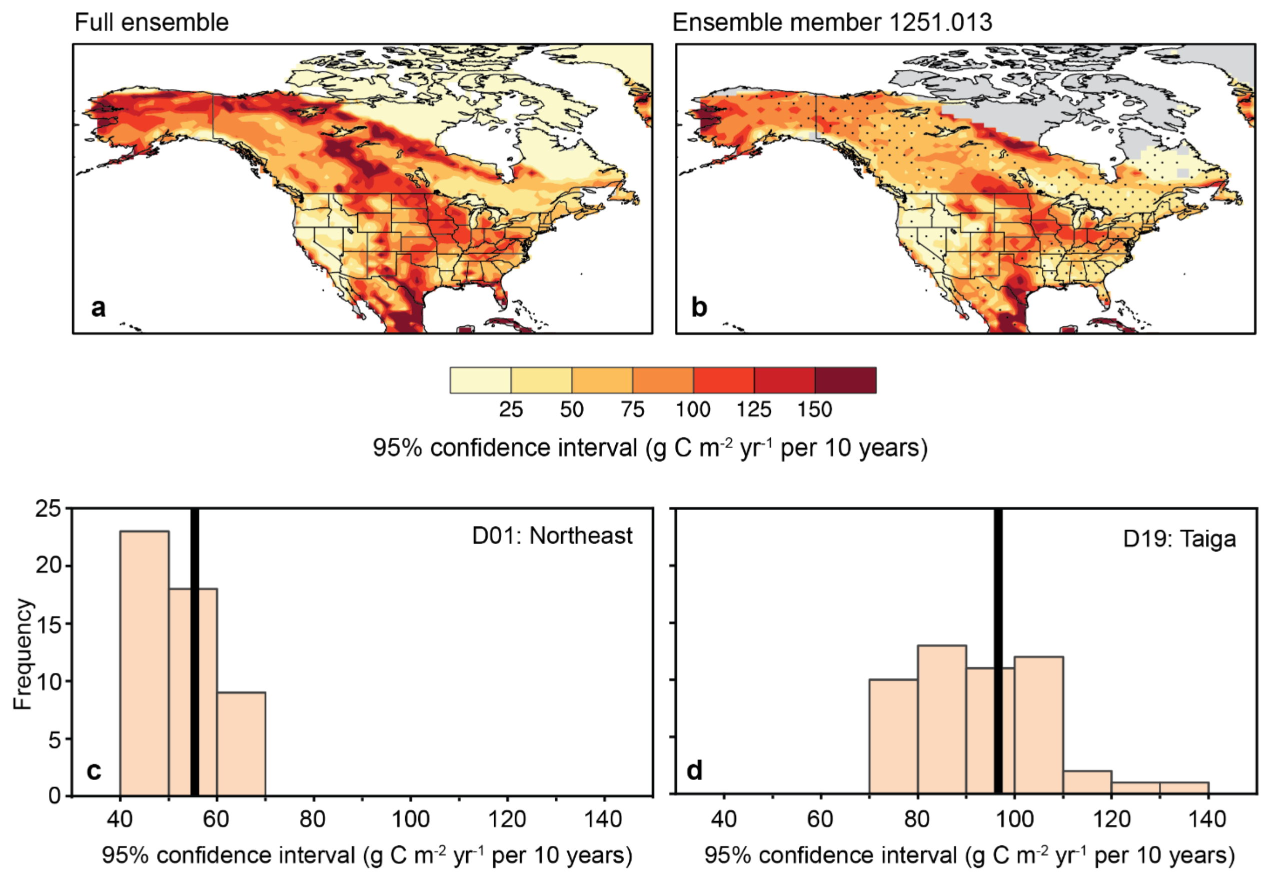

Fig. 4. 95% confidence interval in annual GPP trends for 1991–2020. (**a**) The 95% confidence interval obtained directly from the 50-member ensemble. It is the range of trends (*n* = 48) after excluding the smallest and largest trends for each grid cell. (**b**) The 95% confidence interval obtained from the standard error of the regression trend (s_{b1}). The confidence interval is $2 * 2.048 * s_{b1}$, where $t_{0.975,28} = 2.048$ is the critical t-value for $n = 30$ years of data. Shown is an ensemble member chosen at random. Stippling shows where the trend is statistically significant (*n* = 30 years; p ≤ 0.05). (**c**) Frequency distribution for the 50-member ensemble of the 95% confidence interval obtained from s_{b1} at the grid cell corresponding to the D01: Northeast location. The confidence interval for each ensemble member is

calculated as in (**b**). The thick black line is the 95% confidence interval obtained directly from the 50-member ensemble as in (**a**). (**d**) As in (**c**), but for D19: Taiga.

Correlation of GPP with temperature and precipitation

The CESM2 Large Ensemble also has internal variability in temperature and precipitation, which manifests in the GPP trends. Although all regions of North America have a statistically significant warming trend in the ensemble mean (i.e., the forced trend), the amount of warming varies across the 50-member ensemble due to internal variability (Supplementary Fig. 7). Trends over the 30-year period 1991–2020 are non-significant across Alaska and Northwest Canada in ensemble member 1301.013 but exceed 2°C warming across much of North America (and greater than 3°C in some regions) in ensemble member 1011.001. The signal-to-noise ratio exceeds two over much of North America. Annual precipitation increases in some regions of North America in the ensemble mean, but with considerable variability among ensemble members (Supplementary Fig. 8). Notably, precipitation in Southeast US, which increases significantly in the ensemble mean, decreases in some members and increases in others. The signal-to-noise ratio for precipitation is less than one over much of North America. The 30-year trends for GPP and temperature are positively correlated in seasonally cold climates and negatively correlated in the dry climates of the interior plains region (Fig. 5a). The GPP trends are positively correlated with precipitation trends across much of North America, with largest correlations in the interior region of the US (Fig. 5b). In this region, warm years tend to have low rainfall and vice versa (Fig. 5c).

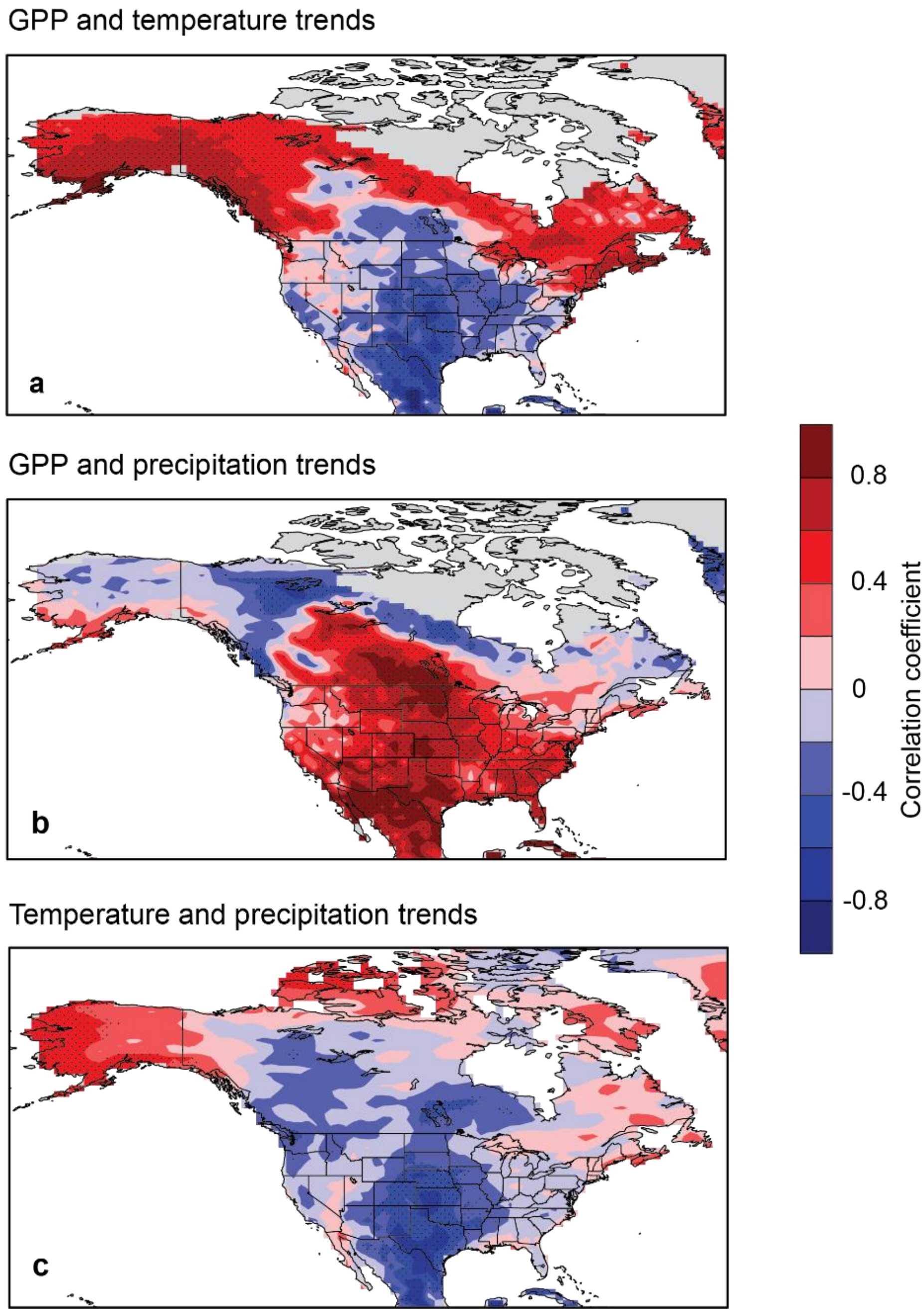

Fig. 5. Correlation across the 50 ensemble members of the 30-year trends (1991–2020). (**a**) GPP and surface air temperature, (**b**) GPP and precipitation, and (**c**) temperature and precipitation. Stippling denotes statistically significant correlations (n = 50, $p \leq 0.05$).

Internal variability of observed trends

Annual GPP at the AmeriFlux Harvard Forest EMS eddy covariance flux tower (US-Ha1; 42.5378°N, 72.1715°W) increased at a rate of 127.1 ± 41.3 g C m^{-2} yr^{-1} per 10 years for the period 1992–2020 (Fig. 6a). The CESM2 Large Ensemble underestimates annual GPP at the grid cell corresponding to Harvard Forest over the 1992–2020 observational period (Fig. 6b). The trend across the 50 members is 73.0 ± 14.1 g C m^{-2} yr^{-1} per 10 years (mean ± standard deviation), with a range of 48–114 g C m^{-2} yr^{-1} per 10 years. Although the mean trend is less than the observations, the distribution of trends obtained from the ensemble falls within the observational uncertainty (Fig. 6c). However, the variability of CESM2 trends (14.1 g C m^{-2} yr^{-1} per 10 years) is one-third the observed variability (41.3 g C m^{-2} yr^{-1} per 10 years).

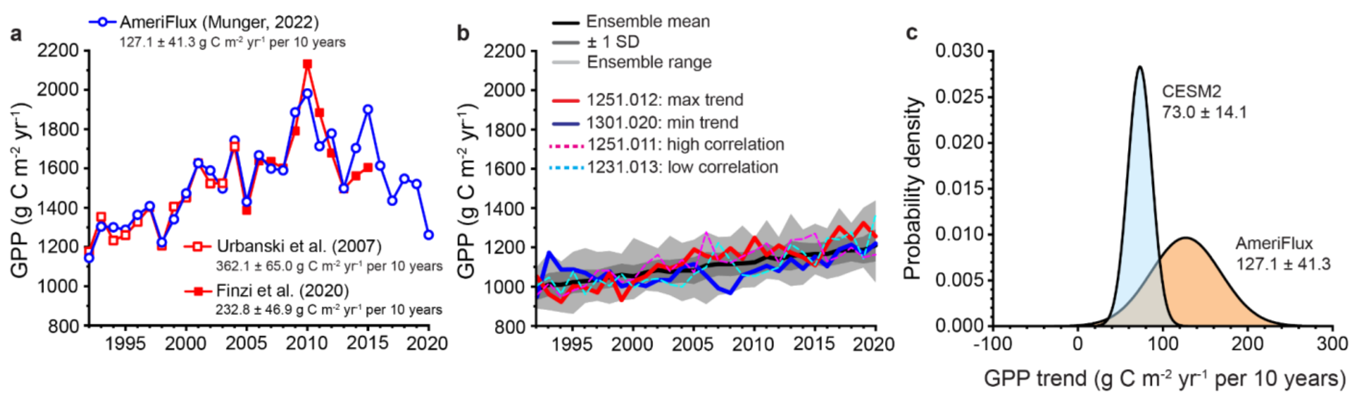

Fig. 6. Observed and simulated annual gross primary production (GPP) at the AmeriFlux US-Ha1 (Harvard Forest) flux tower. (**a**) Observed time series at Harvard Forest published by Urbankski et al. (ref. 13) for 1992–2004, Finzi et al. (ref. 19) for 1992–2015, and the AmeriFlux data (ref. 52) for 1992–2020. The Urbanski et al. data are indistinguishable from the Finzi et al. data over the same time period. Shown are the linear regression slope ± standard error for the three datasets. See Supplementary Table 1 for the data. (**b**) Simulated time series from the 50-member CESM2 Large Ensemble for the grid cell corresponding to the Harvard Forest tower location. The black line is the ensemble mean, the dark gray

shading shows ± one standard deviation across all ensemble members, and the light shading shows the ensemble range. Also shown are four ensemble members. The red line is the ensemble member with the largest trend, and the blue line is the ensemble member with the smallest trend. The dashed magenta and cyan lines are the ensemble members with high and low temporal correlation with the AmeriFlux data, respectively. (**c**) Statistical distribution of trends from the CESM2 Large Ensemble in comparison with the AmeriFlux data for 1992–2020. The model trends are normally distributed (mean ± standard deviation, 73.0 ± 14.1 g C m^{-2} yr^{-1} per 10 years). Also shown is the trend estimated using the AmeriFlux data (127.1 ± 41.3 g C m^{-2} yr^{-1} per 10 years).

Comparable analyses at Morgan-Monroe State Forest (US-MMS; 39.3232°N, 86.4131°W) show broad overlap between model and observed GPP trends (Fig. 7a), but not at Howland Forest (US-Ho1; 45.2041°N, 68.7402°W) (Fig. 7b). At both locations, the variability of trends in the CESM2 Large Ensemble is comparable to the observed variability.

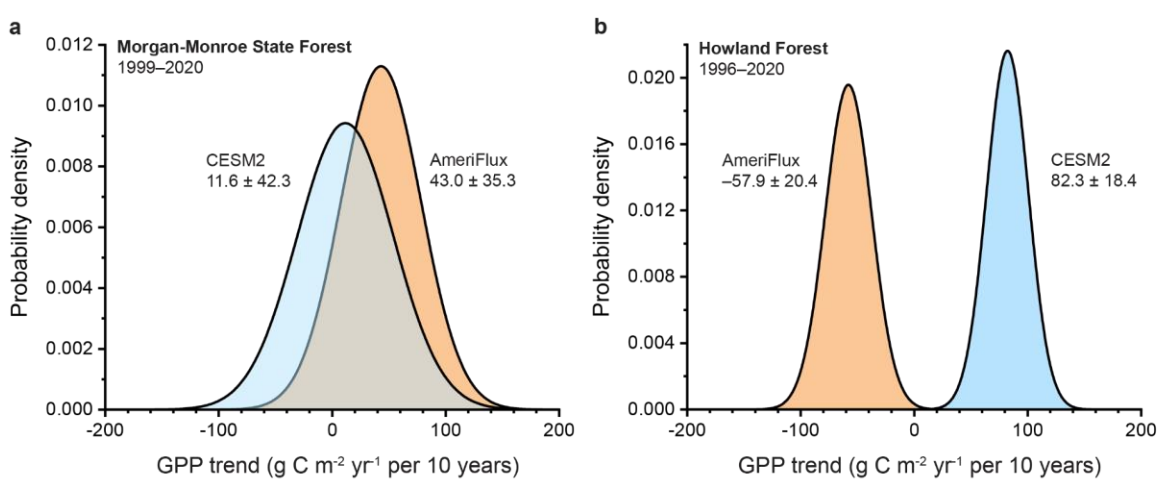

Fig. 7. Observed and simulated annual GPP trends at two AmeriFlux sites. (**a**) Statistical distribution of trends at US-MMS (Morgan-Monroe State Forest) for 1999–2020. Trends from the CESM2 Large Ensemble are normally distributed (mean ± standard deviation, 11.6 ± 42.3 g C m^{-2} yr^{-1} per decade).

Also shown is the trend estimated using the AmeriFlux data (ref. 53). See Supplementary Table 2 for the data. (**b**) As in (**a**), but for US-Ho1 (Howland Forest) for 1996–2020 with observations from Hollinger et al. (ref. 20). See Supplementary Table 3 for the data.

The Harvard Forest data show considerable variability in GPP trends depending on the time period sampled (Fig. 6a). Annual GPP increased over the period 1992–2004 at a rate of 362.1 ± 65.0 g C m^{-2} yr^{-1} per 10 years using data reported by Urbanski et al. (ref. 13). A subsequent dataset by Finzi et al. (ref. 19) that extends the observations to 2015 has a smaller trend for 1992–2015 (232.8 ± 46.9 g C m^{-2} yr^{-1} per 10 years). We used Monte Carlo methods to determine the conditional probability of obtaining these two GPP trends given the long-term forced trend. We calculated the probability of obtaining a trend of 362.1 g C m^{-2} yr^{-1} per 10 years over the 13-year period 1992–2004 and a trend of 232.8 g C m^{-2} yr^{-1} per 10 years over the 24-year period 1992–2015 if the long-term forced trend is 127.1 ± 41.3 g C m^{-2} yr^{-1} per 10 years.

Fig. 8a shows annual GPP from 1992 to 2004 in two time series that draw GPP for each year as a random deviate about the long-term forced trend. Both time series have a forced trend of 127.1 g C m^{-2} yr^{-1} per 10 years, but annual GPP decreases by –143.7 g C m^{-2} yr^{-1} per 10 years in one time series and increases by 397.9 g C m^{-2} yr^{-1} per 10 years in the other. Fig. 8b shows the statistical distribution of trends obtained from Monte Carlo simulations with 100,000 randomly sampled time series. The mean (127.0 g C m^{-2} yr^{-1} per 10 years) is comparable to the forced trend, and the standard deviation is larger (138.2 vs. 41.3 g C m^{-2} yr^{-1} per 10 years) because of the smaller number of years sampled (see equation 3). The 95% confidence interval spans –144 to 398 g C m^{-2} yr^{-1} per 10 years (the time series shown in Fig. 8a are the 2.5 and 97.5 percentiles). The observed trend of 362.1 g C m^{-2} yr^{-1} per 10 years falls within the 95% confidence interval. There is a 4.4% chance of obtaining a trend equal to or greater than the observed trend if the forced trend is 127.1 g C m^{-2} yr^{-1} per 10 years. There is a 10% chance that the trend equals or exceeds 304 g C m^{-2} yr^{-1} per 10 years and a 5% change of a value equal to or greater than 354 g C m^{-2} yr^{-1} per 10 years. With a longer time series spanning 1992–2015, the 95% confidence interval for trends is 20–235 g C m^{-2} yr^{-1} per 10 years (Fig. 8c). The observed trend of 232.8 g C m^{-2} yr^{-1} per 10 years for this time period falls within the uncertainty range (Fig. 8d). There is a 2.7% chance of obtaining a trend equal to or greater than the observed trend. The 10% and 5% thresholds are 197 and 217 g C m^{-2} yr^{-1} per 10 years, respectively.

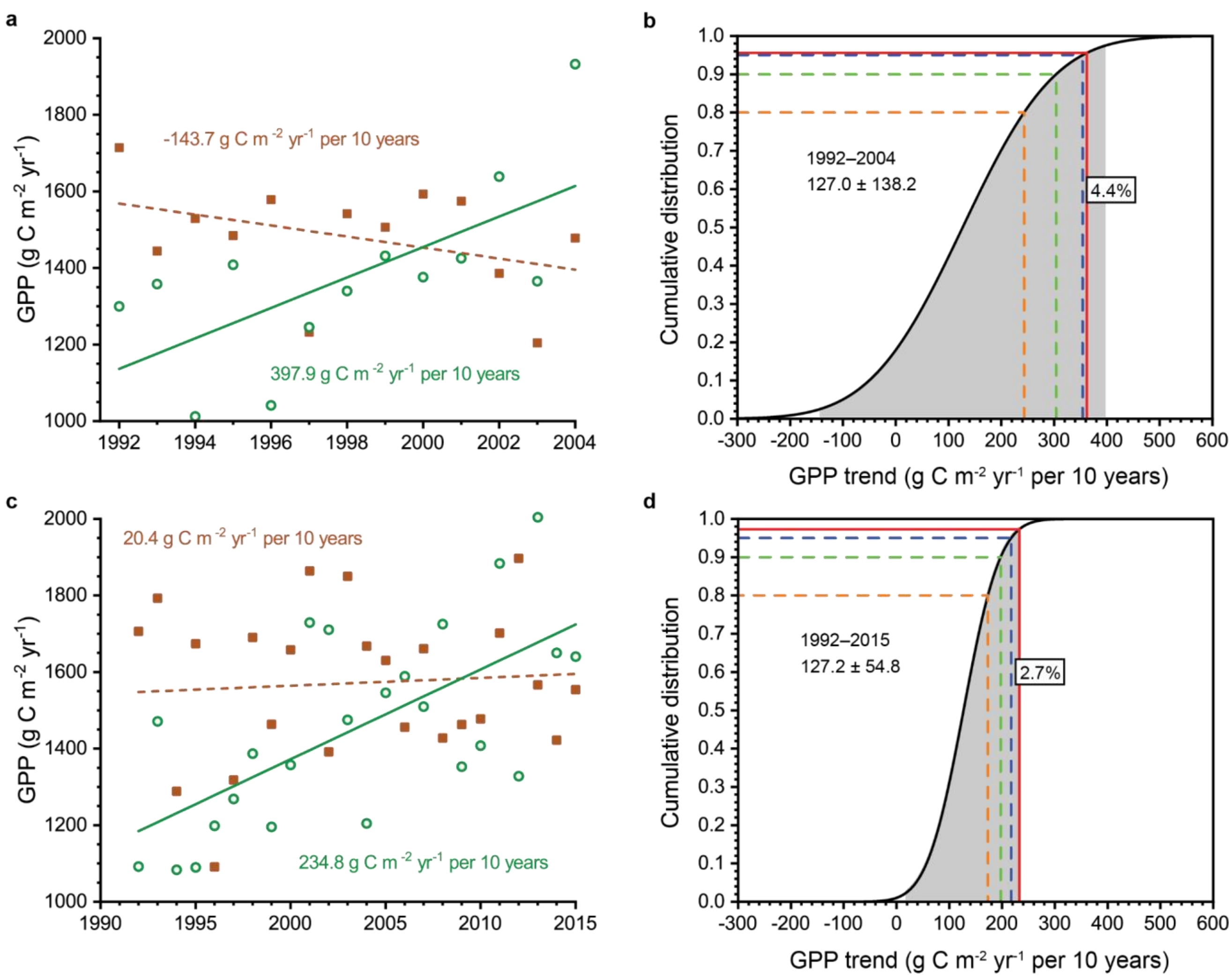

Fig. 8. Conditional probability of GPP trends at the AmeriFlux US-Ha1 (Harvard Forest) tower. (a) Annual GPP for 1992–2004 for two time series in which GPP for each year is chosen as a random deviate

about the 1992–2020 forced trend. The time series are the endpoints of the 95% confidence interval in the Monte Carlo simulations. The brown squares and dashed line show the 2.5th percentile, and the dark green open circles and solid line show the 97.5th percentile. (**b**) Conditional probability distribution of trends. Shown is the cumulative distribution of trends for 1992–2004 obtained from 100,000 randomly sampled time series. The trends are normally distributed with a mean and standard deviation of 127.0 ± 138.2 g C m^{-2} yr^{-1} per 10 years. The gray shading is the 95% confidence interval, and the two time series in panel (**a**) show the endpoints. The red line is the probability of a trend greater than that observed for 1992–2004. Dashed lines show the values for which there is a 20% (orange line), 10% (green line), and 5% (blue line) chance of a greater trend. (**c**)–(**d**) Same as (**a**) and (**b**), but for 1992–2015.

Annual GPP observations at Morgan-Monroe also show variability in trend estimates. Dragoni et al. (ref. 14) found that carbon storage increased over the 10-year period 1999–2008. Our analysis of the AmeriFlux dataset for Morgan-Monroe (ref. 53) finds that annual GPP increased by 208.7 ± 89.9 g C m^{-2} yr^{-1} per 10 years during 1999–2008, decreasing to 43.0 ± 35.3 g C m^{-2} yr^{-1} per 10 years for the full 22-year time series spanning 1999–2020 (Fig. 9a). Monte Carlo analysis similar to those at Harvard Forest show that a forced trend of 43.0 ± 35.3 g C m^{-2} yr^{-1} per 10 years has a 95% confidence interval of –184 to 270 g C m^{-2} yr^{-1} per 10 years when sampled over the 10-year period 1999–2008 (Fig. 9b). The observed trend for 1999–2008 falls within the 95% uncertainty range. There is a 7.6% chance of obtaining a trend equal to or greater than the observed trend if the forced trend is 43.0 g C m^{-2} yr^{-1} per 10 years. There is a 10% chance that the trend exceeds 191 g C m^{-2} yr^{-1} per 10 years and a 5% chance of a value greater than 233 g C m^{-2} yr^{-1} per 10 years.

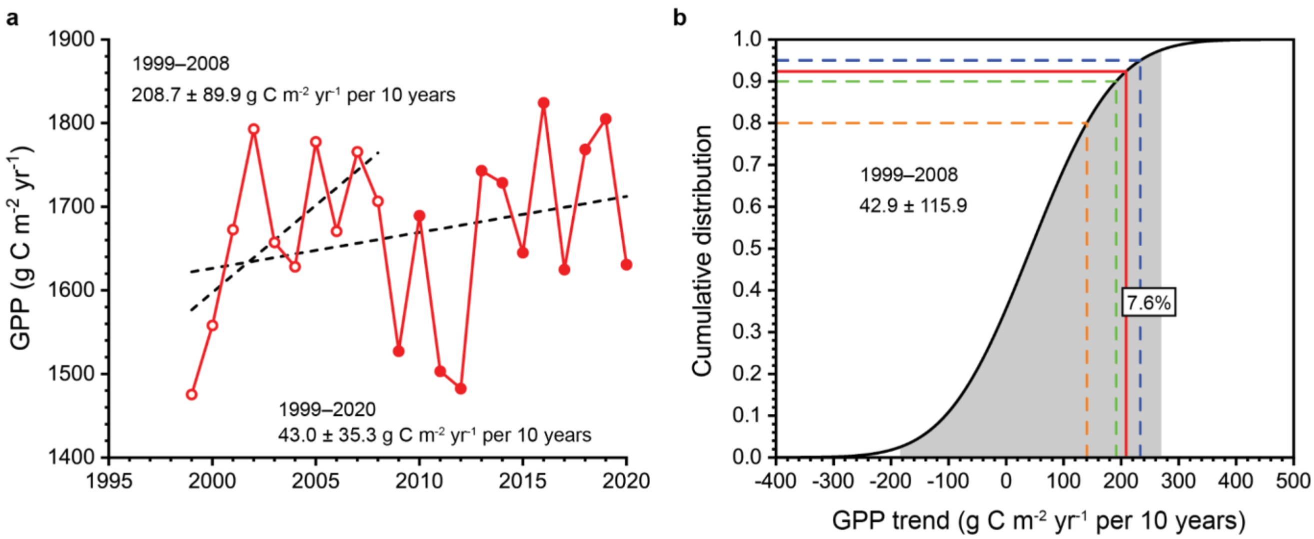

Fig. 9. Annual GPP trends at the AmeriFlux US-MMS (Morgan-Monroe State Forest) tower. (**a**) The full 1999–2020 AmeriFlux time series (ref. 53). Open circles show the years 1999–2008 and closed circles extend the dataset to 2020. Shown are the linear regression (dashed lines) with the regression slope ± standard error for the two time periods. See Supplementary Table 2 for the data. (**b**) Conditional probability distribution of trends. Shown is the cumulative distribution of trends for 1999–2008 obtained with Monte Carlo methods using a forced trend of 43.0 ± 35.3 g C m^{-2} yr^{-1} per 10 years. The trends are normally distributed with a mean and standard deviation of 42.9 ± 115.9 g C m^{-2} yr^{-1} per 10 years. The gray shading is the 95% confidence interval. The red line is the probability of a trend greater than that observed for 1999–2008. Dashed lines show the values for which there is a 20% (orange line), 10% (green line), and 5% (blue line) chance of a greater trend.

Discussion

Our analysis of the 50-member CESM2 Large Ensemble shows that internal variability creates ambiguity in the magnitude and sign of GPP trends when only a single model realization is analyzed. The ensemble mean, however, reflects the forced response. The key inference pertains to how to interpret carbon cycle trends, both in model simulations and in observations.

Internal variability necessitates caution when comparing a single model realization to the observational record. At the model grid cell corresponding to Harvard Forest, the ensemble average GPP trend over 1992–2020 is 73 g C m^{-2} yr^{-1} per 10 years and the range across ensemble members is 48–114 g C m^{-2} yr^{-1} per 10 years (Fig. 6c). A single realization at the low end of the distribution would lead to a conclusion that the model is biased low compared with the observed trend of 127 g C m^{-2} yr^{-1} per 10 years, whereas a simulation at the high end would suggest closer fidelity to the observations. In fact, the distribution of trends across the 50-member ensemble broadly overlaps with the observed trend and its uncertainty. Similar ambiguity arises in comparison with observations at Morgan-Monroe State Forest (Fig. 7a). The ensemble mean trend (12 g C m^{-2} yr^{-1} per 10 years) suggests the model is biased low compared with the observations (43 g C m^{-2} yr^{-1} per 10 years), but the statistical distribution of trends from the large ensemble broadly encompasses the observed trend. Conversely, the high bias at Howland Forest is robust across all ensemble members, and we can confidently conclude the model fails to capture the observed decline in GPP (Fig. 7b).

CESM2 can, in some locations, produce a large positive GPP trend, no trend, and even a negative trend depending on the sequence of internal variability, which is superimposed on the forced response (Fig. 1, Supplementary Fig. 1). Improving the component land model's process parameterizations or adjusting parameters so that a single realization better matches observations risks overfitting, with consequent spurious performance in another realization. Likewise, land models are commonly evaluated in uncoupled simulations forced with meteorological observations[55,56], but alternative reconstructions of historical meteorology, which can be thought of as samples of observational uncertainty, produce different carbon cycle trends[57,58]. A probabilistic comparison of model simulations and observations is needed, with the goal of identifying whether a model is plausible rather than singularly right or wrong[26].

Internal variability also complicates interpretation of the observational record. Harvard Forest is an aggrading forest that is accumulating carbon as it recovers from past agricultural land use, hurricane damage, and wood harvesting[13,19]. Warmer temperature, a longer growing season, and greater precipitation have contributed to increased productivity between 1992 and 2015 (ref. 19). Our analysis does not dispute this understanding of the carbon cycle at Harvard Forest. Rather, we simply interpret the changing carbon cycle in the context of internal variability superimposed upon a forced climate response to anthropogenic emissions. Care needs to be taken in attributing the changing carbon cycle to forced climate change, as indeed is evident in analysis of trends in the physical climate system[59]. The conclusion that forest productivity has increased at Harvard Forest is robust, but the magnitude is uncertain and is influenced by internal variability. Our results show that the large GPP trends for 1992–2004 and 1992–2015 (Fig. 6a) are a manifestation of internal variability and are consistent with a smaller long-term forced trend (Fig. 8b,d). Likewise, there is a long-term positive trend in carbon accumulation at Morgan-Monroe, which can be attributed in part to longer growing seasons[12,14], but which was reduced by severe drought in 2012 (ref. 60). Within this long-term trend, internal variability generates random variability, seen, for example, in a wide range of positive and negative GPP trends (Fig. 7a). The large positive trend found for 1999–2008 is consistent with a much smaller long-term forced trend (Fig. 9b).

The observational record of GPP is one sample from a distribution of possible trajectories. The standard error of the regression trend (s_{b1}) provides an estimate of internal variability for temperature and precipitation[54], and similarly for GPP (Fig. 4). Still unknown, however, is whether the observed trend at Harvard Forest and Morgan-Monroe is a central estimate for the forced response or if it is more representative of end-members of the statistical distribution of trends. Our calculations of conditional probabilities are predicated on the long-term observations as representative of the forced response (Fig. 8, Fig. 9). Other more advanced statistical techniques are available to estimate the observational internal variability for temperature and precipitation[37,38,39]. Similar methods have been used to create an observational ensemble of ocean chlorophyll, for which internal variability creates a wide range of possible trends[49]. Whether the same methods can be applied to create an observational ensemble for the terrestrial carbon cycle is unclear. Nonetheless, our results demonstrate a need to emphasize the standard error, not just the trend, as a key metric of carbon cycle uncertainty.

Interannual variability is one way in which internal variability manifests in the observational record. Interannual variability allows for empirical analysis of carbon cycle responses to temperature and precipitation anomalies, which provides a key constraint on carbon–climate feedbacks[10,61,62]. Our study provides further evidence of the importance of interannual variability for analyzing the carbon cycle. The interannual variability about the forced anthropogenic trend in GPP is a measure of the magnitude of internal variability. CESM2 underestimates interannual variability in GPP compared with observations[63,64], meaning that the importance of internal variability for Earth system model simulations of the terrestrial carbon cycle may be greater than that identified in our study. Our analyses provide qualified findings as to whether CESM2 adequately samples the observational internal variability. The ensemble spread in GPP trends is one-third the observational uncertainty at Harvard Forest (Fig. 6c), but comparable to the observations at Morgan-Monroe and Howland Forest (Fig. 7). Greater effort must be given to quantifying the internal variability of the terrestrial carbon cycle in Earth system models and in estimating the internal variability of the observational record.

The large range in simulated land carbon cycle trends in response to anthropogenic climate change, and the failure to reduce the spread across model generations, has led to focused efforts to reduce model uncertainty[65,66]. Internal variability in air temperature and precipitation trends has been interpreted as irreducible uncertainty in climate projections because of the limited memory in the atmosphere and surface ocean[23,25,26,67]. Similar internal variability, and consequently irreducible uncertainty, occurs in the terrestrial carbon cycle. Further studies are needed to quantify the internal variability of the carbon cycle in both models and observations; to develop the necessary probabilistic framework to understand the changing carbon cycle; and to guide efforts to reduce model uncertainty.

Methods

CESM2 Large Ensemble

We analyzed 50 members of the CESM2 Large Ensemble that differ only in initial conditions[29]. The simulations extend over the period 1850–2100 using historical forcings (1850–2014) and SSP3-7.0 CMIP6 forcings (2015–2100). We used the BB_CMIP6_SM simulations (ensemble members 51–100), in which the prescribed biomass burning emissions were temporally smoothed over the years 1990–2020. The smoothing corrects a discontinuity in the magnitude of interannual variability of the biomass burning emissions used in ensemble members 1–50 that produces spurious warming in northern high latitudes[29,68,69]. CESM2 has a nominal 1° horizontal resolution with active atmosphere, ocean, sea ice, and land component models. The model was initialized from particular years of a preindustrial control simulation and with macro- and micro-perturbations to the initial conditions. The 10-member macro-initializations started from years 1011, 1031, 1051, 1071, 1091, 1111, 1131, 1151, 1171, and 1191. Four sets of 10-member micro-initializations started from years 1231, 1251, 1281, and 1301. Ten members were run for each micro-initialization start year in which spread among the 10 members was generated by a small random perturbation to the atmosphere temperature field at initialization. The start years for the micro-initializations were chosen to sample different states of the Atlantic Meridional Overturning Circulation (AMOC). Rodgers et al. (ref. 29) provide further details of the model configuration, initialization, and forcings. Evaluation of the terrestrial carbon cycle can be found elsewhere[55].

We analyzed the period 1991–2020 to discern trends in annual gross primary production (GPP), surface air temperature, and precipitation for each ensemble member. Memory of initial conditions is minimal at this time period in that the different initializations in 1850 generate similar ensemble variability of GPP trends (Supplementary Fig. 9). Similar to studies of climate trends[23–26,59], we estimated the trend as the linear fit to the 1991–2020 time series using ordinary least squares regression. Statistical significance was determined by regression slopes with $p \leq 0.05$ (n = 30 years). We further analyzed the statistical distribution of GPP trends across the 50-member ensemble at individual model grid cells corresponding to the location of core terrestrial sites in the National Ecological Observatory Network[51].

We quantified the effect of internal variability on the GPP trends using two metrics. The standard deviation of trends across the 50-member ensemble is a direct measure of ensemble variability. The standard error of the regression trend obtained for a single ensemble member, which depends on the interannual variability about the trend, also estimates interval variability, as shown previously for temperature and precipitation[54]. We likewise used the model simulations to assess whether the standard error of the trend obtained from the regression analysis provides an estimate of the internal variability in GPP trends. We compared the standard error of the trend (and the 95% confidence interval for the trend) obtained from individual ensemble members with the actual distribution of trends across the n = 50 ensemble members.

Observational data

We estimated the internal variability in the observational record using long-term annual GPP data obtained from eddy covariance flux towers in the AmeriFlux database. We analyzed GPP at the AmeriFlux US-Ha1 Harvard Forest EMS tower (42.5378°N, 72.1715°W) for the 29-year period 1992–2020 (Supplementary Table 1). We used the AmeriFlux FLUXNET data product[52], which was processed using the ONEFlux processing codes[70] to derive GPP from the measured net ecosystem exchange (NEE). The processing includes friction velocity (ustar) threshold filtering, gap-filling of flux variables, and partitioning of NEE into GPP and ecosystem respiration. We used the GPP_NT_VUT_REF estimate, calculated with nighttime flux partitioning (NT) of NEE to obtain GPP with variable ustar threshold (VUT) and using the most representative NEE after filtering with multiple ustar thresholds (REF). The product compares well to annual GPP data published by Finzi et al. (ref. 19) for 1992–2015 (Supplementary Fig. 10).

We fit a linear regression to the AmeriFlux data (1992–2020) to estimate the long-term annual trend:

$$x_i = b_0 + b_1 * t_i \quad (1)$$

where x_i is annual GPP (g C m^{-2} yr^{-1}) and t_i is year (1992, 1993, …, 2020). The fitted regression for the *n* = 29 year time series is: b_0= –23954.19 g C m^{-2} yr^{-1}, b_1= 12.71 g C m^{-2} yr^{-2}, F = 9.44, p = 0.0048, and R^2 = 0.259. The standard deviation of the residuals is:

$$s_e = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2} = 186.3 \text{ g C m}^{-2}\text{ yr}^{-1} \quad (2)$$

where $\hat{x}_i$ is the predicted GPP for year i using equation (1). The standard error of b_1 is:

$$s_{b1} = s_e \,/\, \sqrt{\sum_{i=1}^{n}(t_i - \bar{t})^2} = s_e \,/\, \sqrt{(n^3 - n)/12} \quad (3)$$

The right-most equation for s_{b1} is the form given by Thompson et al. (ref. 54) when time (t_i) is expressed as n consecutive integers (1992, 1993, …, 2020).

To assess the internal variability of the GPP trend, we used a Monte Carlo approach that statistically samples the observations assuming random interannual variability in GPP. Based on the statistical distribution of the residuals (s_e; supplementary Fig. 11a), we sampled each of the 29 years of data from a random Gaussian deviation about the trend in which GPP for year i is:

$$x_i' = \hat{x}_i + \varepsilon_i * s_e \quad (4)$$

where $\hat{x}_i$ is the predicted GPP for year i using the the linear regression in equation (1), ε_i is a random Gaussian deviate with mean zero and standard deviation of one, and s_e is the standard deviation of the residuals (186.3 g C m^{-2} yr^{-1}). The regression slope (b_1') of the randomly sampled x_i' time series is an estimate of the random variability in the observed trend. We repeated this process 100,000 times to obtain the statistical distribution of b_1'. The resulting probability density function provides the internal variability for the trend. The distribution of b_1', obtained with the assumption of random interannual variability, has a mean (127.0 g C m^{-2} yr^{-1} per 10 years) and standard deviation (41.3 g C m^{-2} yr^{-1} per 10 years) comparable to b_1 and its standard error (Supplementary Fig. 11b).

We then used the statistical model to estimate the conditional probability of obtaining a trend of 362.1 g C m^{-2} yr^{-1} per 10 years for the time period 1992–2004 and 232.8 g C m^{-2} yr^{-1} per 10 years for 1992–2015 (Fig. 6a). In this analysis, we used equation (4), but only sampled the years 1992–2004 and 1992–2015 in the Monte Carlo simulations to obtain the probability density functions for the trend over these two time periods given the long-term trend of 127.1 ± 41.3 g C m^{-2} yr^{-1} per 10 years (Fig. 8b,d). The mean trend is comparable to the long-term trend, and the standard deviation is similar to that expected from equation (3) with $n = 13$ and $n = 24$ years.

We performed the same analysis at the AmeriFlux US-MMS Morgan-Monroe State Forest tower (39.3232°N, 86.4131°W) for the 22-year period 1999–2020 using the AmeriFlux FLUXNET data product (Supplementary Table 2) (ref. 53). Here, we used the daytime flux partitioning product GPP_DT_VUT_REF as in Dragoni et al. (ref. 14). We obtained the linear regression from the observations for the n = 22 years (Fig. 9a; b_0= –6976.29 g C m^{-2} yr^{-1}, b_1= 4.30 g C m^{-2} yr^{-2}, F = 1.48, p = 0.237, R^2 = 0.069, s_e = 105.1 g C m^{-2} yr^{-1}) and used the regression model in the Monte Carlo simulations to sample the years 1999–2008 as in Dragoni et al. (ref. 14). We determined the probability that a trend of 208.7 g C m^{-2} yr^{-1} per 10 years can be found for the period 1999–2008 given the long-term trend of 43.0 ± 35.3 g C m^{-2} yr^{-1} per 10 years (Fig. 9b). The mean trend is comparable to the long-term trend, and the standard deviation is similar to that expected from equation (3) with $n = 10$ years.

We compared GPP trends from the CESM2 Large Ensemble for the grid cell corresponding to Harvard Forest and Morgan-Monroe with the observed trend (Fig. 6c, Fig. 7a). We supplemented this model–observation comparison with GPP data for the AmeriFlux US-Ho1 Howland Forest tower (45.2041°N, 68.7402°W) for 1996–2020 (Supplementary Table 3) (ref. 20). We compared the model and observed trends (Fig. 7b), but did not sub-sample for specific years because only the full 25-year time series has been previously analyzed.

Reporting Summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data Availability

The CESM2 Large Ensemble data that support the findings of this study are available at https://www.cesm.ucar.edu/projects/community-projects/LENS2/data-sets.html. The GPP data for Harvard Forest, Morgan-Monroe State Forest, and Howland Forest are available in the supplement.

Code Availability

The NCAR Command Language (NCL) version 6.4.0 was used for plotting CESM2 data. The Monte Carlo simulations were created using Python version 3.9.12 using Python packages: pandas 1.4.2, numpy 1.21.5, and statsmodels 0.13.2. The code is described in detail in Methods.

Acknowledgments

This material is based upon work supported by the NSF National Center for Atmospheric Research (NCAR), which is a major facility sponsored by the National Science Foundation (NSF) under Cooperative Agreement No. 1852977. The CESM2 Large Ensemble Community Project and supercomputing resources were provided by the IBS Center for Climate Physics in South Korea, DOI:10.5194/esd-12-1393-2021. Funding for the AmeriFlux data portal was provided by the U.S. Department of Energy Office of Science.

Author contributions

G.B. devised the concept for the paper and carried out the analysis. C.D. and W.W. contributed to the analysis. All authors contributed to data interpretation and writing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available online.

References

1. Keeling, C. D. et al. Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii. *Tellus* **28**, 538–551 (1976).

2. Friedlingstein, P. et al. Global carbon budget 2023. *Earth System Science Data* **15**, 5301–5369 (2023).

3. Hansen, J., Ruedy, R., Sato, M. & Lo, K. Global surface temperature change. *Rev. Geophys.* **48**, RG4004 (2010).

4. Rohde, R. A. & Hausfather, Z. The Berkeley Earth land/ocean temperature record. *Earth Syst. Sci. Data* **12**, 3469–3479 (2020).

5. Morice, C. P. et al. An updated assessment of near-surface temperature change from 1850: the HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres* **126**, e2019JD032361 (2021).

6. Box, J. E. et al. Key indictors of Arctic climate change: 1971–2017. *Environmental Research Letters* **14**, 045010 (2019).

7. Zhu, Z. et al. Greening of the Earth and its drivers. *Nature Climate Change* **6**, 791–795 (2016).

8. Piao, S. et al. Characteristics, drivers and feedbacks of global greening. *Nature Reviews Earth & Environment* **1**, 14–27 (2020).

9. Ruehr, S. et al. Evidence and attribution of the enhanced land carbon sink. *Nature Reviews Earth & Environment* **4**, 518–534 (2023).

10. Baldocchi, D., Chu, H. & Reichstein, M. Inter-annual variability of net and gross ecosystem carbon fluxes: a review. *Agric. For. Meteorol.* **249**, 520–533 (2018).

11. Baldocchi, D. D. How eddy covariance flux measurements have contributed to our understanding of *Global Change Biology*. *Glob. Change Biol.* **26**, 242–260 (2020).

12. Baldocchi, D., Novick, K., Keenan, T. & Torn M. AmeriFlux: its impact on our understanding of the 'breathing of the biosphere', after 25 years. *Agricultural and Forest Meteorolog*y **348**, 109929 (2024).

13. Urbanski, S. et al. Factors controlling CO_2 exchange on timescales from hourly to decadal at Harvard Forest. *J. Geophys. Res.* **112**, G02020 (2007).

14. Dragoni, D. et al. Evidence of increased net ecosystem productivity associated with a longer vegetated season in a deciduous forest in south-central Indiana, USA. *Global Change Biology* **17**, 886–897 (2011).

15. Pilegaard, K., Ibrom, A., Courtney, M. S., Hummelshøj, P. & Jensen, N. O. Increasing net CO_2 uptake by a Danish beech forest during the period from 1996 to 2009. *Agricultural and Forest Meteorology* **151**, 934–946 (2011).

16. Keenan, T. F. et al. Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature* **499**, 324–327 (2013).

17. Froelich, N., Croft, H., Chen, J. M., Gonsamo, A. & Staebler, R. M. Trends of carbon fluxes and climate over a mixed temperate–boreal transition forest in southern Ontario, Canada. *Agricultural and Forest Meteorology* **211–212**, 72–84 (2015).

18. Jiang, Y. et al. Trends and controls on water-use efficiency of an old-growth coniferous forest in the Pacific Northwest. *Environmental Research Letters* **14**, 074029 (2019).

19. Finzi, A. C. et al. Carbon budget of the Harvard Forest Long-Term Ecological Research site: pattern, process, and response to global change. *Ecological Monographs* **90**, e01423 (2020).

20. Hollinger, D. Y. et al. Multi-decadal carbon cycle measurements indicate resistance to external drivers of change at the Howland Forest AmeriFlux site. *J. Geophys. Res. Biogeosci.* **126**, e2021JG006276 (2021).

21. Chen, C., Riley, W. J., Prentice, I. C. & Keenan, T. F. CO_2 fertilization of terrestrial photosynthesis inferred from site to global scales. *Proceedings of the National Academy of Sciences* **119**, e2115627119 (2022).

22. Launiainen, S. et al. Does growing atmospheric CO_2 explain increasing carbon sink in a boreal coniferous forest? *Global Change Biology* **28**, 2910–2929 (2022).

23. Deser, C., Knutti, R., Solomon, S. & Phillips, A. S. Communication of the role of natural variability in future North American climate. *Nature Climate Change* **2**, 775–779 (2012).

24. Deser, C., Phillips, A. S., Alexander, M. A. & Smoliak, B. V. Projecting North American climate over the next 50 years: uncertainty due to internal variability. *Journal of Climate* **27**, 2271–2296 (2014).

25. Deser, C. et al. Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change* **10**, 277–286 (2020).

26. Lehner, F. & Deser, C. Origin, importance, and predictive limits of internal climate variability. *Environmental Research: Climate* **2**, 023001 (2023).

27. Kay, J. E. et al. The Community Earth System Model (CESM) large ensemble project: a community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Met. Soc.* **96**, 1333–1349 (2015).

28. Maher, N., Lehner, F. & Marotzke, J. Quantifying the role of internal variability in the temperature we expect to observe in the coming decades. *Environmental Research Letters* **15**, 054014 (2020).

29. Rodgers, K. B. et al. Ubiquity of human-induced changes in climate variability. *Earth System Dynamics* **12**, 1393–1411 (2021).

30. Martel, J.-L., Mailhot, A., Brissette, F. & Caya, D. Role of natural climate variability in the detection of anthropogenic climate change signal for mean and extreme precipitation at local and regional scales. *Journal of Climate* **31**, 4241–4263 (2018).

31. Dai, A. & Bloecker, C. E. Impacts of internal variability on temperature and precipitation trends in large ensemble simulations by two climate models. *Climate Dynamics* **52**, 289–306 (2019).

32. Deser, C. & Phillips, A. S. A range of outcomes: the combined effects of internal variability and anthropogenic forcing on regional climate trends over Europe. *Nonlin. Processes Geophys.* **30**, 63–84 (2023).

33. Hawkins, E. & Sutton, R. The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteorol. Soc.* **90**, 1095–1107 (2009).

34. Lehner, F. et al. Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics* **11**, 491–508 (2020).

35. Hawkins, E. & Sutton, R. Time of emergence of climate signals. *Geophys. Res. Lett.* **39**, L01702 (2012).

36. Lehner, F., Deser, C. & Terray, L. Toward a new estimate of “time of emergence” of anthropogenic warming: insights from dynamical adjustment and a large initial-condition model ensemble. *J. Clim.* **30**, 7739–7756 (2017).

37. McKinnon, K. A., Poppick, A., Dunn-Sigouin, E. & Deser, C. An "observational large ensemble" to compare observed and modeled temperature trend uncertainty due to internal variability. *Journal of Climate* **30**, 7585–7598 (2017).

38. McKinnon, K. A. & Deser, C. Internal variability and regional climate trends in an observational large ensemble. *Journal of Climate* **31**, 6783–6802 (2018).

39. McKinnon, K. A. & Deser, C. The inherent uncertainty of precipitation variability, trends, and extremes due to internal variability, with implications for Western U.S. water resources. *Journal of Climate* **34**, 9605–9622 (2021).

40. Swart, N. C., Fyfe, J. C., Hawkins, E., Kay, J. E. & Jahn, A. Influence of internal variability on Arctic sea-ice trends. *Nat. Clim. Change* **5**, 86–89 (2015).

41. Kirchmeier-Young, M. C., Zwiers, F. W. & Gillett, N. P. Attribution of extreme events in Arctic sea ice extent. *J. Clim.* **30**, 553–571 (2017).

42. DuVivier, A. K. et al. Going with the floe: tracking CESM Large Ensemble sea ice in the Arctic provides context for ship-based observations. *Cryosphere* **14**, 1259–1271 (2020).

43. Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y. & Diffenbaugh, N. S. The potential for snow to supply human water demand in the present and future. *Environ. Res. Lett.* **10**, 114016 (2015).

44. Wieder, W. R. et al. Pervasive alterations to snow-dominated ecosystem functions under climate change. *Proceedings of the National Academy of Sciences* **119**, e2202393119 (2022).

45. Hu, A. & Deser, C. Uncertainty in future regional sea level rise due to internal climate variability. *Geophys. Res. Lett.* **40**, 2768–2772 (2013).

46. Becker, M., Karpytchev, M. & Hu, A. Increased exposure of coastal cities to sea-level rise due to internal climate variability. *Nat. Clim. Chang.* **13**, 367–374 (2023).

47. Rodgers, K. B., Lin, J. & Frölicher, T. L. Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model. *Biogeosciences* **12**, 3301–3320 (2015).

48. Schlunegger, S. et al. Emergence of anthropogenic signals in the ocean carbon cycle. *Nat. Clim. Change* **9**, 719–725 (2019).

49. Elsworth, G. W., Lovenduski, N. S. & McKinnon, K. A. (2021). Alternate history: a synthetic ensemble of ocean chlorophyll concentrations. *Global Biogeochemical Cycles* **35**, e2020GB006924 (2021).

50. Lombardozzi, D., Bonan, G. B. & Nychka, D. W. The emerging anthropogenic signal in land atmosphere carbon-cycle coupling. *Nat. Clim. Change* **4**, 796–800 (2014).

51. Bonan, G. B., Lombardozzi, D. L. & Wieder, W. R. The signature of internal variability in the terrestrial carbon cycle. *Environmental Research Letters* **16**, 034002 (2021).

52. Munger, J. W. AmeriFlux FLUXNET-1F US-Ha1 Harvard Forest EMS Tower (HFR1), Ver. 3-5, AmeriFlux AMP, (Dataset). https://doi.org/10.17190/AMF/1871137 (2022).

53. Novick, K. & Phillips, R. AmeriFlux FLUXNET-1F US-MMS Morgan Monroe State Forest, Ver. 3-5, AmeriFlux AMP, (Dataset). https://doi.org/10.17190/AMF/1854369 (2022).

54. Thompson, D. W. J., Barnes, E. A., Deser, C., Foust, W. E. & Phillips, A. S. Quantifying the role of internal climate variability in future climate trends. *Journal of Climate* **28**, 6443–6456 (2015).

55. Lawrence, D. M. et al. The Community Land Model version 5: description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems* **11**, 4245–4287 (2019).

56. Collier, N. et al. The International Land Model Benchmarking (ILAMB) system: design, theory, and implementation. *Journal of Advances in Modeling Earth Systems* **10**, 2731–2754 (2018).

57. Bonan G. B. et al. Model structure and climate data uncertainty in historical simulations of the terrestrial carbon cycle (1850–2014). *Global Biogeochemical Cycles* **33**, 1310–1326 (2019).

58. Hardouin, L., et al. Uncertainty in land carbon budget simulated by terrestrial biosphere models: the role of atmospheric forcing. *Environ. Res. Lett.* **17**, 094033 (2022).

59. Jain, S. et al. Importance of internal variability for climate model assessment. *npj Clim. Atmos. Sci.* **6**, 68 (2023).

60. Roman, D. T. et al. The role of isohydric and anisohydric species in determining ecosystem-scale response to severe drought. *Oecologia* **179**, 641–654 (2015).

61. Cox, P. M. et al. Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature* **494**, 341–344 (2013).

62. Piao, S. et al. Interannual variation of terrestrial carbon cycle: issues and perspectives. *Global Change Biology* **26**, 300–318 (2020).

63. Wozniak, M. C., Bonan, G. B., Keppel-Aleks, G. & Steiner, A. L. Influence of vertical heterogeneities in the canopy microenvironment on interannual variability of carbon uptake in temperate deciduous forests. *Journal of Geophysical Research: Biogeosciences* **125**, e2020JG005658 (2020).

64. Wieder, W. R., Butterfield, Z., Lindsay, K., Lombardozzi, D. L. & Keppel-Aleks, G. Interannual and seasonal drivers of carbon cycle variability represented by the Community Earth System Model (CESM2). *Global Biogeochemical Cycles* **35**, e2021GB007034 (2021).

65. Friedlingstein, P. et al. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate* **27**, 511–526 (2014).

66. Arora, V. K. et al. Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences* **17**, 4173–4222 (2020).

67. Hawkins, E., Smith, R. S., Gregory, J. M. & Stainforth, D. A. Irreducible uncertainty in near-term climate projections. *Clim. Dyn.* **46**, 3807–3819 (2016).

68. Fasullo, J. T. et al. Spurious late historical-era warming in CESM2 driven by prescribed biomass burning emissions. *Geophysical Research Letters* **49**, e2021GL097420 (2022).

69. Kim, J.-E. et al. Interannual fires as a source for subarctic summer decadal climate variability mediated by permafrost thawing. *npj Climate and Atmospheric Science*, **6**, 84 (2023).

70. Pastorello, G. et al. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data* **7**, 225 (2020).