

<https://doi.org/10.1038/s41612-024-00878-w>

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

Check for updates

Gordon B. Bonan<sup>1</sup> ✉, Clara Deser<sup>1</sup>, William R. Wieder<sup>1,2</sup>, Danica L. Lombardozzi<sup>1,3</sup> & Flavio Lehner<sup>1,4,5</sup>

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

Climate change is evident in numerous disparate observations of the Earth system. The increase in atmospheric CO<sub>2</sub> measured at Mauna Loa, Hawaii, since 1959, is one of the iconic records of global change<sup>1,2</sup>, as is the planetary warming seen in the time series of surface temperature measurements<sup>3–5</sup> and reflected in Arctic sea-ice loss<sup>6</sup>. Multidecadal changes in the biosphere, both greening and browning, are found in satellite-derived vegetation indices<sup>7–9</sup>. Further evidence for a changing biosphere is obtained from the worldwide network of eddy covariance flux towers, which measure energy, water, and CO<sub>2</sub> exchange between the biosphere and atmosphere<sup>10–12</sup>. Analyses of flux tower measurements find temporal increases in terrestrial productivity at many locations<sup>13–22</sup>. 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<sup>19,20</sup>. Although the trends have been interpreted in terms of changes in climate, CO<sub>2</sub> concentration, and other forcings, the extent to which unforced variability internal to the climate system influences carbon cycle trends is unknown.

The changing state of the Earth system reflects the forced change in response to rising greenhouse gas concentrations and other anthropogenic forcings<sup>23</sup>. Superimposed on the forced response is naturally occurring variation in climate at timescales from several days to decades<sup>24–27</sup>. This

variation, known as internal climate variability, is unforced and arises from chaotic behavior intrinsic to the coupled atmosphere, ocean, sea ice, and land system. Internal climate variability is random behavior overlaid onto the long-term response to anthropogenic forcings. It can be thought of as the climate that is realized in contrast to the climate that is expected from the forcings alone, and it means there can be periods of cooling or minimal warming during a long-term warming trend or periods of accelerated warming that exceed the expected warming<sup>28</sup>.

Internal climate variability is studied using climate simulations over the historical era and projections of future climate change over the 21st century. The sensitivity of weather forecasts to small perturbations in initial conditions is the behavior of nonlinear deterministic systems identified by Edward Lorenz in his foundational study of chaos theory and popularized by a butterfly flapping its wings<sup>29</sup>. Similar behavior occurs in climate simulations, where small perturbations of the initial atmosphere temperature field (e.g., 10<sup>-14</sup> K) produce different climate trajectories<sup>24,26,30</sup>. Large ensembles (typically 30–100 members) performed with a single model quantify the range of outcomes due to internal climate variability<sup>24,26,30–32</sup>. Each ensemble member has identical anthropogenic forcings, but they differ in initial atmosphere and/or ocean states. Each simulation is a unique expression of the random sequence of unforced variability, and each is an equally plausible depiction of climate. The spread among ensemble members is due solely to internal climate variability, but the ensemble mean,

<sup>1</sup>Climate and Global Dynamics Laboratory, NSF National Center for Atmospheric Research, Boulder, CO, USA. <sup>2</sup>Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, CO, USA. <sup>3</sup>Department of Ecosystem Science & Sustainability, Colorado State University, Fort Collins, CO, USA. <sup>4</sup>Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, USA. <sup>5</sup>Polar Bears International, Bozeman, MT, USA. ✉e-mail: [bonan@ucar.edu](mailto:bonan@ucar.edu)

which averages the random variability among ensemble members, shows the forced response.

Internal climate variability produces a large spread among ensemble members in multidecadal temperature and precipitation trends at regional scales<sup>24–27,30,31,33–35</sup>. It is a large source of uncertainty in climate projections at regional spatial scales and decadal timescales<sup>27,31,36,37</sup>. Internal climate variability can mask anthropogenic influences on climate, seen in the concept of time of emergence, which is the time when the forced signal emerges from the noise<sup>38,39</sup>. 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 climate variability<sup>35,40–42</sup>.

Internal climate variability imparts unforced variability in other components of the Earth system including sea ice<sup>43–45</sup>, snowmelt<sup>46,47</sup>, sea level rise<sup>48,49</sup>, and ocean biogeochemistry<sup>50–52</sup>. Less studied is whether internal climate variability manifests in the terrestrial carbon cycle. Previous studies have found a signature of internal climate variability in the terrestrial carbon cycle<sup>53</sup>, which masks the forced signal of change<sup>54</sup> and confounds the interpretation of observed changes in the carbon cycle<sup>55</sup>.

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 the variability in GPP trends. Annual GPP increased over the period 1992–2004 at a rate of 36.3 g C m<sup>-2</sup> yr<sup>-1</sup> per year<sup>13</sup>. A longer time series that extends the observations to 2015 has a smaller annual trend equal to 23.3 g C m<sup>-2</sup> yr<sup>-1</sup> per year over the 24-year period<sup>19</sup>. One interpretation is that the post-2004 data evidence a change in the carbon cycle<sup>11</sup>. An alternative but untested interpretation is that the two trends differ as a result of internal climate 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. 32) to examine how internal climate 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<sup>12</sup>. 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

variability in trends across all 50 ensemble members. We apply this finding to estimate the internal variability of the AmeriFlux GPP data for Harvard Forest over the period 1992–2020 (ref. 56) and calculate the probability of obtaining the different trends reported for 1992–2004 and 1992–2015. We demonstrate that internal climate 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. 57) and Howland Forest (US-Ho1; 45.2041°N, 68.7402°W) for 1996–2020 (ref. 20).

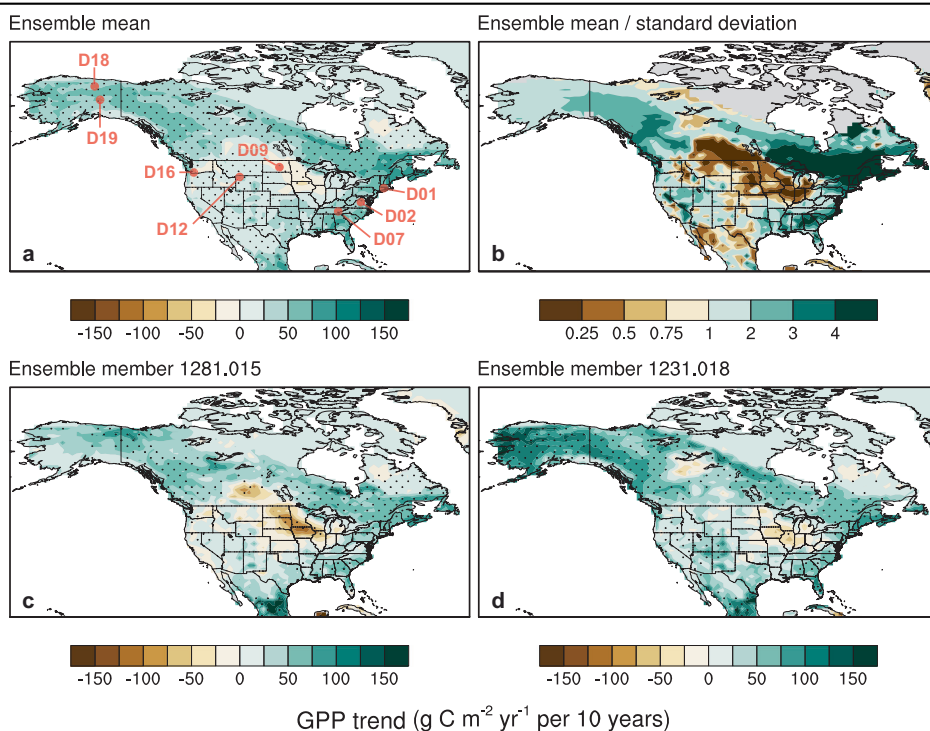
## Results

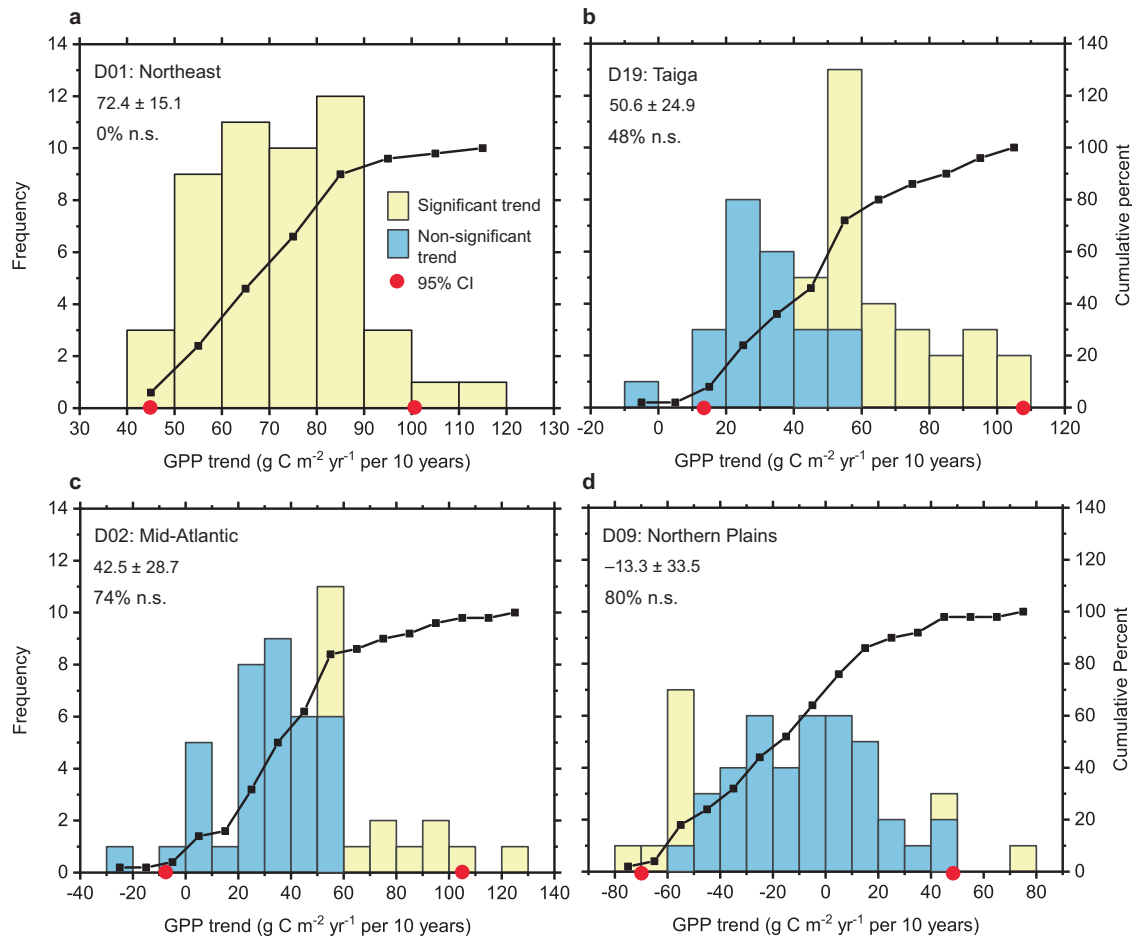
### Simulated GPP trends

Across much of North America, the CESM2 ensemble mean, which is indicative of the model’s forced response to anthropogenic emissions, has a statistically significant increase in annual GPP from 1991 to 2020 (Fig. 1a). However, there is considerable variability among 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 is 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.





**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 ( $n = 30$  members;  $p \leq 0.05$ ), and light blue shading shows non-significant trends. The mean  $\pm$  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.

Histograms of trends across the 50 ensemble members for individual grid cells further illustrate the ensemble variability (Fig. 2). In the Northeast, all of the members have a statistically significant trend, but the 95% confidence interval of the trend (obtained from the ensemble spread) ranges from 45 to 100  $\text{g C m}^{-2} \text{ yr}^{-1}$  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  $\text{g C m}^{-2} \text{ yr}^{-1}$  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  $\text{g C m}^{-2} \text{ yr}^{-1}$  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). The variability in GPP trends is comparable whether the ensemble members differ in start year or differ only in a  $10^{-14}$  K perturbation to the initial air temperature field (Supplementary Fig. 3).

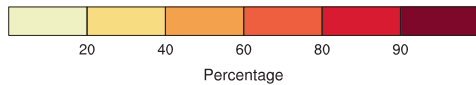
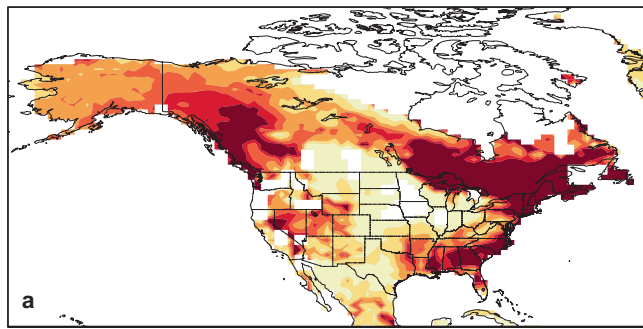
In British Columbia, eastern portions of Canada, the Northeast US, and parts of the Southeast US, more than 90% of the members have a statistically significant positive GPP trend (Fig. 3a). Other regions show large variability among ensemble members. In Alaska, the ensemble mean trend is statistically 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 the Midwest US is statistically significant in only 10–30% of the members (Fig. 3b).

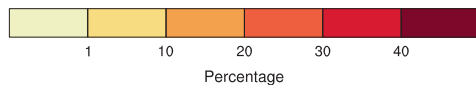
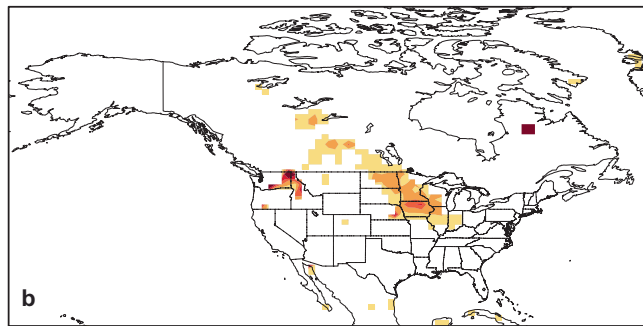
The 95% confidence interval for annual GPP trends obtained from the 50-member ensemble shows a wide range of trends among members (Fig. 4a). The confidence interval spans more than 100  $\text{g C m}^{-2} \text{ yr}^{-1}$  per 10 years across portions of Alaska, northern Canada, the Canadian prairie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

The standard error of the linear regression trend ( $s_{b1}$ , Eq. 3), which quantifies the interannual variability of 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 climate variability has been shown previously for temperature and precipitation<sup>58</sup>, and we find that a similar result pertains to GPP. The 95% confidence interval of the GPP trend obtained using  $s_{b1}$  for a single ensemble member (Fig. 4b) approximates the 95% confidence interval of the trend for the 50-member ensemble (Fig. 4a). This is also evident for other ensemble members (Supplementary Fig. 5). Differences between the 95% confidence interval of the 50-member ensemble and the  $s_{b1}$ -estimated confidence interval are mostly within  $\pm 25 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years (Supplementary Fig. 6). A prominent exception

Significant positive trends

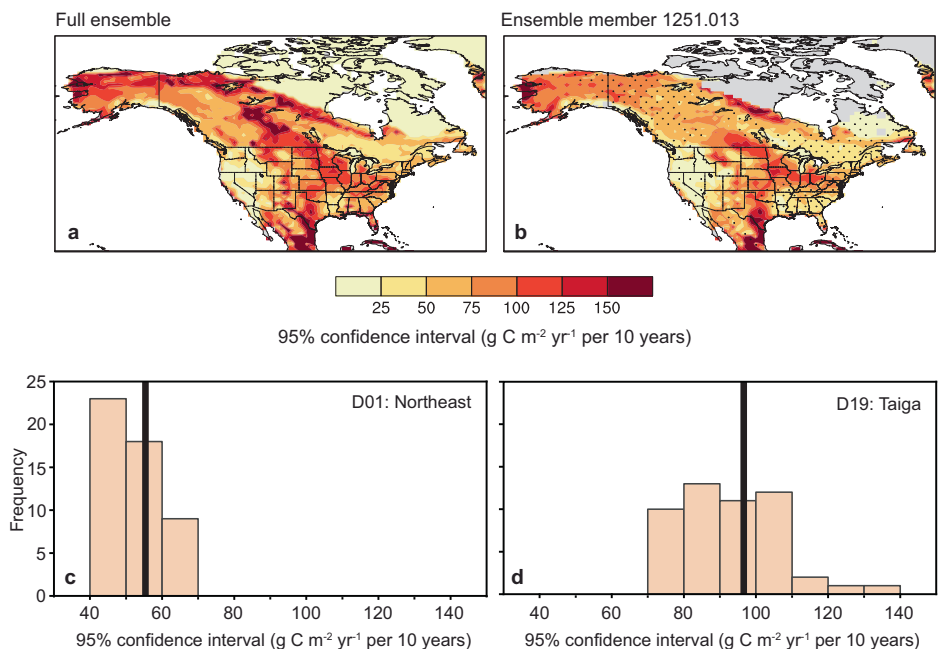


Significant negative trends



**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.

**Fig. 4 | 95% confidence interval in annual GPP trends for 1991–2020.** **a** The 95% confidence interval was obtained 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 ( $s_{b1}$ , Eq. 3). 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;  $p \leq 0.05$ ). **c** Frequency distribution of the 95% confidence interval obtained from  $s_{b1}$  for the  $n = 50$  ensemble members 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 from the 50-member ensemble as in (a). **d** As in (c), but for D19: Taiga.



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, and therefore the estimated 95% confidence interval obtained from the linear regression varies among 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. 7).

**Correlation of GPP with temperature and precipitation**

The CESM2 Large Ensemble 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 climate variability (Supplementary Fig. 8). Annual precipitation increases in some regions of North America in the ensemble mean, but with considerable variability among ensemble members (Supplementary Fig. 9). 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).

**Internal variability of observed GPP 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  $127.1 \pm 41.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  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, as well as the interannual variability (Fig. 6b). The trend across the 50

members is  $73.0 \pm 14.1 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years (mean  $\pm$  standard deviation), with a range of  $48\text{--}114 \text{ g C m}^{-2} \text{ yr}^{-1}$  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,

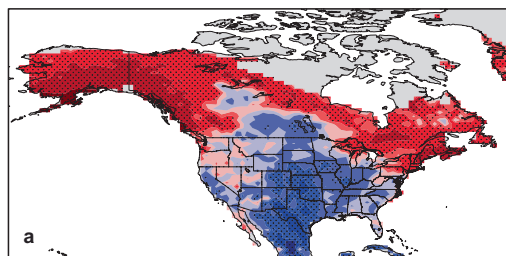
the variability of CESM2 trends ( $14.1 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years) is one-third the observed variability ( $41.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years).

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

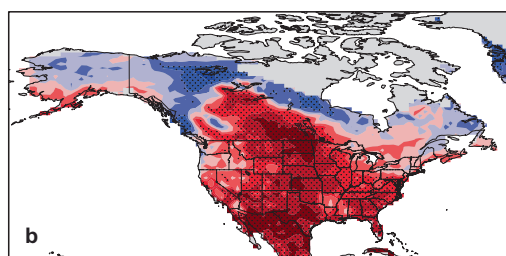
The Harvard Forest data show considerable variability in GPP trends depending on the time period sampled (Fig. 6a). Annual GPP increased over the period 1992–2004 at a rate of  $362.1 \pm 65.0 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years using data reported by Urbanski et al. (ref. 13). A subsequent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

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

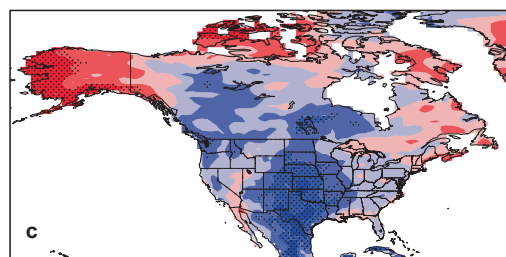
GPP and temperature trends



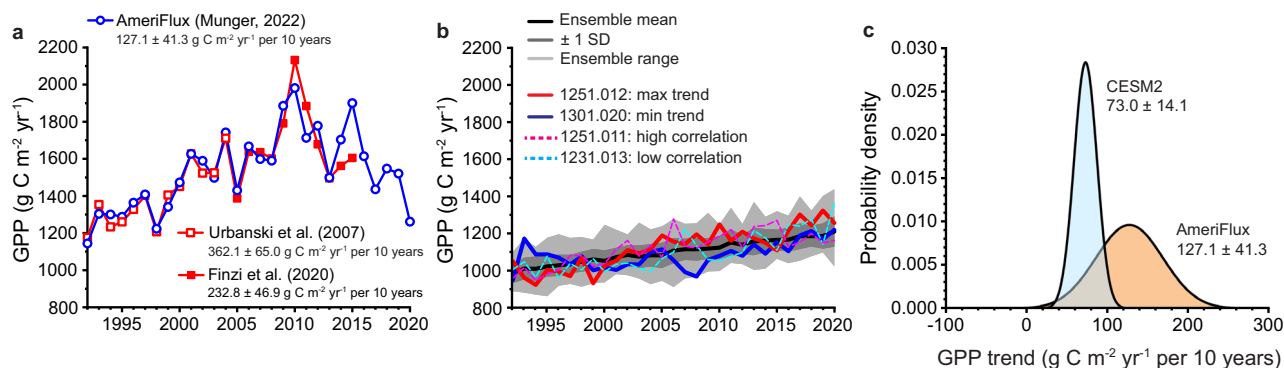
GPP and precipitation trends



Temperature and precipitation trends



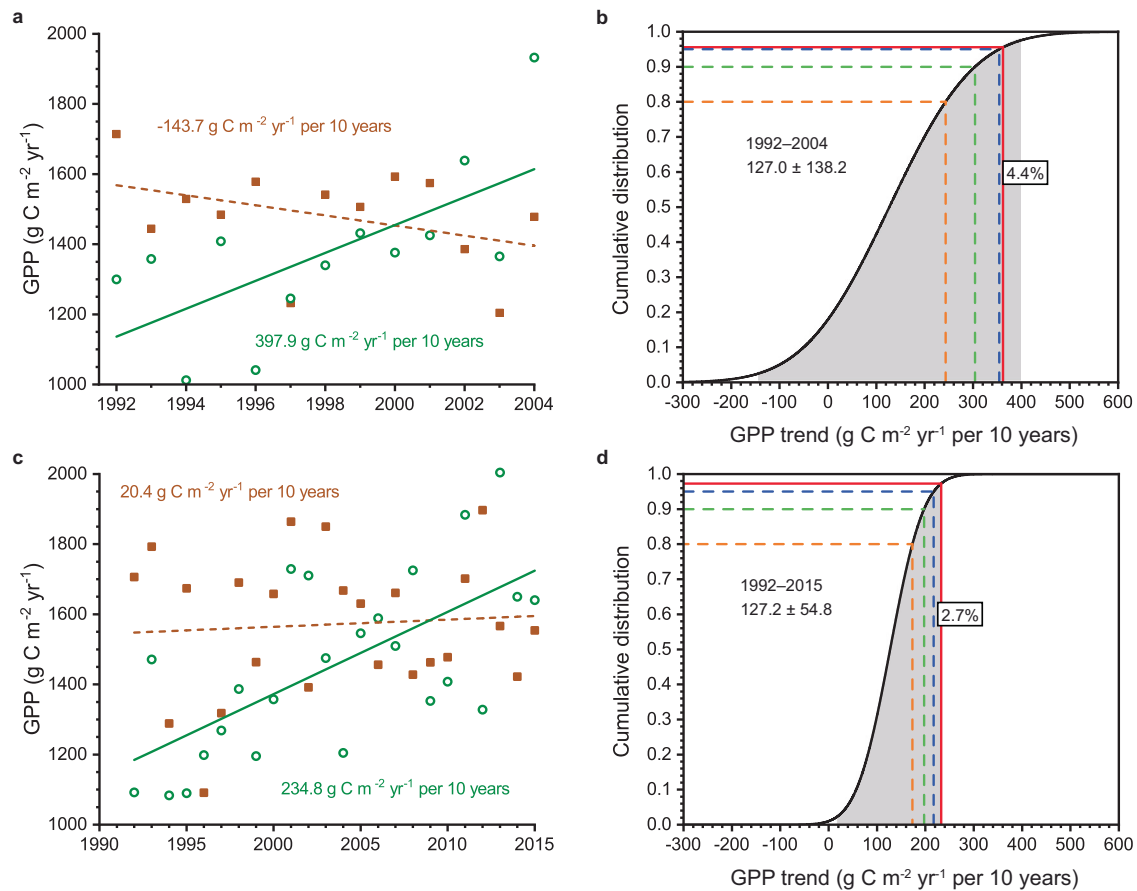
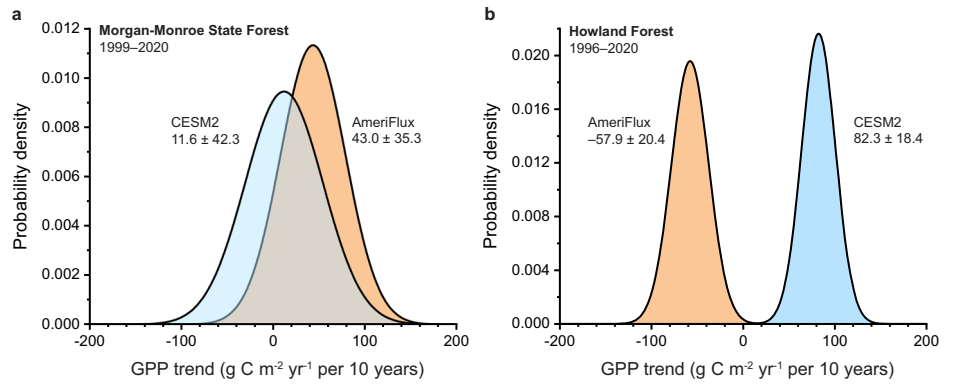
**Fig. 5 | Correlation across the 50 ensemble members of the 30-year trends (1991–2020).** a GPP and surface air temperature, b GPP and precipitation, and c temperature and precipitation. Stippling denotes statistically significant correlations ( $n = 50$ ,  $p \leq 0.05$ ).



**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 Urbanski et al. (ref. 13) for 1992–2004, Finzi et al. (ref. 19) for 1992–2015, and the AmeriFlux data (ref. 56) 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  $\pm$  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  $\pm$  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  $\pm$  standard deviation,  $73.0 \pm 14.1 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years). Also shown is the trend estimated using the AmeriFlux data ( $127.1 \pm 41.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years).

**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  $\pm$  standard deviation,  $11.6 \pm 42.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years). Also shown is the trend estimated using the AmeriFlux data (ref. 57). 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.



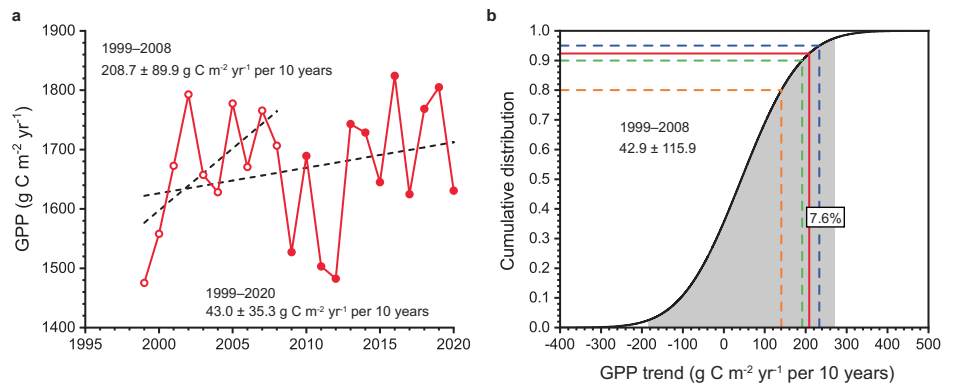
**Fig. 8 | Conditional probability of GPP trends at the AmeriFlux US-Ha1 (Harvard Forest) tower obtained using Monte Carlo methods. a** Annual GPP for 1992–2004 for two Monte Carlo simulations in which GPP for each year is chosen as a random deviate about the 1992–2020 forced trend. The time series are the end-points of the 95% confidence interval for trends in 100,000 Monte Carlo simulations. The brown squares and dashed line show the simulation with a trend at the 2.5th percentile, and the dark green open circles and solid line show the simulation with a trend at the 97.5th percentile. **b** Conditional probability distribution of trends.

Shows the cumulative distribution of trends for 1992–2004 obtained from the 100,000 Monte Carlo simulations. The trends are normally distributed with a mean and standard deviation of  $127.0 \pm 138.2 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years. The gray shading is the 95% confidence interval, and the two time series in panel (a) show the end-points. 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.

trend equal to or greater than the observed trend. The 5% and 10% thresholds are  $217$  and  $197 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years, respectively. Despite a forced trend of  $127.1 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years, interannual variability in GPP can by chance give rise to the larger trends observed for both the 1992–2004 ( $362.1 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years) and the 1992–2015 ( $232.8 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years) time periods.

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. 57) finds that annual GPP increased by  $208.7 \pm 89.9 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years during 1999–2008, decreasing to  $43.0 \pm 35.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years for the full 22-year time series spanning

**Fig. 9 | Annual GPP trends at the AmeriFlux US-MMS (Morgan–Monroe State Forest) tower.** **a** The full 1999–2020 AmeriFlux time series (ref. 57). 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  $\pm$  standard error for the two time periods. See Supplementary Table 2 for the data. **b** Conditional probability distribution of trends obtained using Monte Carlo methods. Shown is the cumulative distribution of trends for 1999–2008 obtained using a forced trend of  $43.0 \pm 35.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years. The trends are normally distributed with a mean and standard deviation of  $42.9 \pm 115.9 \text{ g C m}^{-2} \text{ yr}^{-1}$  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.



1999–2020 (Fig. 9a). Monte Carlo analysis similar to those at Harvard Forest show that a forced trend of  $43.0 \pm 35.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years has a 95% confidence interval of  $-184$  to  $270 \text{ g C m}^{-2} \text{ yr}^{-1}$  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  $43.0 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years. There is a 5% chance of a value greater than  $233 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years and a 10% chance that the trend exceeds  $191 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years.

## Discussion

Our analysis of the 50-member CESM2 Large Ensemble shows that internal climate 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 climate 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  $73 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years and the range across ensemble members is  $48$ – $114 \text{ g C m}^{-2} \text{ yr}^{-1}$  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  $127 \text{ g C m}^{-2} \text{ yr}^{-1}$  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 ( $12 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years) suggests the model is biased low compared with the observations ( $43 \text{ g C m}^{-2} \text{ yr}^{-1}$  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 temporal sequence of internal climate 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<sup>59,60</sup>, but alternative reconstructions of historical meteorology, which can be thought of as samples of observational uncertainty, produce different carbon cycle trends<sup>61,62</sup>. 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<sup>27</sup>.

Internal climate variability 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<sup>13,19</sup>. 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 show that internal climate variability superimposed upon a long-term forced response can produce short-term unforced changes in GPP trends that are consistent with the observations. Care needs to be taken to distinguish forced vs. unforced components of GPP trends, as indeed is evident in analysis of trends in the physical climate system<sup>24,30,63</sup>. The conclusion that forest productivity has increased at Harvard Forest is robust, but the magnitude is uncertain and is influenced by internal climate variability. Our results show that the large GPP trends for 1992–2004 and 1992–2015 (Fig. 6a) can be found by chance despite 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<sup>12,14</sup>, but which was reduced by severe drought in 2012 (ref. 64). Within this long-term trend, internal climate 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<sup>58</sup>, 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 (Figs. 8 and 9). Other more advanced statistical techniques are available to estimate the observational internal variability for temperature and precipitation<sup>40–42</sup>. Similar methods have been used to create an observational ensemble of ocean chlorophyll, for which internal climate variability creates a wide range of possible trends<sup>52</sup>. Whether the same methods can be applied to create an observational ensemble for the terrestrial carbon cycle is unclear.

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<sup>65</sup>, meaning that the importance of internal climate 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 of GPP. 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.

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<sup>26,27,66</sup>. Internal climate variability generates comparable irreducible uncertainty 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<sup>32</sup>. 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<sup>32,67</sup>. 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 members was generated by a random perturbation to the atmosphere temperature field at initialization of order 10<sup>-14</sup> K. The start years for the micro-initializations were chosen to sample different states of the Atlantic Meridional Overturning Circulation. Ensemble members are identified by start year and ensemble number (e.g., 1281.015 is the 15th member of the micro-initializations at start year 1281; note that the BB\_CMIP6\_SM simulations are members 11–20). Rodgers et al. (ref. 32.) provide details of the model configuration, initialization, and forcings.

Ensemble members with the same start year (e.g., the 10-member micro-initialization at year 1231) have the same initial carbon states. Members with different start years have different initial carbon states, but each start year is drawn from the preindustrial control and the differences are small<sup>53</sup>. Memory of initial conditions is minimal by 1991–2020 in that the different initializations in 1850 generate similar ensemble variability of GPP trends (Supplementary Fig. 3).

We analyzed the 30-year period 1991–2020 to discern trends in annual gross primary production (GPP) for each ensemble member. Similar to studies of climate trends<sup>24–27,63</sup>, we estimated the trend as the linear fit to the 1991–2020 time series using ordinary least squares regression:

$$x_i = b_0 + b_1 * t_i \tag{1}$$

where  $x_i$  is annual GPP (g C m<sup>-2</sup> yr<sup>-1</sup>) and  $t_i$  is year (1991, 1992, ..., 2020). The standard deviation of the residuals is:

$$s_e = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \tag{2}$$

where  $\hat{x}_i$  is the predicted GPP from the linear regression. The standard error of the trend ( $b_1$ ) is:

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

The right-most equation for  $s_{b1}$  is the form given by Thompson et al. (ref. 58) when time ( $t_i$ ) is expressed as  $n$  consecutive integers ( $n = 30$ ; 1991, 1992, ..., 2020). Statistical significance was determined by regressions with  $p \leq 0.05$ .

We estimated the 95% confidence interval for the GPP trend ( $b_1$ ) using two methods. (1) We obtained the 95% confidence interval 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. (2) The standard error of the regression trend ( $s_{b1}$ ) obtained for a single ensemble member also estimates ensemble variability, as shown previously for temperature and precipitation<sup>58</sup>. The 95% confidence interval for the trend in this method is  $b_1 \pm 2.048 s_{b1}$ , where  $t_{0.975,28} = 2.048$  is the critical t-value for  $n = 30$  years of data.

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<sup>53</sup>.

### 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<sup>56</sup>, which was processed using the ONEFlux processing codes<sup>68</sup> 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 as in Eq. (1). The fitted regression for the  $n = 29$ -year time series is:  $b_0 = -23954.19$  g C m<sup>-2</sup> yr<sup>-1</sup>,  $b_1 = 12.71$  g C m<sup>-2</sup> yr<sup>-2</sup>,  $F = 9.44$ ,  $p = 0.0048$ , and  $R^2 = 0.259$ . The standard deviation of the residuals is:  $s_e = 186.3$  g C m<sup>-2</sup> yr<sup>-1</sup>.

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:

$$x'_i = \hat{x}_i + \varepsilon_i * s_e \tag{4}$$

where  $\hat{x}_i$  is the predicted GPP for year  $i$  using the linear regression in Eq. (1),  $\varepsilon_i$  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<sup>-2</sup> yr<sup>-1</sup>). 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 of the GPP trend. The distribution of  $b'_1$ , obtained from the Monte Carlo simulations with the assumption of random interannual variability (Supplementary Fig. 11b), has a mean ( $127.0 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years) and standard deviation ( $41.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years) comparable to  $b_1$  and its standard error obtained from the observed GPP time series (Fig. 6a;  $127.1 \pm 41.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years).

We used the statistical model to estimate the conditional probability of obtaining a trend of  $362.1 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years for the time period 1992–2004 and  $232.8 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years for 1992–2015 (Fig. 6a). In this analysis, we used Eq. (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 the two time periods given the long-term trend of  $127.1 \pm 41.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  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 Eq. (3) with  $n = 13$  and  $n = 24$  years.

We performed the same analysis at the AmeriFlux US-MMS Morgan–Monroe State Forest tower ( $39.3232^\circ\text{N}$ ,  $86.4131^\circ\text{W}$ ) for the 22-year period 1999–2020 using the AmeriFlux FLUXNET data product (Supplementary Table 2) (ref. 57). 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;  $b_0 = -6976.29 \text{ g C m}^{-2} \text{ yr}^{-1}$ ,  $b_1 = 4.30 \text{ g C m}^{-2} \text{ yr}^{-2}$ ,  $F = 1.48$ ,  $p = 0.237$ ,  $R^2 = 0.069$ ,  $s_e = 105.1 \text{ g C m}^{-2} \text{ yr}^{-1}$ ) 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  $208.7 \text{ g C m}^{-2} \text{ yr}^{-1}$  per 10 years can be found for the period 1999–2008 given the long-term trend of  $43.0 \pm 35.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  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 Eq. (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 (Figs. 6c, 7a). We supplemented this model–observation comparison with GPP data for the AmeriFlux US-Ho1 Howland Forest tower ( $45.2041^\circ\text{N}$ ,  $68.7402^\circ\text{W}$ ) 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 and is available upon request from corresponding author G.B.

Received: 31 May 2024; Accepted: 7 December 2024;

Published online: 21 December 2024

## References

- Keeling, C. D. et al. Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii. *Tellus* **28**, 538–551 (1976).
- Friedlingstein, P. et al. Global carbon budget 2023. *Earth Syst. Sci. Data* **15**, 5301–5369 (2023).
- Hansen, J., Ruedy, R., Sato, M. & Lo, K. Global surface temperature change. *Rev. Geophys.* **48**, RG4004 (2010).
- Rohde, R. A. & Hausfather, Z. The Berkeley Earth land/ocean temperature record. *Earth Syst. Sci. Data* **12**, 3469–3479 (2020).
- Morice, C. P. et al. An updated assessment of near-surface temperature change from 1850: the HadCRUT5 data set. *J. Geophys. Res.* **126**, e2019JD032361 (2021).
- Box, J. E. et al. Key indicators of Arctic climate change: 1971–2017. *Environ. Res. Lett.* **14**, 045010 (2019).
- Zhu, Z. et al. Greening of the Earth and its drivers. *Nat. Clim. Change* **6**, 791–795 (2016).
- Piao, S. et al. Characteristics, drivers and feedbacks of global greening. *Nat. Rev. Earth Environ.* **1**, 14–27 (2020).
- Ruehr, S. et al. Evidence and attribution of the enhanced land carbon sink. *Nat. Rev. Earth Environ.* **4**, 518–534 (2023).
- Baldocchi, D., Chu, H. & Reichstein, M. Inter-annual variability of net and gross ecosystem carbon fluxes: a review. *Agric. Meteorol.* **249**, 520–533 (2018).
- Baldocchi, D. D. How eddy covariance flux measurements have contributed to our understanding of *Global Change Biology*. *Glob. Change Biol.* **26**, 242–260 (2020).
- Baldocchi, D., Novick, K., Keenan, T. & Torn, M. AmeriFlux: its impact on our understanding of the ‘breathing of the biosphere’, after 25 years. *Agric. For. Meteorol.* **348**, 109929 (2024).
- Urbanski, S. et al. Factors controlling CO<sub>2</sub> exchange on timescales from hourly to decadal at Harvard Forest. *J. Geophys. Res.* **112**, G02020 (2007).
- 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. *Glob. Change Biol.* **17**, 886–897 (2011).
- Pilegaard, K., Ibrom, A., Courtney, M. S., Hummelshøj, P. & Jensen, N. O. Increasing net CO<sub>2</sub> uptake by a Danish beech forest during the period from 1996 to 2009. *Agric. For. Meteorol.* **151**, 934–946 (2011).
- Keenan, T. F. et al. Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature* **499**, 324–327 (2013).
- 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. *Agric. For. Meteorol.* **211–212**, 72–84 (2015).
- Jiang, Y. et al. Trends and controls on water-use efficiency of an old-growth coniferous forest in the Pacific Northwest. *Environ. Res. Lett.* **14**, 074029 (2019).
- Finzi, A. C. et al. Carbon budget of the Harvard Forest Long-Term Ecological Research site: pattern, process, and response to global change. *Ecol. Monogr.* **90**, e01423 (2020).
- 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).
- Chen, C., Riley, W. J., Prentice, I. C. & Keenan, T. F. CO<sub>2</sub> fertilization of terrestrial photosynthesis inferred from site to global scales. *Proc. Natl Acad. Sci. USA* **119**, e2115627119 (2022).
- Launiainen, S. et al. Does growing atmospheric CO<sub>2</sub> explain increasing carbon sink in a boreal coniferous forest? *Glob. Change Biol.* **28**, 2910–2929 (2022).
- IPCC [Intergovernmental Panel on Climate Change]. Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the Intergovernmental Panel on Climate Change (Cambridge University Press, 2021).

24. Deser, C., Knutti, R., Solomon, S. & Phillips, A. S. Communication of the role of natural variability in future North American climate. *Nat. Clim. Change* **2**, 775–779 (2012).
25. 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. *J. Clim.* **27**, 2271–2296 (2014).
26. Deser, C. et al. Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change* **10**, 277–286 (2020).
27. Lehner, F. & Deser, C. Origin, importance, and predictive limits of internal climate variability. *Environ. Res.* **2**, 023001 (2023).
28. Deser, C. Certain uncertainty: the role of internal climate variability in projections of regional climate change and risk management. *Earth's Fut.* **8**, e2020EF001854 (2020).
29. Lorenz, E. N. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963).
30. 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. Am. Met. Soc.* **96**, 1333–1349 (2015).
31. Maher, N., Lehner, F. & Marotzke, J. Quantifying the role of internal variability in the temperature we expect to observe in the coming decades. *Environ. Res. Lett.* **15**, 054014 (2020).
32. Rodgers, K. B. et al. Ubiquity of human-induced changes in climate variability. *Earth Syst. Dyn.* **12**, 1393–1411 (2021).
33. 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. *J. Clim.* **31**, 4241–4263 (2018).
34. Dai, A. & Bloecker, C. E. Impacts of internal variability on temperature and precipitation trends in large ensemble simulations by two climate models. *Clim. Dyn.* **52**, 289–306 (2019).
35. 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. Process. Geophys.* **30**, 63–84 (2023).
36. Hawkins, E. & Sutton, R. The potential to narrow uncertainty in regional climate predictions. *Bull. Am. Meteorol. Soc.* **90**, 1095–1107 (2009).
37. Lehner, F. et al. Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth Syst. Dyn.* **11**, 491–508 (2020).
38. Hawkins, E. & Sutton, R. Time of emergence of climate signals. *Geophys. Res. Lett.* **39**, L01702 (2012).
39. 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).
40. 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. *J. Clim.* **30**, 7585–7598 (2017).
41. McKinnon, K. A. & Deser, C. Internal variability and regional climate trends in an observational large ensemble. *J. Clim.* **31**, 6783–6802 (2018).
42. 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. *J. Clim.* **34**, 9605–9622 (2021).
43. 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).
44. 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).
45. 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).
46. 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).
47. Wieder, W. R. et al. Pervasive alterations to snow-dominated ecosystem functions under climate change. *Proc. Natl Acad. Sci. USA* **119**, e2202393119 (2022).
48. Hu, A. & Deser, C. Uncertainty in future regional sea level rise due to internal climate variability. *Geophys. Res. Lett.* **40**, 2768–2772 (2013).
49. 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).
50. 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).
51. Schlunegger, S. et al. Emergence of anthropogenic signals in the ocean carbon cycle. *Nat. Clim. Change* **9**, 719–725 (2019).
52. Elsworth, G. W., Lovenduski, N. S. & McKinnon, K. A. Alternate history: a synthetic ensemble of ocean chlorophyll concentrations. *Glob. Biogeochem. Cycles* **35**, e2020GB006924 (2021). (2021).
53. Bonan, G. B., Lombardozzi, D. L. & Wieder, W. R. The signature of internal variability in the terrestrial carbon cycle. *Environ. Res. Lett.* **16**, 034002 (2021).
54. 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).
55. Li, N. et al. Enhanced global carbon cycle sensitivity to tropical temperature linked to internal climate variability. *Sci. Adv.* **10**, ead16155 (2024).
56. 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/18711137> (2022).
57. 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).
58. 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. *J. Clim.* **28**, 6443–6456 (2015).
59. Lawrence, D. M. et al. The Community Land Model version 5: description of new features, benchmarking, and impact of forcing uncertainty. *J. Adv. Modeling Earth Syst.* **11**, 4245–4287 (2019).
60. Collier, N. et al. The International Land Model Benchmarking (ILAMB) system: design, theory, and implementation. *J. Adv. Model. Earth Syst.* **10**, 2731–2754 (2018).
61. Bonan, G. B. et al. Model structure and climate data uncertainty in historical simulations of the terrestrial carbon cycle (1850–2014). *Glob. Biogeochem. Cycles* **33**, 1310–1326 (2019).
62. 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).
63. Jain, S. et al. Importance of internal variability for climate model assessment. *npj Clim. Atmos. Sci.* **6**, 68 (2023).
64. 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).
65. 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). *Glob. Biogeochem. Cycles* **35**, e2021GB007034 (2021).
66. Hawkins, E., Smith, R. S., Gregory, J. M. & Stainforth, D. A. Irreducible uncertainty in near-term climate projections. *Clim. Dyn.* **46**, 3807–3819 (2016).

67. Fasullo, J. T. et al. Spurious late historical-era warming in CESM2 driven by prescribed biomass burning emissions. *Geophys. Res. Lett.* **49**, e2021GL097420 (2022).
68. Pastorello, G. et al. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Sci. Data* **7**, 225 (2020).

### Acknowledgements

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, <https://doi.org/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** The online version contains supplementary material available at <https://doi.org/10.1038/s41612-024-00878-w>.

**Correspondence** and requests for materials should be addressed to Gordon B. Bonan.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024