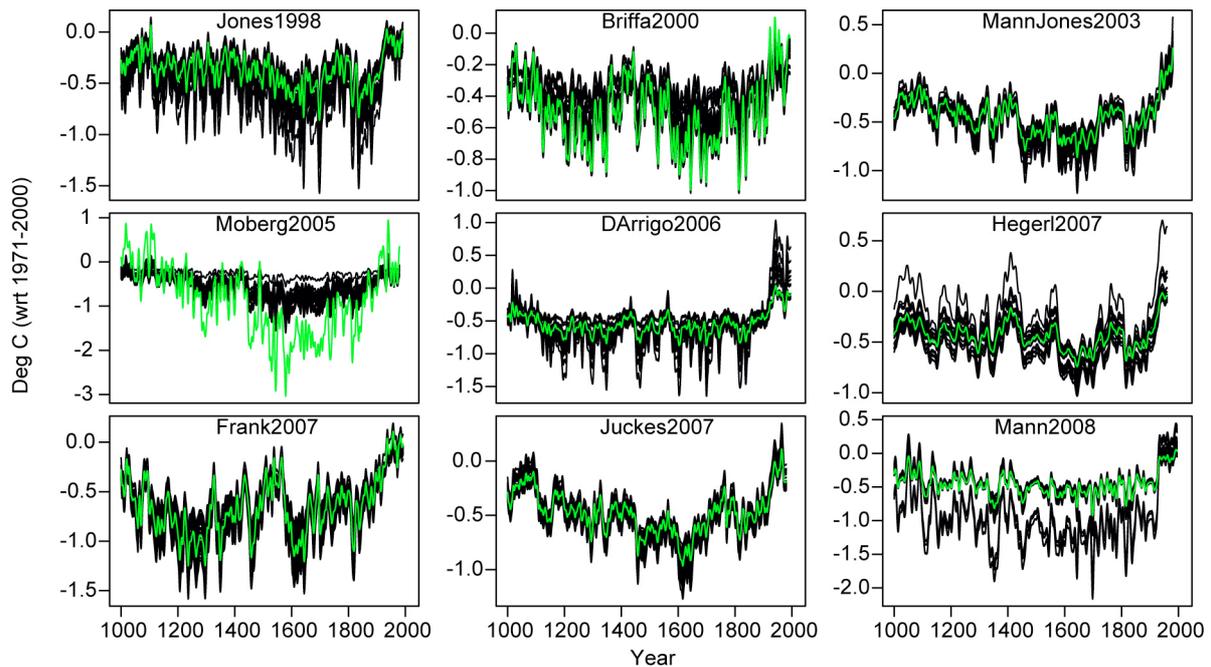**Figure S1**. *The calibration of large-scale temperature reconstructions is fundamental to translate proxy measures into degree Celsius temperature change. The ensemble calibration approach utilized here empirically considers uncertainties related to the calibration period as shown by the black curves for each of the nine large-scale temperature reconstructions (see methods). Potential biases in both the proxy and instrumental data make it difficult to prefer one calibration period to another. For example, in the case of the instrumental target data, the increasing error back in time, and potential biases in early land[1] and mid 20th century sea-surface[2] temperatures, complicate potential weighting schemes to reflect which calibration periods and lengths are most appropriate. In addition to the influence of the calibration period and the spatial domain and seasonality of the instrumental target (see figure S2), the actual method used for calibration has been subject to considerable debate. A comprehensive analysis[2] suggested that "when compared at decadal or lower resolution, we find that most methods can provide satisfactory and similar results" and also that most of the regression methods applied to smoothed data (e.g., the variance matching applied herein, as well as forward, total-least squares, and inverse regressions; see equations 1-3 in ref. 2, respectively) "provide almost identical results". We illustrate inverse regression for the common 1850-1960 calibration period in this figure, by the green curves. These tests suggest that the calibration applied serves as a consistent framework to represent a large fraction of the uncertainty that can be attributed to accepted calibration methodologies and periods. For all but one reconstruction (Moberg2005), the inverse regression results are within the range of the scaling calibration ensembles. Tests of smoothed inverse regression with the up to ten decadal intervals for each reconstruction revealed only the Briffa2000 and Moberg2005 values extended outside the ensemble scaling uncertainties towards larger reconstructed amplitudes. However, inverse regression with shorter calibration windows become erratic with, for example, the majority of reconstructions showing physically implausible changes of sign of regression coefficients. Improvements to the instrumental measurements[3], proxy reconstructions[4,5], and calibration techniques[6] will hopefully reduce calibration uncertainties, and may result in systematic changes to the regression coefficients, the large-scale temperature amplitude, and estimates of the carbon cycle climate sensitivity.*
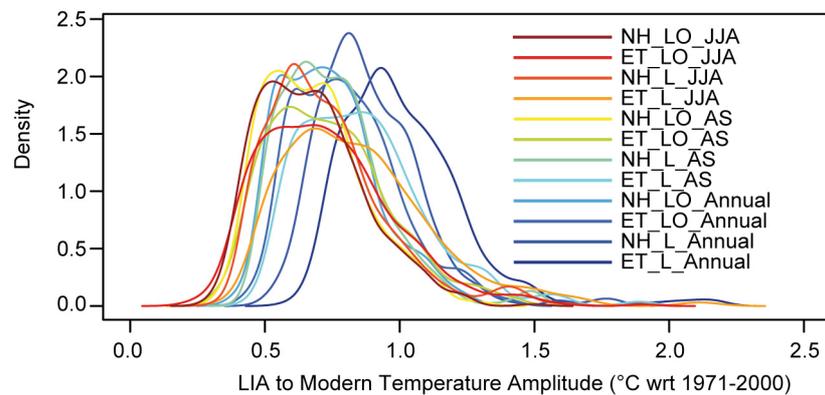
Figure S2. *Large-scale temperature reconstructions have been historically calibrated to a variety of instrumental targets with a variety of different methods. Researcher specific choices include the seasonal window (e.g., June-August (JJA), April-September (AS), Annual (Ann)), for land only (L) or land and ocean temperatures (LO), the calibration time period, latitudinal bands (e.g., the full Northern Hemisphere (NH) or NH extra-tropics (ET), and using a variety of different methods. This multitude of decisions has contributed to the spread of amplitudes. In the present contribution we consider the uncertainty inherent to these choices by employing ensemble calibration techniques, whereby proxy records are calibrated over a variety of periods to the same instrumental target (i.e., NH Annual LO), using a method shown to well preserve the absolute amplitude. In this figure, we address the large-scale amplitude that may be derived should a different instrumental target be chosen. Median amplitudes range from about 1.0 to 0.7 degrees C, with the smallest amplitude for the Annual NH temperature averaged over the land and sea-surfaces. The Little Ice Age (LIA) to modern temperature amplitude is quantified as the temperature difference between the coldest (1601-1630) and warmest (1971-2000) 30-year climatological periods.*
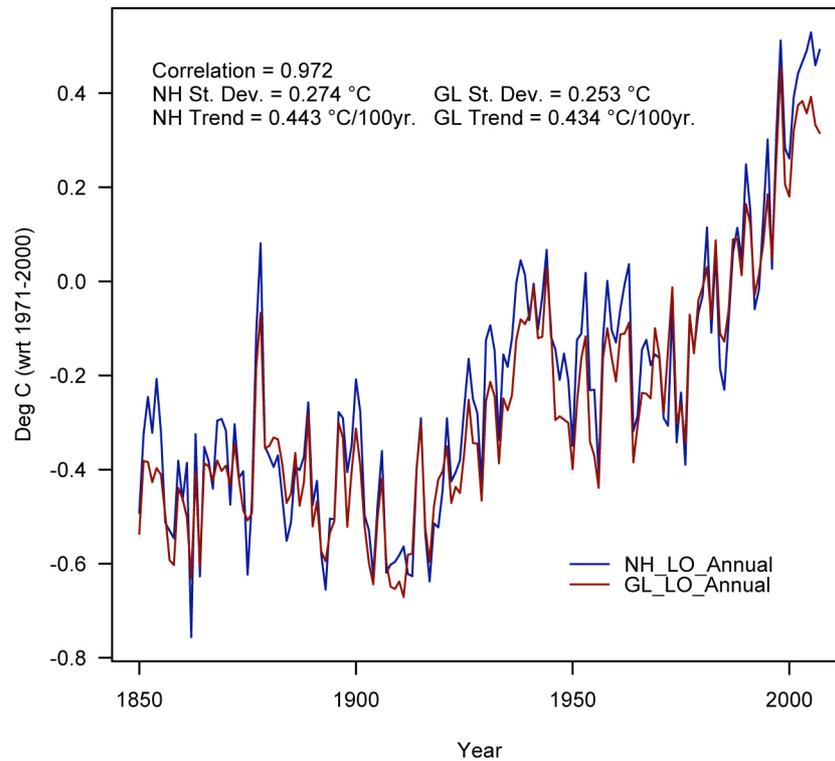
Figure S3. *Large-scale temperature reconstructions are dominated by tree-ring data from the Northern Hemisphere, whereas the CO_2 variations recorded in the Antarctic ice cores represent a more global mixture of land and oceanic processes. The temperature reconstructions were calibrated to Northern Hemisphere Land and Ocean temperatures[7], and in this figure, we show that this instrumental target used for defining γ is highly representative for global conditions. Nevertheless, the carbon cycle-climate sensitivity is quantified in terms of NH temperature change (also for the climate models).*
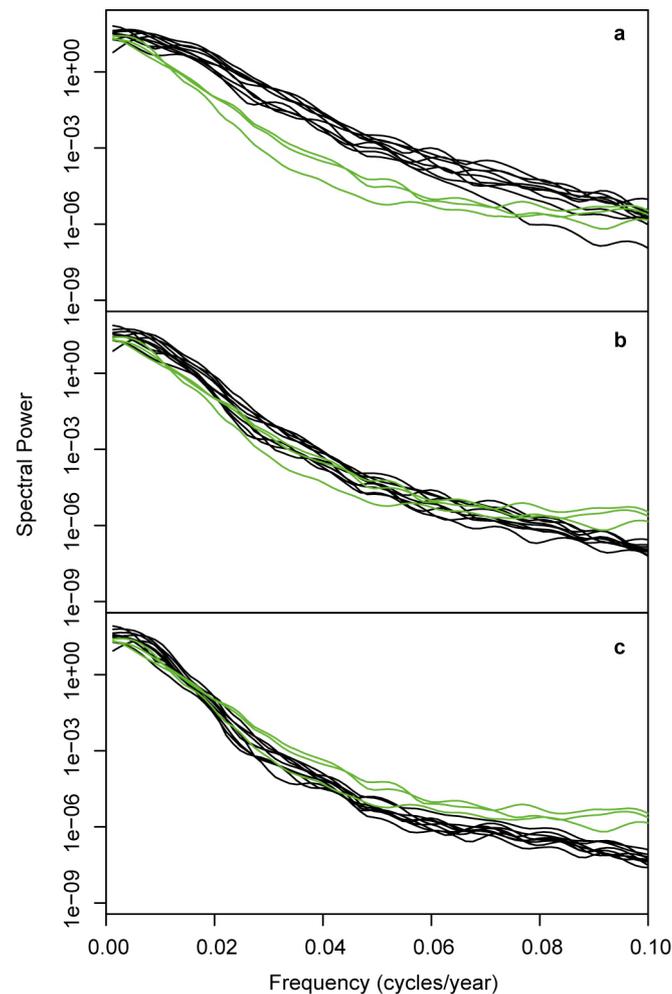
Figure S4. *Variation in the temporal resolution and precision of the ice-core records and temperature reconstructions complicates direct assessment of carbon-cycle feedbacks across timescales. While the temperature reconstructions are generally annually resolved, CO_2 variations are smoothed over decades due to the enclosure processes of gases into ice and the ice archive itself is prone to some dating uncertainties. In this figure, the frequency characteristics of the proxy archives are shown for various smoothings. The CO_2 smoothing is kept constant with a 50-year spline for three records (green lines), while the smoothing for the nine temperature reconstructions is in **a,** 50-years, **b,** 80-years, **c,** 110 years (black lines). Strongest agreement of the spectral loadings is found with 80-year smoothing for the temperature reconstructions. See also figure S5.*
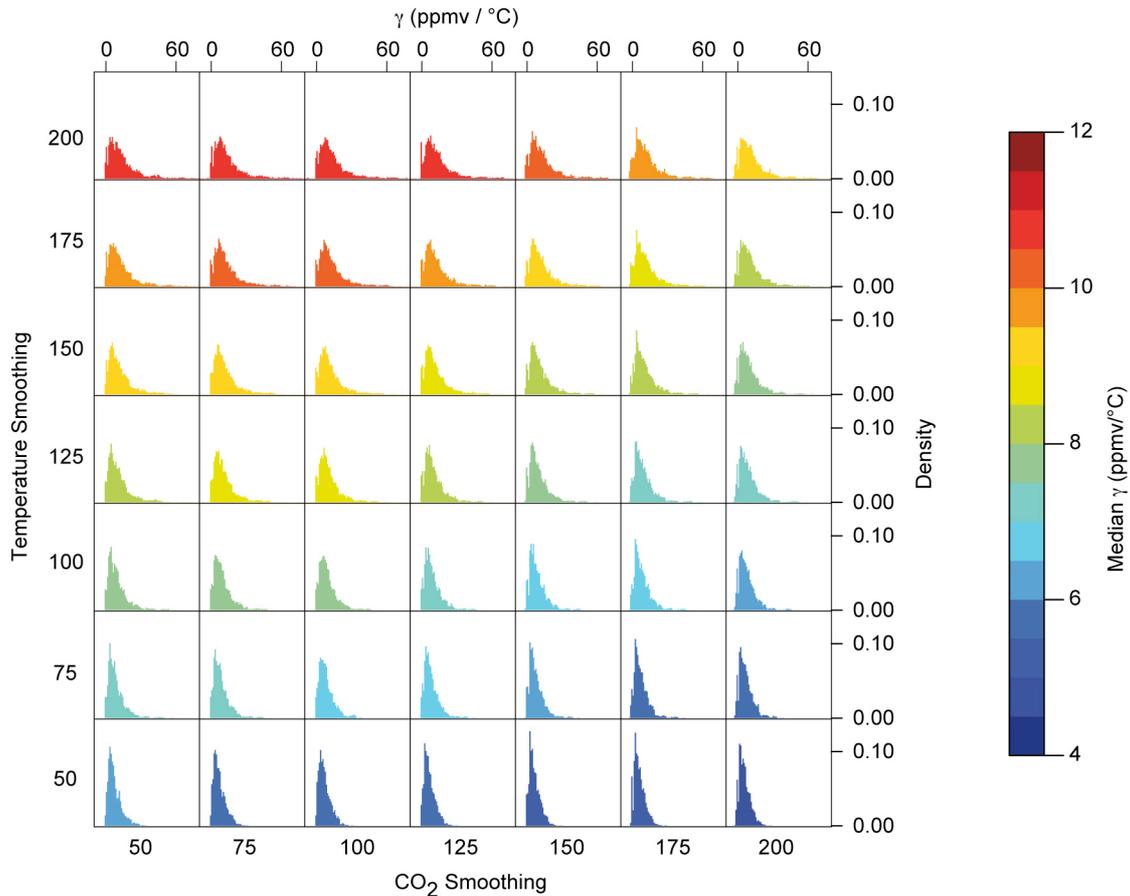
Figure S5. *The degree to which the temperature and CO_2 data are smoothed, differentially influence estimates of γ (see also figure 4). To provide details on these relationships, we plot histograms of γ, conditioned on temperature and CO_2 smoothing (50, 75, ..., 200-year splines). Highest median values of γ are found with substantial temperature smoothing and minimal CO_2 smoothing – towards the upper left corner of the figure. And conversely lowest values are found with highly smoothed CO_2 data and minimally smoothed temperature data. The reduced sensitivity to the CO_2 smoothing possibly reflects the already smoothed nature of the ice archive. This might also be the cause for roughly constant median sensitivities of γ to be found for temperature: CO_2 smoothing ratios smaller than 1:1 (i.e., to have a shallower slope than unity across the figure matrix). See also figure S4.*
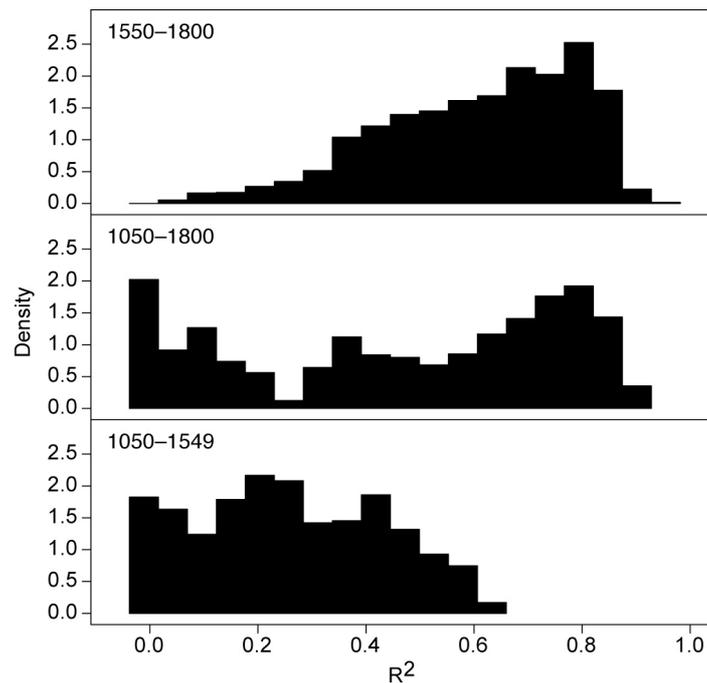
Figure S6. *The high correspondence between the Little Ice Age CO_2 dip and temperature has been long recognized[8]. In this figure we show the distributions of all correlations between the temperature and CO_2 data (after lagging) for **a**,1550-1800, **b**, 1050-1800, and **c**, 1050-1549. Both the high correlations in the late period containing the Little Ice Age dip and the low correlations between temperature and CO_2 during the early period are evident. The degree to which different factors are responsible for increased or decreased correspondence between temperature and CO_2 data throughout time remains unclear. Decreased temperature reconstruction quality, long-time scale carbon feedback processes, anthropogenic influences, covariation of temperature and precipitation, and complex climate-ecological processes (e.g., fire) may all contribute variation in the observed coupling within the carbon cycle-climate system.*
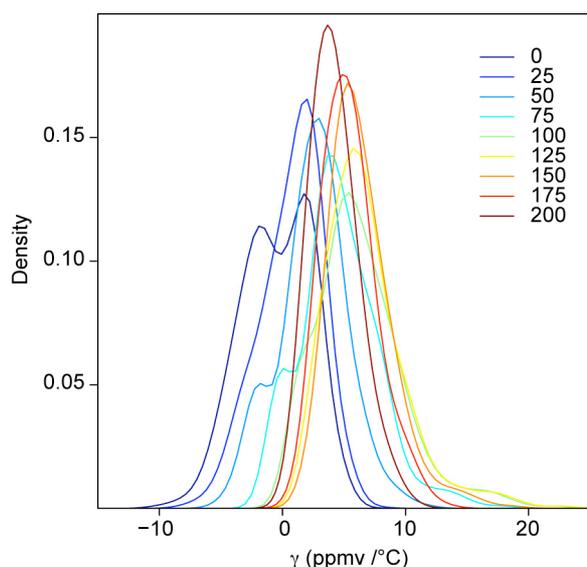
Figure S7. *For 100 year-spline smoothed reconstructed CO_2 and temperature, we systematically lagged the CO_2 data by 0, 25, 50, …, 200 years and performed regressions between the CO_2 and temperature records using data between 1050 - 1549. Coloured lines show density plots of γ for the various lags. These tests were performed due to the generally low agreement between and among temperature and CO_2 records during the first half of the past millennium. The temperature and CO_2 correlations during the 1550-1800 period (0.44 and 0.54 at 50-year lag) are higher than those during the early 1050-1549 period (0.10 and 0.27 at 50-year lag). The lower common signal in the early versus late periods is also observed among the three CO_2 records (0.72 vs 0.86) and the nine temperature reconstructions (0.50 vs 0.73). This reduced agreement during the early period provides few degrees of freedom to assess lags in the CO_2 -temperature relationships. Median responses for a particular lag range from 0.0 to 6.2 ppmv/°C. The total range of all values is -10.4 to 23.2 ppmv/°C. While these results support the lower early period γ, it should be noted that the highest sensitivities were found with a 125 year lag. High quality and resolution temperature and CO_2 reconstructions for the past two millennia might allow a more robust assessment of the (multi)centennial-scale evolution of the coupled climate- carbon cycle.*
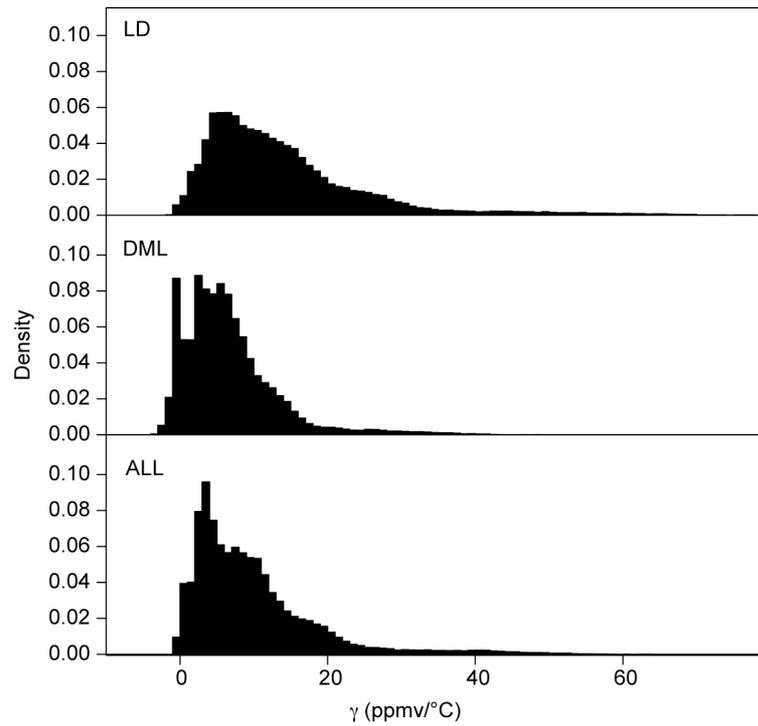
Figure S8. *The individual amplitudes and shapes of CO$_2$ time-series derived from the various ice core records result in differences in the feedback strength. Distributions of γ are shown for the Law Dome (LD), Dronning Maud Landing (DML), and ALL CO$_2$ data, where the latter also includes data from the South Pole. The LD record, with its greater amplitude – and also the prominent LIA dip – yields estimates for the strongest carbon cycle feedbacks.*
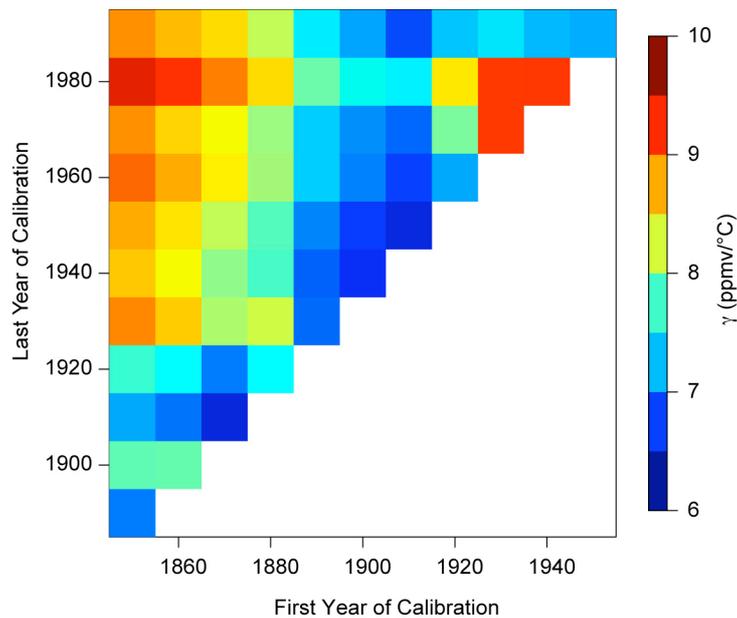
Figure S9. *The period used for calibrating large-scale temperature reconstructions is known to affect the temperature amplitude due to the varying and time dependent proxy and target fit. These calibration windows thus consequently affect estimates of the carbon cycle feedback strength. Here we show the median feedback as a function of the starting and ending decade of calibration. Median γ's between 6 and 10 ppmv/°C are always obtained, however patterns related to the calibration window are evident, with for example lowest feedbacks obtained for periods ending before 1920 and also those starting between 1890 and 1910. The dependence upon particular time windows and limitations to select an optimal calibration interval (see figure S1) suggests utility in ensemble calibration methods and risks associated with application of a single conventional (e.g., 1901-2000) window.*
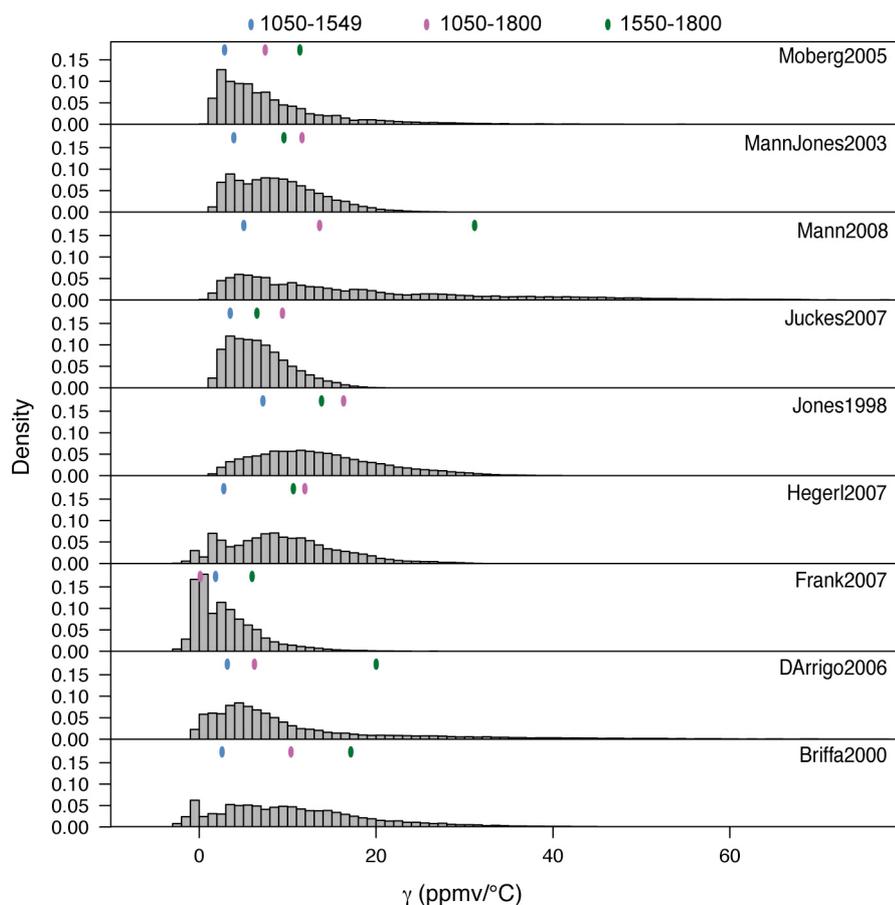
Figure S10. *Despite the ensemble calibration methods applied and the consistent instrumental target, the choice of a particular large-scale temperature reconstruction over another yields a wide variety in estimates of γ. In general all distributions have a very long upper tail, with largest γ obtained for Mann2008 and smallest γ for Frank2007. Some of the reconstructions tend to yield more closely spaced median estimates (colored dots) for the different time periods (e.g., MannJones2003) whereas others tend to display high sensitivity to the chosen time period (e.g., Mann2008). The bimodal distribution for the Frank2007 record is a result of the generally weaker correlation between the CO_2 and temperature over the full 1050-1800 period – a characteristic which results from the early "exit" from warm Medieval conditions in comparison to the CO_2 peak ~ 1200. Similar features are seen in Briffa2000 and Hegerl2007, for example.*
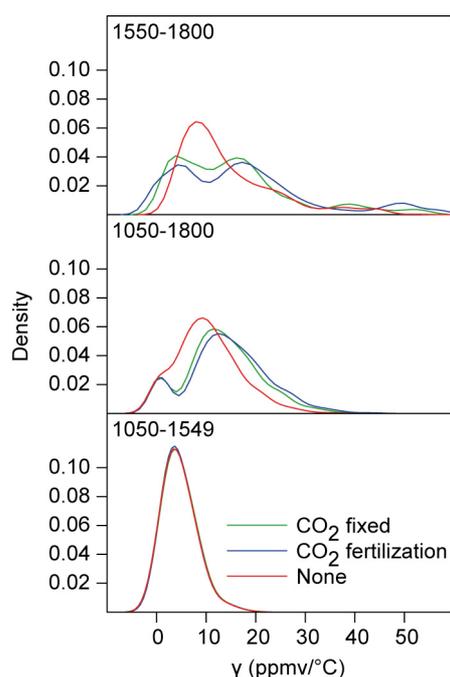
Figure S11. *Pre-industrial era land-use changes, by resulting in net CO_2 uptake or release on decadal to millennial scales, may influence estimates of the carbon cycle sensitivity to climate. As a preliminary investigation of the sensitivity of γ to changes in land-use, we use results of two simulations with the Bern Carbon Cycle-Climate model[9] in which the spatial and temporal evolution of the area of pasture, cropland, and built-up land is prescribed[10] over the Holocene[11]. Model output consists of simulations including and excluding fertilization of the terrestrial biosphere in response to varying atmospheric CO2 as driven by changes in land-use[11]. Ref. 11 provides the first such simulations of the impacts of land-use change on atmospheric CO_2 concentration for the entire Holocene. We here utilize two simulations of atmospheric CO_2 concentration changes including and excluding modeled fertilization effects of the terrestrial biosphere due to land-use changes. Model output consists of smoothed data of atmospheric CO_2 concentration with estimates available every five years. We here linearly interpolate between data points and difference these records with the 100-year smoothed reconstruction of CO_2 concentrations from ice-cores. Regression coefficients between the "land-use corrected" CO_2 data and the 100-year smoothed temperature reconstructions were performed for the late (1550-1800; upper panel), full (1050-1800; middle panel), and early (1050-1549; lower panel) periods as in the main text. In this figure, we show density plots for no ("None", i.e. the original reconstructed temperature and CO_2) land-use changes and those based upon plant response at fixed CO_2 levels ("CO_2 fixed"), and where plant response is allowed to vary as a function of land-use driven changes in CO_2 concentration ("CO_2 fertilization"). While these tests must be regarded with caution due to the low temporal resolution and uncertainties in the land-use data particularly during the early portion of the past millennium, overall a fairly low sensitivity in the estimates of γ is found. The land-use effects after ~ 1700 do impact the fit between temperature reconstructions and CO_2 data. The 1550-1800 late period shows increased spread towards both higher and lower γ's, with a few ppmv increase in the median. For the early 1050-1549 period, changes are negligible. The full period shows characteristics from both the early and late periods with an unchanged lower limit and increased spread towards higher values of γ. Overall, these sensitivity tests suggest that the differences in γ observed (Fig. 3) between the early and late periods cannot be readily explained by changes in land-use, and that the overall effect of land-use change has a fairly small impact in the evaluation of the sensitivity of the carbon cycle.*

Percentile	MWP to LIA	LIA to RW	MWP to RW
0%	0.05	0.30	-0.18
5%	0.14	0.47	0.09
10%	0.18	0.51	0.14
15%	0.21	0.54	0.17
20%	0.24	0.57	0.19
25%	0.27	0.59	0.21
30%	0.29	0.61	0.23
35%	0.31	0.63	0.25
40%	0.33	0.65	0.27
45%	0.36	0.67	0.29
50%	0.38	0.70	0.31
55%	0.41	0.72	0.32
60%	0.44	0.75	0.34
65%	0.48	0.79	0.36
70%	0.51	0.82	0.38
75%	0.55	0.86	0.40
80%	0.60	0.90	0.42
85%	0.66	0.94	0.45
90%	0.72	1.01	0.49
95%	0.81	1.12	0.57
100%	1.26	1.64	1.08

Table S1. *Percentiles for the amplitude difference in °C between the Medieval Warm Period (MWP) and Little Ice Age (LIA), LIA to Recent Warmth (RW), and the MWP to RW as estimated by the pairwise difference between ensemble distributions as shown in figure 2b in the main text. The MWP, LIA, and RW are respectively defined by 30-year periods as: 1071-1100, 1601-1630, 1971-2000. Amplitudes are given as positive numbers except for the zeroth percentile (lowermost range) for the MWP to RW, where this negative value is associated with a pairwise case when a temperature estimate of the MWP exceeded RW.*

Percentile	1050-1549	1050-1800	1550-1800	All
0%	-3.8	-3.2	0.8	-3.8
5%	-0.2	-0.2	4.1	0.2
10%	0.4	1.0	5.1	1.7
15%	1.3	3.6	6.0	2.5
20%	1.8	4.9	6.8	3.2
25%	2.1	5.9	7.5	3.9
30%	2.5	6.7	8.3	4.6
35%	2.7	7.5	9.2	5.4
40%	3.0	8.2	10.1	6.1
45%	3.3	8.9	11.1	6.9
50%	3.6	9.7	12.1	7.7
55%	4.0	10.5	13.2	8.6
60%	4.3	11.3	14.6	9.6
65%	4.7	12.2	16.1	10.7
70%	5.2	13.1	18.0	12.0
75%	5.8	14.2	20.3	13.4
80%	6.5	15.5	23.3	15.2
85%	7.3	17.0	27.4	17.7
90%	8.5	19.2	33.4	21.4
95%	10.7	23.1	42.9	28.7
100%	31.7	40.7	135.8	135.8

Table S2. *Percentile estimates of γ for the early (1050-1549), full (1050-1800), and late (1550-1800) analysis periods. Values for the early and late periods correspond to distributions in figure 3. All is the composite distribution derived by consideration of all estimates.*

Model	Span	N years ($\Delta T \leq 0.38°C$)	N years ($\Delta T \leq 0.70°C$)	γ (ppmv/°C) ($\Delta T \leq 0.38°C$)	γ (ppmv/°C) ($\Delta T \leq 0.70°C$)
HadCM3LC	1861-2100	116	143	11.2	12.3
IPSL-CM2C	1861-2100	99	129	2.4	6.0
MPI	1861-2100	111	140	4.8	6.5
CSM-1	1861-2100	189	216	5.1	6.3
IPSL-CM4-LOOP	1861-2100	101	134	3.7	2.1
BERN-CC	1765-2100	219	250	13.7	15.6
LLNL	1871-2100	97	116	-1.9	6.4
FRCGC	1901-2100	83	106	14.4	8.6
CLIMBER2-LPJ	1901-2100	89	111	5.0	6.6
Uvic-2.7	1861-2100	102	132	13.6	14.4
Mean				7.2	8.5
St.Dev.				5.6	4.3

Table S3. *Estimates of γ for the ten coupled carbon cycle-climate models range from 2.1 to 15.6 ppmv/°C for temperature changes ≤ 0.70 °C and -1.9 to 14.4 ppmv/°C for changes in temperature ≤ 0.38 °C. These two different thresholds help constrain assessments to the more typical variations experienced during the past millennium and generally exclude the 21st century projected warming from the analysis. The number of years available for these various model runs ranges from 83 to 250 – far too short to assess lagged responses and multi-decadal to century-scale changes in the temperature/ CO_2 coupling. For more details on the models, see ref. 12.*

References

1   Frank, D., Büntgen, U., Böhm, R., Maugeri, M. & Esper, J. Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. *Quat. Sci. Rev.* 26, 3298-3310 (2007).

2   Lee, T., Zwiers, F. & Tsao, M. Evaluation of proxy-based millennial reconstruction methods. *Clim. Dyn.* 31, 263-281 (2008).

3   Böhm, R. *et al.* The early instrumental warm-bias: a solution for long central European temperature series 1760 – 2007. *Clim. Change* (in press).

4   Frank, D., Esper, J. & Cook, E. R. Adjustment for proxy number and coherence in a large-scale temperature reconstruction. *Geophys. Res. Lett.* 34, L16709 doi:16710.11029/12007gl030571, doi:L16709 doi:10.1029/2007gl030571 (2007).

5   Büntgen, U., Frank, D. C., Nievergelt, D. & Esper, J. Summer temperature variations in the European Alps, AD 755-2004. *J. Clim.* 19, 5606-5623 (2006).

6   Mann, M., Rutherford, S., Wahl, E. & Ammann, C. Testing the fidelity of methods used in proxy-based reconstructions of past climate. *J. Clim.* 18, 4097-4107 (2005).

7   Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. & Jones, P. D. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.* 111, doi:10.1029/2005jd006548 (2006).

8   Woodwell, G. M. *et al.* Biotic feedbacks in the warming of the earth. *Clim. Change* 40, 495-518 (1998).

9   Strassmann, K. M., Joos, F. & Fischer, G. Simulating effects of land use changes on carbon fluxes: past contributions to atmospheric CO_2 increases and future commitments due to losses of terrestrial sink capacity. *Tellus B* 60, 583-603 (2008).

10   Klein Goldewijk, K. & Van Drecht, G. in *Integrated modelling of global environmental change. An overview of IMAGE 2.4*   (eds A F Bouwman, T Kram, & K Klein Goldewijk) (Netherlands Environmental Assessment Agency (MNP), Bilthoven, 2006).

11   Stocker, B. *Transient Simulations of Land Use Change in the Holocene - Separating the Human Impact from Natural Drivers of the Carbon Cycle* M.S. thesis, University of Bern, (2009).

12   Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis: Results from the (CMIP)-M-4 model intercomparison. *J. Clim.* 19, 3337-3353 (2006).