LETTER

Widespread seasonal compensation effects of spring warming on northern plant productivity

Wolfgang Buermann[1,2]*, Matthias Forkel[3], Michael O'Sullivan[1], Stephen Sitch[4], Pierre Friedlingstein[5], Vanessa Haverd[6], Atul K. Jain[7], Etsushi Kato[8], Markus Kautz[9], Sebastian Lienert[10,11], Danica Lombardozzi[12], Julia E. M. S. Nabel[13], Hanqin Tian[14,15], Andrew J. Wiltshire[16], Dan Zhu[17], William K. Smith[18] & Andrew D. Richardson[19,20]

Climate change is shifting the phenological cycles of plants[1], thereby altering the functioning of ecosystems, which in turn induces feedbacks to the climate system[2]. In northern (north of 30° N) ecosystems, warmer springs lead generally to an earlier onset of the growing season[3,4] and increased ecosystem productivity early in the season[5]. In situ[6] and regional[7–9] studies also provide evidence for lagged effects of spring warmth on plant productivity during the subsequent summer and autumn. However, our current understanding of these lagged effects, including their direction (beneficial or adverse) and geographic distribution, is still very limited. Here we analyse satellite, field-based and modelled data for the period 1982–2011 and show that there are widespread and contrasting lagged productivity responses to spring warmth across northern ecosystems. On the basis of the observational data, we find that roughly 15 per cent of the total study area of about 41 million square kilometres exhibits adverse lagged effects and that roughly 5 per cent of the total study area exhibits beneficial lagged effects. By contrast, current-generation terrestrial carbon-cycle models predict much lower areal fractions of adverse lagged effects (ranging from 1 to 14 per cent) and much higher areal fractions of beneficial lagged effects (ranging from 9 to 54 per cent). We find that elevation and seasonal precipitation patterns largely dictate the geographic pattern and direction of the lagged effects. Inadequate consideration in current models of the effects of the seasonal build-up of water stress on seasonal vegetation growth may therefore be able to explain the differences that we found between our observation-constrained estimates and the model-constrained estimates of lagged effects associated with spring warming. Overall, our results suggest that for many northern ecosystems the benefits of warmer springs on growing-season ecosystem productivity are effectively compensated for by the accumulation of seasonal water deficits, despite the fact that northern ecosystems are thought to be largely temperature- and radiation-limited[10].

Northern land regions have experienced substantial warming since the early 1970s, which has changed how ecosystems function[11]. One prominent example of emerging ecosystem responses is shifts in plant phenological cycles: earlier spring onset and delayed autumn senescence are lengthening the northern growing season[6,12]. These phenological shifts have altered ecosystem productivity[5,6,8,13,14] and the seasonality of important ecosystem feedbacks to the atmosphere and climate system[6,15].

Warmer and earlier springs may also influence ecosystem function later in the growing season through indirect or lagged effects[16,17].

For example, in situ studies provide evidence for substantial positive lagged effects on ecosystem productivity, whereby the influence of warmer springs may be conveyed to subsequent seasons through the development of larger leaves or increased foliar nitrogen[6]. By contrast, warmer or earlier springs may cause earlier autumn senescence because of the fixed lifespans of leaves[18] or adversely affect plant productivity later in the season through the build-up of water deficits[7–9,19,20]. However, a more comprehensive understanding of lagged productivity responses is still lacking.

Here, we use long-term (spanning the period 1982–2011) satellite data of vegetation greenness (as a proxy for potential photosynthesis)[21], flux-tower and model estimates of CO_2 uptake through photosynthesis (gross primary productivity, GPP)[22,23] and high-resolution climate data[24] to estimate the strength and geographic distribution of lagged effects that capture the influence of spring phenological transitions on plant productivity during the subsequent summer and autumn. Our analysis relies on identifying correlations between spring temperature (which serves as an independent phenological indicator) and satellite greenness or simulated GPP during spring and subsequent seasons to estimate concurrent phenological responsiveness and linked lagged effects (see Methods).

Across northern land, correlations between annual spring temperature and spring greenness show a significantly positive and spatially extensive pattern consistent with the notion of a tight control of spring temperature on concurrent plant productivity: 80% of northern (north of 30° N) vegetated non-agricultural land (total study area of roughly 41×10^6 km^2) exhibits statistically significant ($P < 0.05$ at the grid-cell level) positive correlations (Fig. 1a). To assess lagged effects on plant productivity associated with anomalous spring temperatures, we computed partial correlations between spring temperature and subsequent summer and autumn greenness, whereby covarying effects of concurrent climate on these correlations are controlled for (see Supplementary Information, section 1). Partial correlations between annual spring temperature and subsequent summer greenness show a widespread positive (6%, $P < 0.05$) and negative (6%, $P < 0.05$) pattern (Fig. 1b). Areas of positive partial correlations are predominantly situated in Eurasia, covering vast regions north of 50° N, whereas areas with negative correlations are more localized in western North America, Siberia and temperate eastern Asia. The partial correlation pattern between spring temperature and autumn greenness indicates an extension of the summer pattern of negative correlation (11%, $P < 0.05$; positive correlations cover only 2%, $P < 0.05$), with additional coverage seen
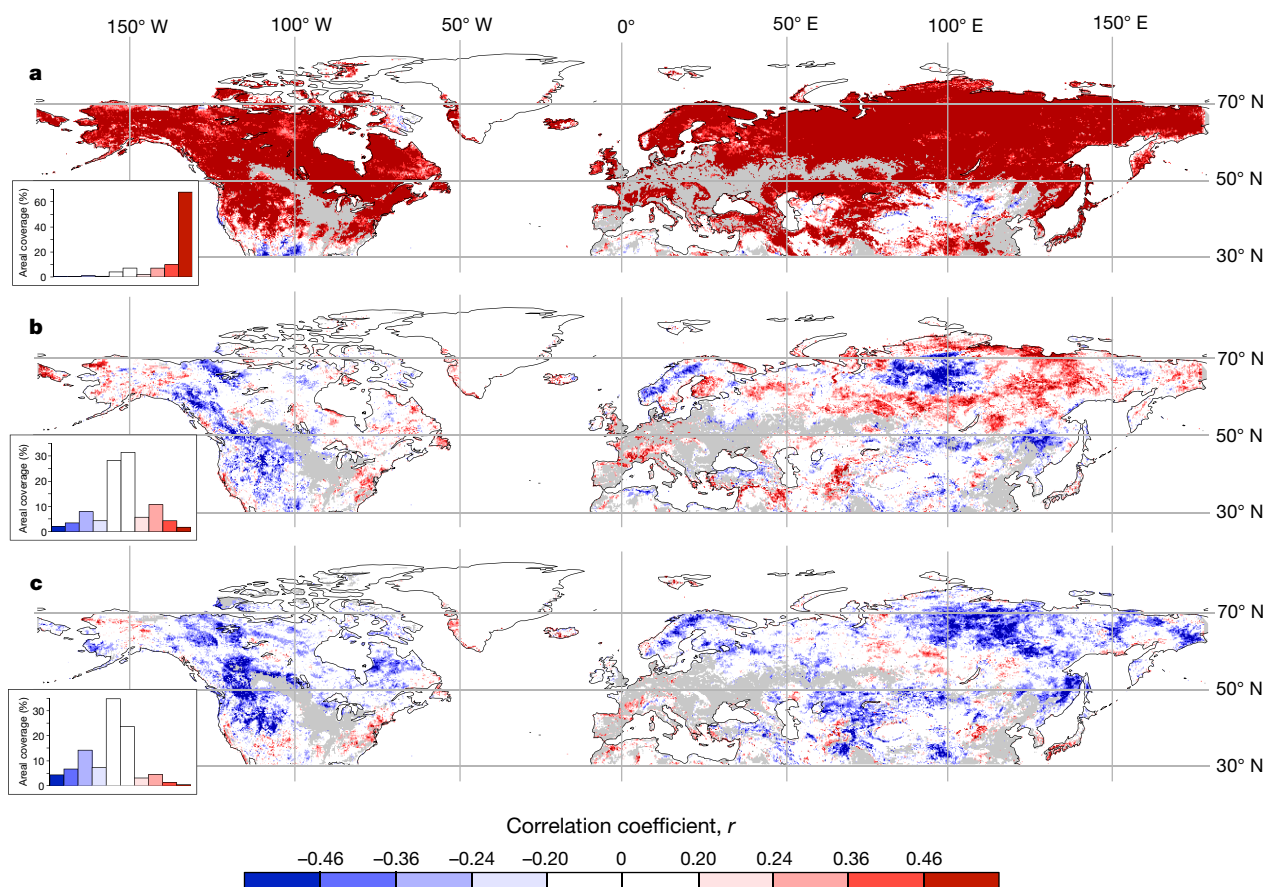
Fig. 1 | Spatial pattern of concurrent and lagged productivity responses to spring warming based on satellite greenness observations. a, Grid-cell correlations between yearly spring temperatures and spring satellite vegetation greenness (expressed through the normalized difference vegetation index, NDVI) for our study period, 1982–2011. **b**, **c**, Partial correlations between annual spring temperature and subsequent summer (**b**) and autumn (**c**) NDVI over this period. In these partial correlations, the covarying influences of summer temperature and precipitation (**b**) and autumn temperature and precipitation (**c**) on the correlations between spring temperature and summer or autumn NDVI have been removed. Seasons are defined using a local adaptive procedure (see Methods). Absolute values of the correlation coefficient (r) correspond to significance levels of $P = 0.3$ ($r = 0.20$), $P = 0.2$ ($r = 0.24$), $P = 0.05$ ($r = 0.36$) or $P = 0.01$ ($r = 0.46$). For each map, frequency histograms showing the areal coverage corresponding to positive and negative correlations, estimated as a fraction of total study area, are also provided (see insets). Areas that are cultivated or managed[30] (light grey) are not included in the analysis.

mainly in northeastern Eurasia and temperate central Asia (Fig. 1b, c). Although long-term trends in temperature and greenness could potentially influence these correlations, a corresponding analysis on detrended data shows that the patterns are similar (Supplementary Information, section 1). This similarity suggests a dominant influence of interannual to quasi-decadal variability on the correlation pattern between spring temperature and satellite greenness during subsequent seasons. A comparison of the strength of these lagged relationships with concurrent climatic influences on greenness pattern shows that at regional scales the influence of spring temperature on summer and autumn greenness can be equally important or even dominant (Supplementary Information, section 1).

To further assess the robustness of the lagged productivity responses that we identified from satellite data, we compared these responses to those inferred from flux-tower measurements of land–atmosphere CO_2 flux (FLUXNET). The results show that, across $n = 16$ tower sites, the strength and direction of relationships between spring temperature and spring and summer greenness derived from satellite data correspond well to those based on spring temperature and spring and summer GPP derived from tower data (Extended Data Fig. 1). However, the agreement between the relationships between spring temperature and autumn satellite greenness and between spring temperature and autumn tower-derived GPP is not as strong (Extended Data Fig. 1). This validation has several caveats, including the small number of available tower sites and the differences in spatial scales for satellite

(coarse) and tower (fine-scale) data; however, the overall consistency in the estimated lagged productivity responses suggests that the estimates based on satellite data are plausible.

The geographic distribution of the relationship between changes in spring temperature and subsequent summer greenness (see Fig. 1b) suggests that some combination of climate, elevation and land cover may explain these patterns. To investigate this, we conducted a random-forest analysis using a set of predictors that encapsulate such factors (see Supplementary Information, section 2). The results show that the partial correlation pattern between spring temperature and summer greenness can be explained with elevation and selected climate variables (such as summer precipitation and precipitation seasonality) acting as the most important variables (Extended Data Fig. 2, Supplementary Information, section 2). Across northern ecosystems, we find that these partial correlations tend to become more negative with higher elevation, but such well-defined directional relationships are not as apparent for important precipitation metrics (Extended Data Fig. 2).

Grouping the lagged productivity responses on the basis of the direction of robust correlations between spring temperature and spring, summer and autumn greenness reveals large clusters of regions with negative lagged effects and more scattered areas with positive lagged effects (Fig. 2a). As a result, negative lagged productivity responses associated with spring warming and greening, coupled with declines in summer or autumn greenness, stretch over vast areas in western North America, Siberia and to some extent eastern temperate Asia, whereas
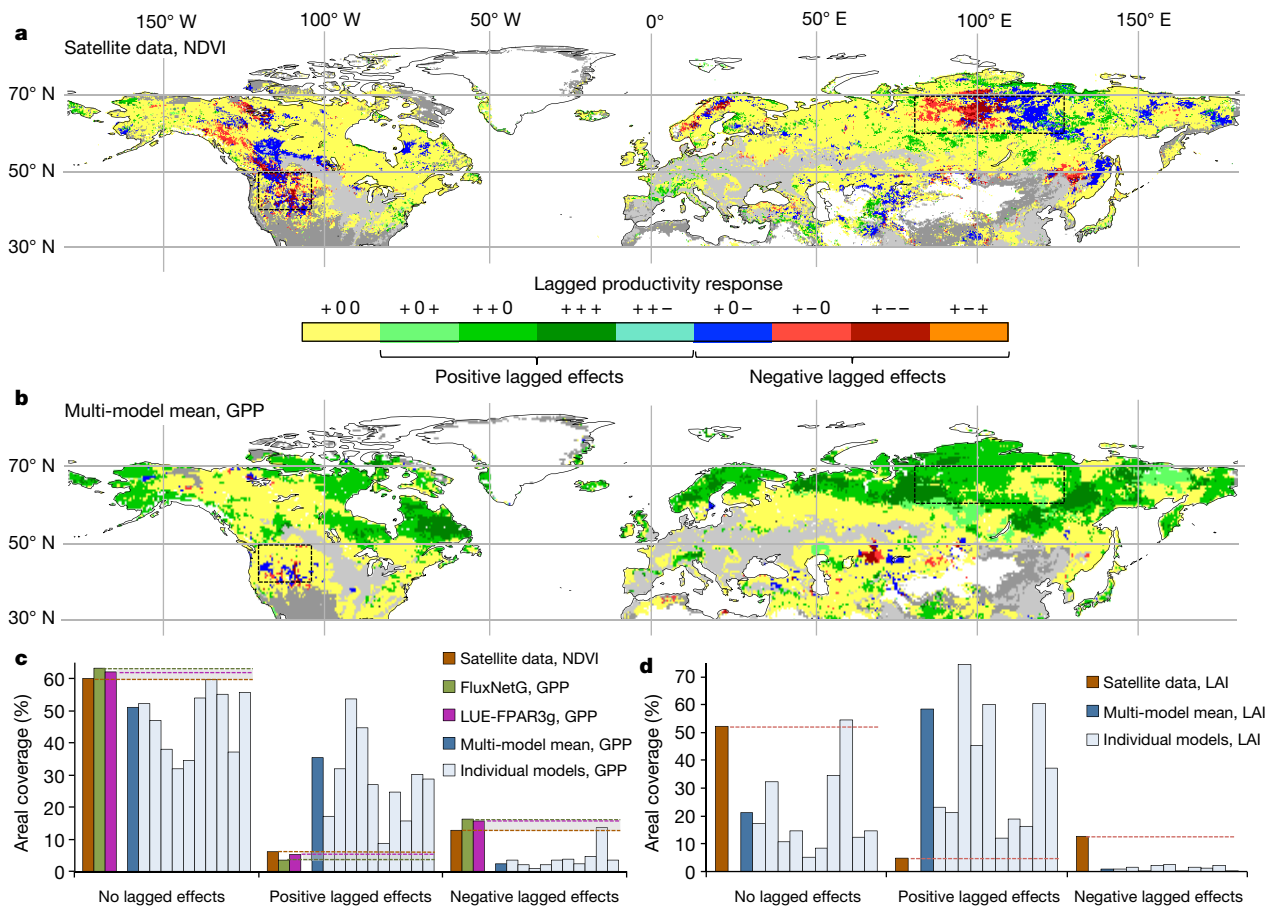
Fig. 2 | Spatial pattern of lagged productivity responses based on satellite greenness observations and models. a, **b**, The maps summarize the direction of robust ($P < 0.05$) grid-cell correlations between annual spring temperature and spring, summer and autumn NDVI determined from satellite data (**a**) or spring, summer and autumn GPP determined from the TRENDYv6 multi-model mean (**b**). For example, the lagged productivity response denoted as '$+-0$' represents positive correlations between spring temperature and spring NDVI or GPP, negative correlations between spring temperature and summer NDVI or GPP, and no correlations between spring temperature and autumn NDVI or GPP. The relationships between spring temperature and summer as well as autumn NDVI or GPP are estimated using partial correlations, whereby effects of covarying concurrent climate influences have been controlled for (see Fig. 1, Methods). The corresponding patterns for individual models are shown in Extended Data Fig. 3. Areas with no robust link between spring temperature and spring NDVI or GPP (dark grey) and areas that are cultivated or managed (light grey) are also shown. The two focal

regions in this study (western USA and Siberia) are indicated by black-dashed rectangles. **c**, Extent of areas with no, positive or negative lagged effects (see definition in **a**) within the study region for satellite NDVI data (brown) and GPP based on TRENDYv6 models (dark blue, multi-model mean; light blue, individual models). Corresponding results from a similar analysis for two satellite-data-constrained GPP datasets, based on upscaled FLUXNET data (FluxNetG; green) and a light-use-efficiency model (LUE-FPAR3g; magenta; see Methods), are also shown (see also Extended Data Fig. 4). The horizontal dashed lines are to aid comparison to the TRENDYv6 model results and the shaded regions encapsulate the spread among the three estimates derived from satellite-based approaches. **d**, Results from a complementary analysis for satellite-data-constrained and modelled LAI (see Methods). Results from the same analysis for detrended data show that the differences between observation- and model-based estimates of the areal fractions of positive and negative lagged effects are similar (Supplementary Information, section 1).

positive lagged effects are more common in eastern Eurasia north of 50° N (except Siberia).

Carbon-cycle models must be able to simulate the responses of vegetation phenology and the corresponding effects on ecosystem productivity and net carbon uptake realistically to estimate climate–carbon feedbacks credibly[25]. We therefore assessed the ability of ten current-generation models that contribute to TRENDYv6[22,23] to replicate the observed lagged productivity responses to spring warming. The results reveal a substantially higher multi-model mean areal coverage of positive lagged effects on plant productivity (and much lower coverage of negative lagged effects) than for the satellite estimates (Fig. 2a, b). Although there are marked differences among the individual models (Extended Data Fig. 3), a notable pattern in the ensemble is the near absence of any negative lagged effects across Siberia and the overall abundance of positive lagged effects that extend over summer and autumn (Fig. 2a, b). Satellite greenness has been used extensively as a proxy for vegetation productivity[3,26], but direct comparisons between greenness and GPP patterns are limited (see Methods). However,

a similar analysis using two satellite-data-constrained GPP datasets (based on upscaled FLUXNET data and a light-use-efficiency model; see Methods) reveals nearly identical lagged productivity patterns to those based on satellite greenness (Extended Data Fig. 4).

Grouping the lagged productivity responses more broadly, into positive and negative lagged effects, yields an areal extent of regions with positive lags of 36% for the TRENDYv6 ensemble (9%–54% for the ten individual TRENDYv6 models) and 4%–6% for the estimates derived from satellite data and satellite-data-constrained approaches (Fig. 2c). The areal coverage of negative lagged effects predicted by the TRENDYv6 ensemble is only 2% (1%–14% for the ten models), whereas that estimated from the satellite-based approaches is 13%–16%. (The ranges for the satellite-based estimates encapsulate the spread among the three different estimates; see shading in Fig. 2c.)

It is not clear why these terrestrial carbon-cycle models cannot adequately replicate the observed spatial pattern of lagged productivity responses to warmer springs. One key factor could be how seasonal vegetation growth is represented in the models. To assess this,
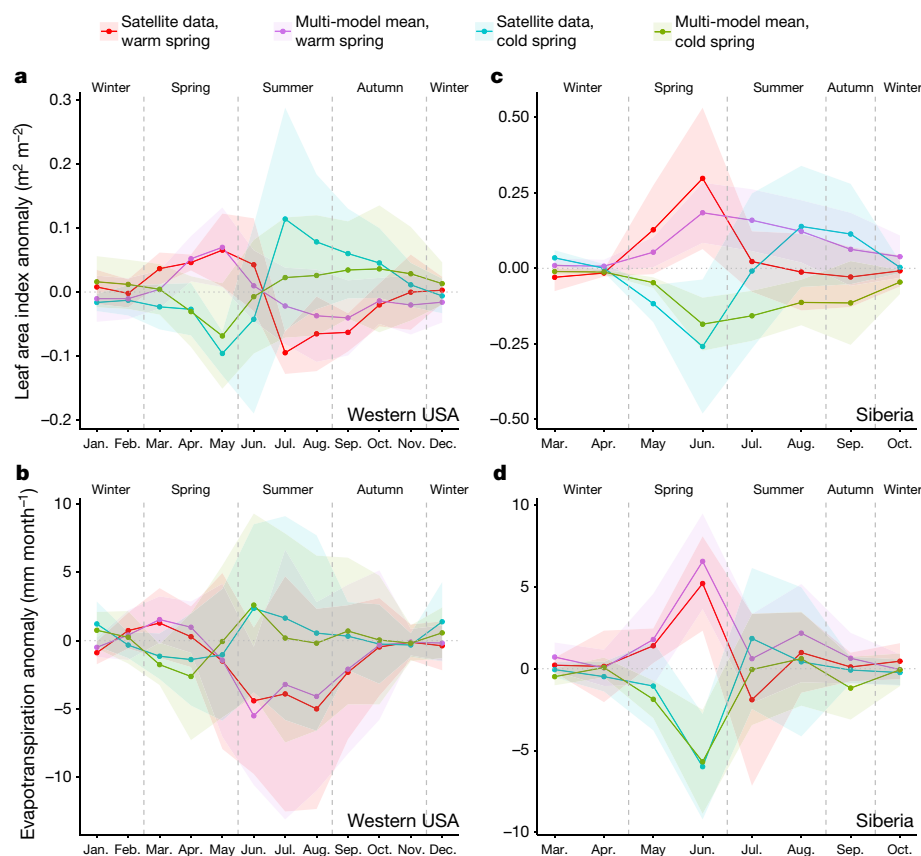
Fig. 3 | Seasonal trajectories of regionally averaged LAI and evapotranspiration anomalies based on observation-constrained and modelling approaches for warm- and cold-spring years.
a–d, Monthly anomalies in spatially averaged and maximum composited LAI (**a**, **c**) and evapotranspiration (**b**, **d**) based on satellite-data-constrained estimates (LAI3g, ET-GLEAM) and model simulations (TRENDYv6 multi-model mean) for western USA (**a**, **b**) and Siberia (**c**, **d**). Western USA encompasses the non-agricultural regions from 120° W to 105° W and 40° N to 50° N, whereas Siberia is defined to be

from 80° E to 125° E and 60° N to 70° N (see also Fig. 2). Anomalies are relative to the mean over the study period, 1982–2011. The monthly maximum composites shown are based on the mean LAI or evapotranspiration of the seven warmest- and coldest-spring years within the study period. The climatological seasons are indicated by the vertical grey dashed lines. Uncertainty bounds (shaded areas) reflect the spread in the monthly LAI or evapotranspiration anomalies within the compositing period (±1 s.d., $n = 7$).

we performed a similar seasonal correlation analysis with satellite and modelled leaf area index (LAI) data (see Methods). The results reveal an even larger discrepancy between the areal proportions of positive and negative lagged responses of LAI to spring warming determined using observation-based and modelling approaches compared to those determined using productivity metrics (Fig. 2c, d, Extended Data Fig. 4). The substantial overestimation of growing-season LAI in the models in response to spring warmth could cause too much new carbon to be allocated in plant tissue, which then enhances GPP.

Limited water availability may cause adverse lagged effects in response to spring warmth and could help to reconcile the differences between observations and models. To further investigate this we performed a regional analysis for western USA and Siberia, for which observation-based and simulated lagged productivity responses show more converging and diverging patterns, respectively (see Fig. 2). For western USA, we find that seasonal trajectories in aggregated satellite-data-constrained and modelled LAI and evapotranspiration display positive anomalies during spring in years with warmer springs and corresponding negative anomalies later in the growing season (suggestive of negative lagged effects associated with a build-up of water stress) (Fig. 3a, b). However, for Siberia, the seasonal trajectories in observation-constrained and modelled LAI for warm-spring years start to diverge substantially during summer and autumn, with the observations displaying more negative anomalies during summer and autumn (again suggestive of water stress) and the opposite pattern predicted by the models (Fig. 3c). Seasonal trajectories of observation-based and modelled evapotranspiration for years with anomalous

spring temperatures are more in agreement, although there is some indication that the models underestimate water stress in summer in warm-spring years (Fig. 3d). The consistency between the observed and modelled responses of LAI and evapotranspiration to spring warmth over western USA, a region that is known for its vulnerability to drought in response to spring warmth[27–29], suggests that the hydrology and phenology schemes included in the models are generally fit for purpose. The strong divergence between observation-based and modelled responses of seasonal vegetation growth to spring warmth over Siberia (which is dominated by needleleaf deciduous forests) may be due to underestimation of the effects of water stress on seasonal canopy development and general omission of fixed leaf lifespans in the models (Extended Data Table 1, Supplementary Information, section 3). We estimate that, owing to the difference between observation-based and modelled productivity responses to anomalous spring temperatures across Siberia, annual GPP for a warm-spring year may be up to four times higher in the TRENDYv6 ensemble (1.7 Pg C yr^{-1}) than an observation-constrained estimate based on upscaled FLUXNET data (0.4 Pg C yr^{-1}) (Extended Data Fig. 5).

Our analysis based on satellite vegetation records over multiple decades provides evidence for widespread positive and negative lagged plant-productivity responses across northern ecosystems associated with warmer springs. The spatially extensive pattern of negative lagged effects that we identified implies substantially reduced benefits for ecosystem productivity and carbon sequestration from longer northern growing seasons under climate change. We have also shown that current terrestrial carbon-cycle models substantially underestimate

(overestimate) negative (positive) lagged effects associated with spring warming. This is possibly because these models inadequately capture the effects of the seasonal build-up of water stress on seasonal vegetation growth. Continued monitoring of emerging ecosystem responses and improved modelling capabilities will therefore be crucial to improve our understanding of the complex interactions between a changing climate, shifts in phenological cycles and effects on energy, water and carbon cycles.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0555-7.

1. Fu, Y. H. et al. Declining global warming effects on the phenology of spring leaf unfolding. *Nature* **526**, 104–107 (2015).
2. Peñuelas, J., Rutishauser, T. & Filella, I. Phenology feedbacks on climate change. *Science* **324**, 887–888 (2009).
3. Myneni, R., Keeling, C., Tucker, C., Asrar, G. & Nemani, R. R. Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* **386**, 698–702 (1997).
4. Menzel, A. et al. European phenological response to climate change matches the warming pattern. *Glob. Change Biol.* **12**, 1969–1976 (2006).
5. Keenan, T. F. et al. Net carbon uptake has increased through warming-induced changes in temperate forest phenology. *Nat. Clim. Change* **4**, 598–604 (2014).
6. Richardson, A. D. et al. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agric. For. Meteorol.* **169**, 156–173 (2013).
7. Grippa, M. et al. The impact of snow depth and snowmelt on the vegetation variability over central Siberia. *Geophys. Res. Lett.* **32**, L21412 (2005).
8. Buermann, W., Bikash, P. R., Jung, M., Burn, D. H. & Reichstein, M. Earlier springs decrease peak summer productivity in North American boreal forests. *Environ. Res. Lett.* **8**, 024027 (2013).
9. Sippel, S. et al. Contrasting and interacting changes in simulated spring and summer carbon cycle extremes in European ecosystems. *Environ. Res. Lett.* **12**, 075006 (2017).
10. Nemani, R. R. et al. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* **300**, 1560–1563 (2003).
11. Settele, J. et al. in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the IPCC* (eds Field, C. B. et al.) 271–359 (Cambridge Univ. Press, Cambridge, 2014).
12. Forkel, M. et al. Codominant water control on global interannual variability and trends in land surface phenology and greenness. *Glob. Change Biol.* **21**, 3414–3435 (2015).
13. Piao, S. L. et al. Net carbon dioxide losses of northern ecosystems in response to autumn warming. *Nature* **451**, 49–52 (2008).
14. Forkel, M. et al. Enhanced seasonal CO$_2$ exchange caused by amplified plant productivity in northern ecosystems. *Science* **351**, 696–699 (2016).
15. Forzieri, G. et al. Satellites reveal contrasting responses of regional climate to the widespread greening of Earth. *Science* **356**, 1180–1184 (2017).
16. Richardson, A. D. et al. Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philos. Trans. R. Soc. Lond. B* **365**, 3227–3246 (2010).
17. Wolf, S. et al. Warm spring reduced carbon cycle impact of the 2012 US summer drought. *Proc. Natl Acad. Sci. USA* **113**, 5880–5885 (2016).
18. Keenan, T. & Richardson, A. D. The timing of autumn senescence is affected by the timing of spring phenology: implications for predictive models. *Glob. Change Biol.* **21**, 2634–2641 (2015).
19. Barnett, T. P., Adam, J. C. & Lettenmaier, D. P. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* **438**, 303–309 (2005).
20. Zhang, K., Kimball, J. S., Kim, Y. & McDonald, K. C. Changing freeze–thaw seasons in northern high latitudes and associated influences on evapotranspiration. *Hydrol. Processes* **25**, 4142–4151 (2011).
21. Pinzon, J. E. & Tucker, C. J. A. Non-stationary 1981–2012 AVHRR NDVI3g time series. *Remote Sens.* **6**, 6929–6960 (2014).
22. Sitch, S. et al. Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences* **12**, 653–679 (2015).
23. LeQuéré, C. et al. Global carbon budget 2017. *Earth Syst. Sci. Data* **10**, 405–448 (2018).
24. Harris, I., Jones, P. D., Osborna, T. J. & Lister, D. H. Updated high-resolution grids of monthly climate observations – the CRU TS3.10 dataset. *Int. J. Climatol.* **34**, 623–642 (2014).
25. Cadule, P. et al. Benchmarking coupled climate-carbon models against long-term atmospheric CO$_2$ measurements. *Glob. Biogeochem. Cycles* **24**, GB2016 (2010).
26. Huang, M. et al. Velocity of change in vegetation productivity over northern high latitudes. *Nat. Ecol. Evol.* **1**, 1649–1654 (2017).
27. Westerling, A. L., Hidalgo, H. G., Cayan, D. R. & Swetnam, T. W. Warming and earlier spring increase western U.S. forest wildfire activity. *Science* **313**, 940–943 (2006).
28. Sacks, W. J., Schimel, D. S. & Monson, R. K. Coupling between carbon cycling and climate in a high-elevation subalpine forest: a model-data fusion analysis. *Oecologia* **151**, 54–68 (2007).
29. Parida, B. R. & Buermann, W. Increasing summer drying in North American ecosystems in response to longer nonfrozen periods. *Geophys. Res. Lett.* **41**, 5476–5483 (2014).
30. Bartholomé, E. & Belward, A. S. GLC2000: a new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sens.* **26**, 1959–1977 (2005).

Author contributions W.B., M.F. and A.D.R. designed the research. W.B., M.F. and M.O. carried out the analysis and W.B. wrote the manuscript with contributions from all authors. S.S., P.F., V.H., A.K.J., E.K., M.K., S.L., D.L., J.E.M.S.N., H.T., A.J.W. and D.Z. contributed to the TRENDY results. W.K.S. contributed to the LUE-FPAR3g results.

Competing interests : The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0555-7.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0555-7.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to W.B.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data sources. For the satellite vegetation data, we used the GIMMS-NDVI version 3g (NDVI3g)[21] and LAI3g[31] products, which are both available at 8-km spatial and 15-day temporal resolution for our study period, 1982–2011. The NDVI3g data stem from optical surface reflectance measurements from a series of NOAA-AVHRR satellites. Effects of orbital drifts, inter-sensor calibration and stratospheric aerosols from volcanic eruption have been corrected for, making this the most consistent long-term satellite vegetation dataset currently available[21]. The LAI3g fields are derived from the NDVI3g data using an artificial neural network model[31]. Gridded monthly climate data were obtained from the Climatic Research Unit (CRU TS3.23) at 0.5° spatial resolution[24] for our study period (1982–2011). As an estimate for the observation-constrained evapotranspiration (ET), we included the Global Land Evaporation Amsterdam Model (GLEAM) dataset, which has a spatial resolution of 0.25° at daily time steps[32]. While the GLEAM approach is based on an empirical model, it is heavily constrained by observations through assimilating satellite microwave vegetation optical depth data as a proxy for water stress[32]. In addition, land-cover data used in this study are based on the GLC2000 land-cover classification[30]. For complementary analyses, we also used site-level GPP data derived from the global FLUXNET tower network (FLUXNET2015, tier 1) and two observation-constrained, gridded monthly GPP datasets. The first includes GPP data (0.5° spatial resolution, available for 1982–2008) estimated from upscaled carbon observations based on FLUXNET (FluxNetG)[33]. FluxNetG is different from the previously published FluxNet-MTE[33] because it has been produced with inputs from only a single satellite vegetation dataset (NDVIg; a predecessor of NDVI3g) to reduce artefacts from using multiple satellite data (the FluxNetG dataset was also used in ref. [8]). Second, we used GPP data (0.5° spatial resolution, available for 1982–2011) derived using the light-use-efficiency (LUE) MODIS GPP algorithm driven by bi-monthly GIMMS FPAR3g (LUE-FPAR3g)[34]. Additional meteorological driver data required as input into the MODIS GPP algorithm were derived from NCEP-DOE Reanalysis II (http://www.esrl.noaa.gov). For more information on the GIMMS3g GPP dataset, see ref. [34].

TRENDYv6 models. We also analysed monthly GPP, LAI and ET simulation outputs for 1982–2011 from ten terrestrial carbon-cycle models that were part of a recent model intercomparison project, TRENDYv6[22,23]. The models included in the analysis here are LPX-Bern, LPJ-GUESS, ISAM, CABLE, VISIT, CLM4.5, DLEM, JSBACH, ORCHIDEE-MICT and JULES. In TRENDYv6, the models were forced with the CRUNCEPv6 climate dataset, which is based on a merged product of the monthly CRU climate data, and to be consistent with the TRENDYv6 ensemble we also used this climate dataset here. In addition, a set of factorial simulations[22] were performed and we analysed outputs from a simulation in which only atmospheric CO_2 and climate were varied (land-use change held fixed; experiment 'S2') because our study focus was on non-agricultural ecosystems. For an overview of the processes included in the models relevant to this study, see Extended Data Table 1. For a more general overview of the models see tables 4a and 5 in ref. [23].

Analysis framework. The satellite bi-monthly GIMMS NDVI3g and LAI3g vegetation data were averaged to a monthly temporal resolution (to be consistent with the TRENDYv6 model outputs). Then, the fine-scale satellite vegetation and coarse-scale CRU temperature fields were (dis)aggregated to a common 0.25° spatial grid on which all correlation analyses were performed. The motivation for this spatial aggregation step is twofold: (i) it retains a certain level of spatial information inherent in the satellite products and (ii) it aligns more closely with the coarser spatial resolutions of the TRENDY carbon-cycle models. Model outputs from TRENDYv6 were either analysed at their native model resolutions spanning grid-cell dimensions from 0.5° to 1.9° (ref. [22]) or resampled to a common 0.5° grid through nearest neighbours (for example, to estimate multi-model means of GPP, LAI and ET at grid-cell levels).

To estimate lagged vegetation growth and productivity responses we first divided the mean seasonal cycle of NDVI or simulated GPP (based on the 30-year study period) into spring, summer and autumn periods for each grid cell. Hereby, the start of spring and the end of autumn are defined by the months in which corresponding temperatures are closest to 0 °C, whereas the start and end of summer are defined by the months in which the NDVI (GPP) is closest to 95% (85%) of the annual maximum NDVI (GPP). Alternative approaches for characterizing

phenological cycles involving start and end dates of the growing season are more ambiguous if based solely on optical vegetation indices[35,36] or when the underlying data have relatively low temporal resolution, as in this study[12].
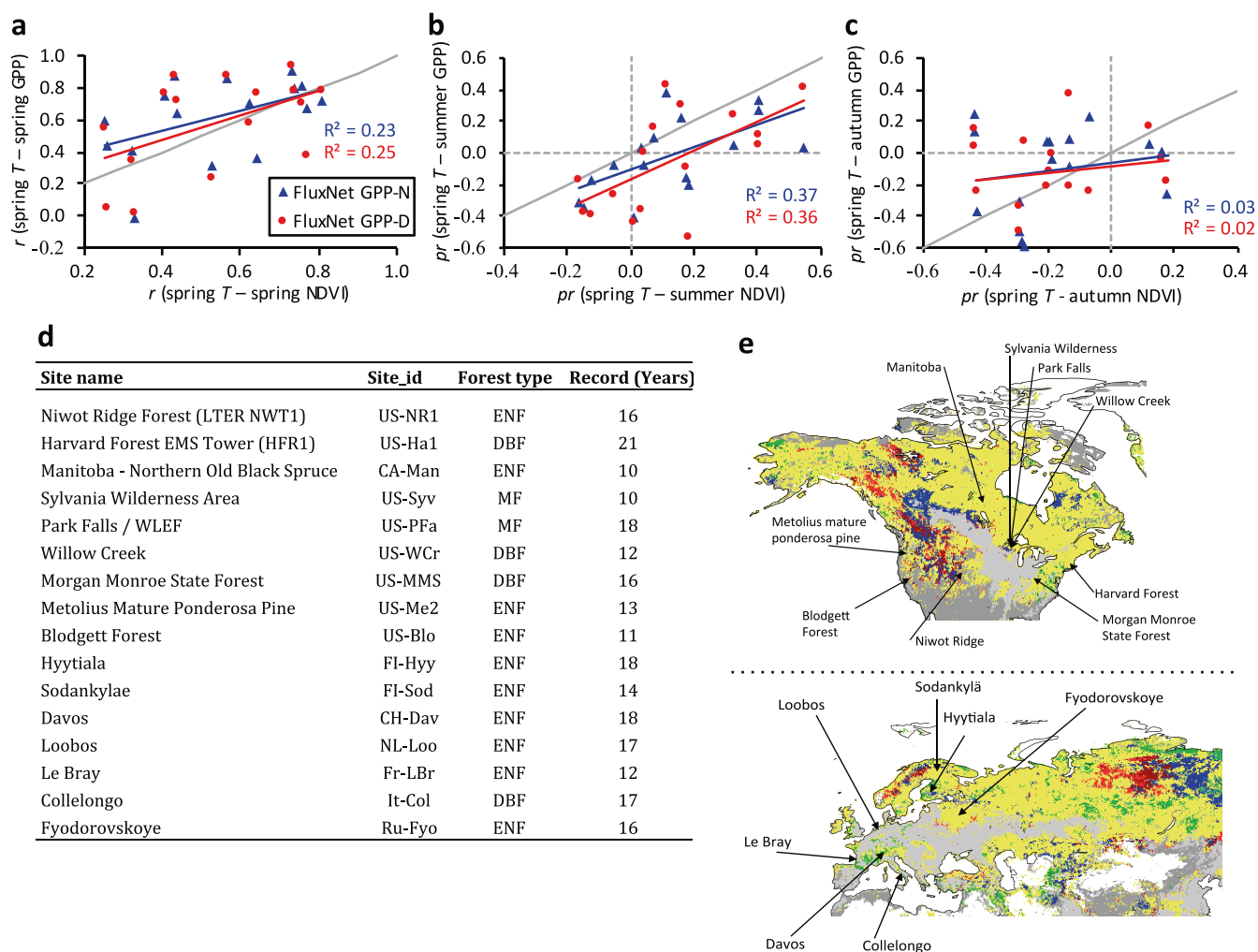
In a next step, we (building on the conceptual model of ref. [16]) classified lagged productivity responses for each grid cell as follows. First, as a minimum requirement for phenological responsiveness to spring warming, we require the spring temperature and the response variable of interest (NDVI, LAI or GPP) to be significantly ($P < 0.05$) positively correlated. Second, we define a lagged productivity (NDVI, GPP) or phenology (LAI) response on the basis of the direction of robust ($P < 0.05$) partial correlations between annual spring temperature (as an independent phenological indicator) and subsequent summer and autumn seasonal means of the response variable of interest; for example, if at a given locality the annual spring temperature is positively correlated with spring NDVI but negatively correlated with subsequent summer NDVI and not robustly correlated with autumn NDVI, then the response label would be '$+-0$', with the type of symbol denoting the direction of correlations and sequence corresponding to spring–spring, spring–summer and spring–autumn relationships (see Fig. 2). Partial correlations are used to control for covarying effects of climate over seasonal timescales, which can confound the correlations between annual spring temperature and subsequent summer and autumn response variables (see Supplementary Information, section 1).

As indicated, the satellite vegetation data (NDVI3g, LAI3g) used here stem from a series of satellites; although this record has been assembled carefully and validated to some extent[31], remaining non-vegetation artefacts in the data cannot be ruled out[37]. Further, satellite greenness (or NDVI) captures the amount of light absorbed by chlorophyll in green leaves[38] and has been used extensively as a proxy for spatially resolved vegetation productivity at continental and multi-decadal scales[3,26]. However, to overcome the limited comparability of directly observed NDVI-based and simulated GPP-based patterns, we also analysed observation-constrained GPP data. The results show good agreement between the lagged productivity patterns at both the site level (using GPP flux-tower data) and across northern ecosystems (using gridded GPP data from up-scaled FLUXNET and a LUE model), providing further support for the robustness of our results (see Extended Data Fig. 4). Finally, we also used satellite and modelled LAI data to probe the mismatch between lagged greenness and modelled (TRENDYv6) GPP responses to spring warmth.
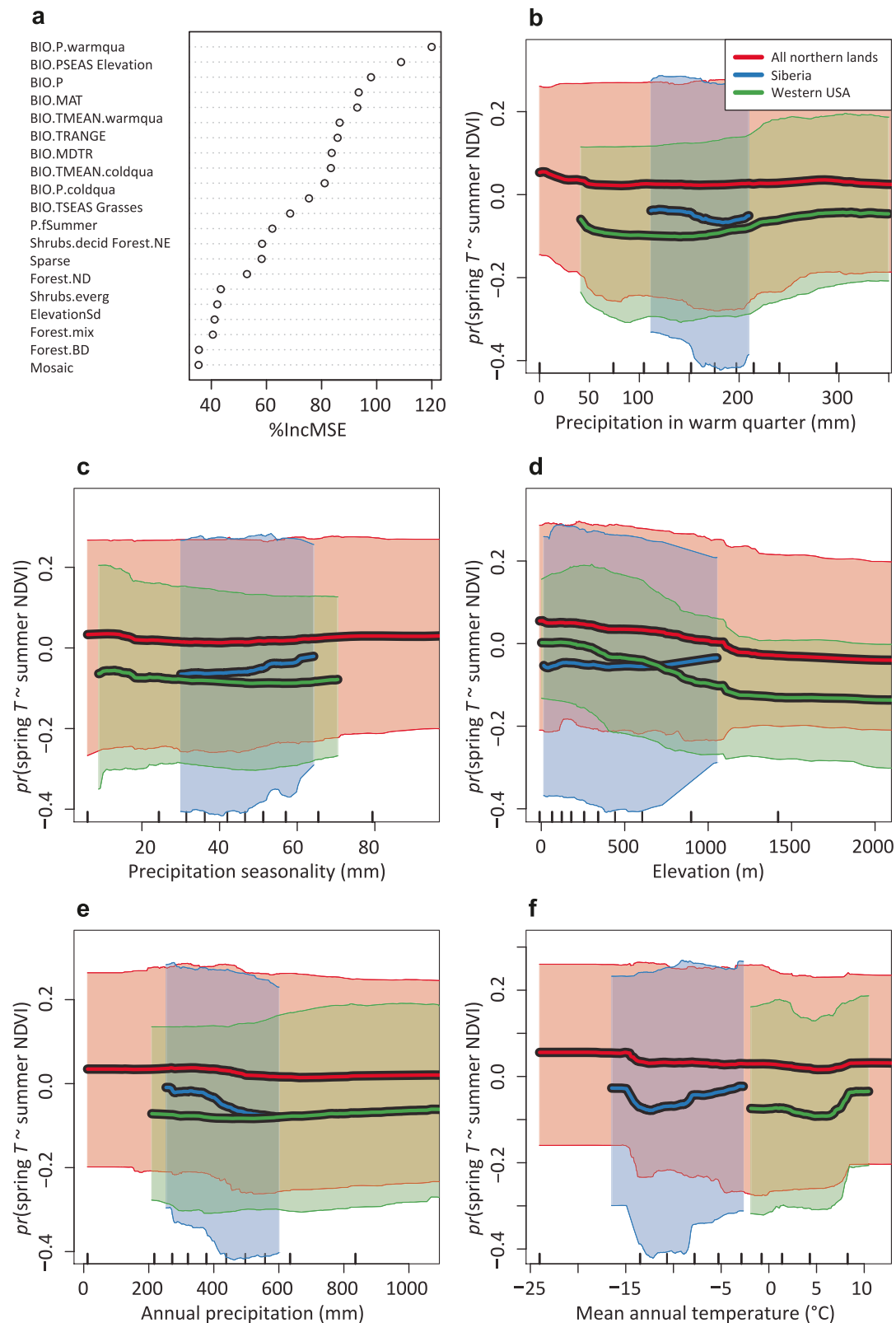
Data availability

The satellite NDVI3g data that support the findings of this study were downloaded from http://ecocast.arc.nasa.gov/data/pub/gimms/3g.v0/. The satellite LAI3g data are available from R. B. Myneni (rmyneni@bu.edu) on reasonable request. The LUE-FPAR3g GPP data can be requested from W.K.S. (wksmith@email.arizona.edu) and the FluxNetG GPP data from M. Jung (mjung@bgc-jena.mpg.de). The TRENDYv6 data are available from S.S. (s.a.sitch@exeter.ac.uk) on reasonable request.

31. Zhu, Z. et al. Global data sets of vegetation leaf area index (LAI) 3g and fraction of photosynthetically active radiation (FPAR) 3g derived from global inventory modeling and mapping studies (GIMMS) normalized difference vegetation index (NDVI3g) for the period 1981 to 2011. *Remote Sens.* **5**, 927–948 (2013).
32. Martens, B. et al. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.* **10**, 1903–1925 (2017).
33. Jung, M. et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J. Geophys. Res. Biogeosci.* **116**, G00J07 (2011).
34. Smith, W. K. et al. Large divergence of satellite and Earth system model estimates of global terrestrial CO_2 fertilization. *Nat. Clim. Change* **6**, 306–310 (2016).
35. Walther, S. et al. Satellite chlorophyll fluorescence measurements reveal large-scale decoupling of photosynthesis and greenness dynamics in boreal evergreen forests. *Glob. Change Biol.* **22**, 2979–2996 (2016).
36. Wu, C. et al. Land surface phenology derived from normalized difference vegetation index (NDVI) at global FLUXNET sites. *Agric. For. Meteorol.* **233**, 171–182 (2017).
37. Jiang, C. et al. Inconsistencies of interannual variability and trends in long-term satellite leaf area index products. *Glob. Change Biol.* **23**, 4133–4146 (2017).
38. Myneni, R. B. et al. The interpretation of spectral vegetation indexes. *IEEE Trans. Geosci. Remote Sens.* **33**, 481–486 (1995).

a

r (spring T − spring GPP) vs r (spring T − spring NDVI)

R² = 0.23
R² = 0.25

▲ FluxNet GPP-N
● FluxNet GPP-D

b

pr (spring T − summer GPP) vs pr (spring T − summer NDVI)

R² = 0.37
R² = 0.36

c

pr (spring T − autumn GPP) vs pr (spring T - autumn NDVI)

R² = 0.03
R² = 0.02

d

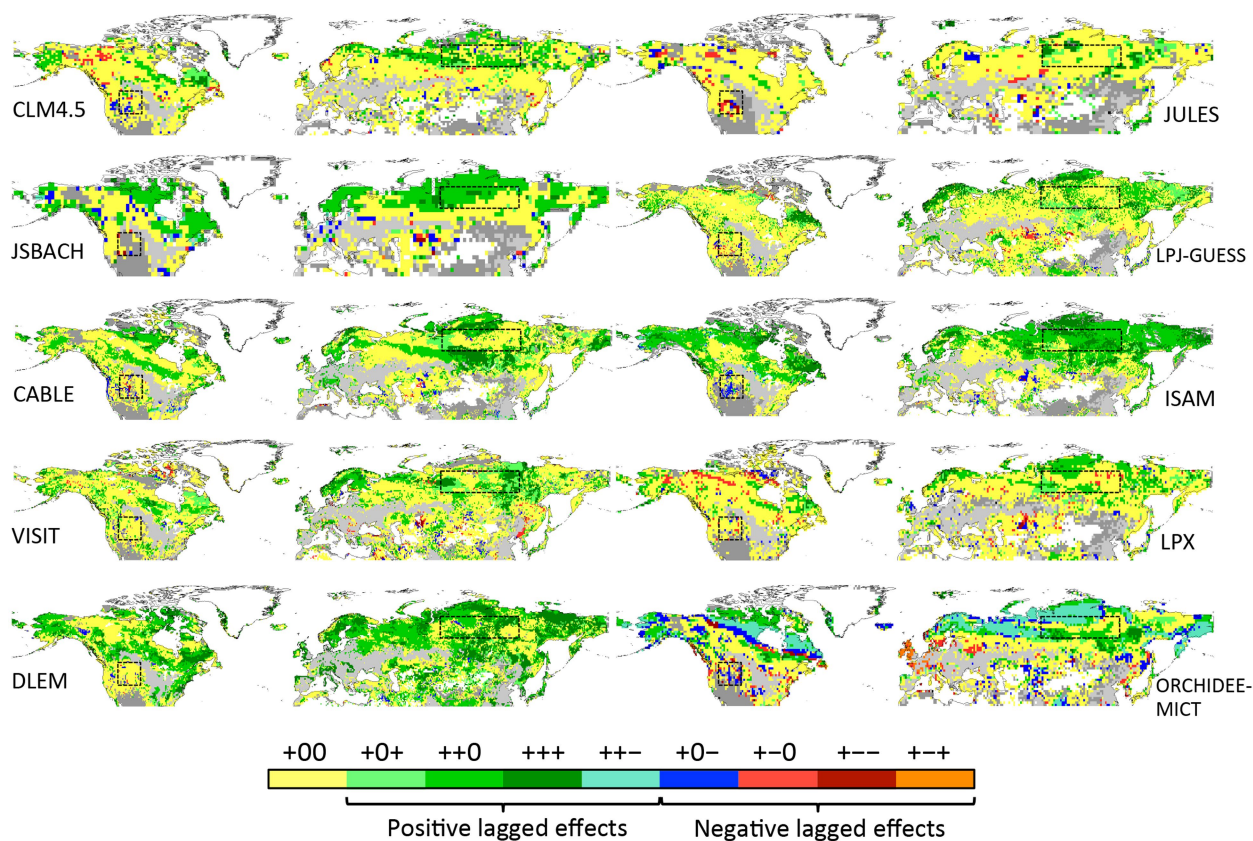| Site name	Site_id	Forest type	Record (Years)
Niwot Ridge Forest (LTER NWT1)	US-NR1	ENF	16
Harvard Forest EMS Tower (HFR1)	US-Ha1	DBF	21
Manitoba - Northern Old Black Spruce	CA-Man	ENF	10
Sylvania Wilderness Area	US-Syv	MF	10
Park Falls / WLEF	US-PFa	MF	18
Willow Creek	US-WCr	DBF	12
Morgan Monroe State Forest	US-MMS	DBF	16
Metolius Mature Ponderosa Pine	US-Me2	ENF	13
Blodgett Forest	US-Blo	ENF	11
Hyytiala	FI-Hyy	ENF	18
Sodankylae	FI-Sod	ENF	14
Davos	CH-Dav	ENF	18
Loobos	NL-Loo	ENF	17
Le Bray	Fr-LBr	ENF	12
Collelongo	It-Col	DBF	17
Fyodorovskoye	Ru-Fyo	ENF	16

e

Extended Data Fig. 1 | Comparison of lagged productivity responses based on satellite greenness observations and in situ estimates of carbon fluxes across selected FLUXNET sites. a–c, Site-specific correlations between spring temperature (T) and spring (**a**), summer (**b**) or autumn (**c**) satellite NDVI (x axis) plotted over the corresponding site-specific correlations between spring temperature and spring (**a**), summer (**b**) or autumn (**c**) flux-tower GPP (y axis). In **b** and **c**, the relationships are based on partial correlations (pr) between spring temperature and subsequent summer (**b**) or autumn (**c**) NDVI or GPP, with covarying effects of summer temperature and precipitation (**b**) and autumn temperature and precipitation (**c**) removed. (Partial) correlations are shown for two estimates of GPP: GPP-N (based on night-time partitioning of net ecosystem exchange) and GPP-D (daytime partitioning). **d,** For this comparison, satellite NDVI time series at 8-km (native) spatial resolution have been extracted for the 16 FluxNet tower sites with at least 10-year data records. Forest types for the tower sites are: ENF, evergeen needleleaf forest; DBF, deciduous broadleaf forest; MF, mixed forest. **e,** Maps showing the approximate locations of the FLUXNET tower sites. FLUXNET data for this comparative analysis are from the FLUXNET2015 dataset (tier 1).

Extended Data Fig. 2 | Random-forest analysis to explain the partial correlation pattern between annual spring temperature and summer satellite greenness on hemispheric and regional scales. a, Ranked importance of a set of explanatory variables in a random-forest model for the whole northern ecosystem study region, encompassing all vegetated non-agricultural land north of 30° N (see Supplementary Information, section 2, for details on the explanatory variables used). The ranking is based on the highest increment in mean squared error (IncMSE) between the observed and random-forest-predicted correlation after permuting the relevant explanatory variable. **b–f**, Individual conditional expectation lines of the random-forest-predicted partial correlation (pr) between spring temperature (T) and summer NDVI for the five most important explanatory variables. Lines and shaded bands reflect the mean (regional-average response) and the 5%–95% percentile range (grid-cell-level responses to environmental predictors) for the northern (north of 30° N, non-managed) study region (red) and for the focus regions (Siberia, blue; western USA, green) (see Supplementary Information, section 2).

Extended Data Fig. 3 | Spatial pattern of lagged productivity responses based on the individual carbon-cycle models included in TRENDYv6. All patterns are based on monthly GPP over the period 1982–2011, using outputs from the ten TRENDYv6 models included in the analysis (see Methods). The maps summarize the direction of statistically significant ($P < 0.05$) correlation between annual spring temperature and spring, summer or autumn GPP. For details on classification scenarios and contour labels, see Fig. 2. Areas with no robust link between spring temperature and spring GPP (dark grey) and areas that are cultivated or managed (light grey) are also shown.

Lagged Productivity Response Scenarios   **Lagged Vegetation Growth Response Scenarios**

+00	+0+	++0	+++	++−	+0−	+−0	+−−	+−+

Positive lagged effects   Negative lagged effects

Extended Data Fig. 4 | Spatial pattern of lagged productivity and vegetation growth responses estimated through satellite-based and modelling approaches. a–f, Summary of the direction of robust ($P < 0.05$) correlations between annual spring temperature and spring, summer or autumn satellite NDVI (**a**), satellite LAI (**b**), satellite upscaled GPP (FluxNetG; **c**), satellite-data-driven LUE-modelled GPP (LUE-FPAR3g; **d**), and multi-model mean GPP (**e**) and LAI (**f**) based on the ten TRENDYv6 models. For details on scenario classifications and contour labels see Fig. 2. Arrows (arrows with strikethroughs) linking panels highlight qualitative agreement (disagreement) between the lagged responses of productivity and vegetation growth based on the various approaches.

Western US

Siberia

Extended Data Fig. 5 | Changes in regional climate, satellite greenness and plant carbon fluxes from observation-constrained and modelling approaches for warm- and cold-spring years. a–f, Monthly anomalies in regionally averaged maximum composited climate (**a, d**), NDVI (**b, e**) and GPP (**c, f**) for warm- and cold-spring years, for the focus regions (**a–c**, western USA; **d–f**, Siberia). The anomalies are relative to the mean of the study period (1982–2011) and are based on maximum composites of monthly means of the seven warmest- and coldest-spring years within the study period. The observation-constrained GPP anomalies (**c, f**) stem from FluxNetG, which combined GPP estimates from flux towers with climate and satellite greenness in a machine-learning framework (see Methods). The boundaries between the climatological seasons are indicated by vertical grey dashed lines. Uncertainty bounds (shaded areas) reflect the spread in the respective monthly anomalies within the compositing period (± 1 s.d., $n = 7$). On the basis of these anomalies, we estimate, for a warm-spring year (relative to mean conditions) in Siberia (area, 2.5×10^6 km^2), annual GPP increases of 0.4 Pg C and 1.7 Pg C for FluxNetG and the TRENDYv6 ensemble, respectively, which corresponds to higher plant carbon uptake in the TRENDYv6 ensemble by a factor of roughly four (**f**). This is, to a large extent (about 64%), because of the overestimation of positive lagged effects in the TRENDYv6 models, but another important factor (36%) is the higher sensitivity of concurrent carbon uptake to spring warming in the TRENDYv6 models (compared to FluxNetG).

Extended Data Table 1 | Comparison of how specific processes relevant to this study are represented in the TRENDYv6 carbon-cycle models

	CLM4.5	LPX-Bern	LPJ-GUESS	ISAM	CABLE	VISIT	DLEM	JSBACH	ORCHIDEE-MICT	JULES
Phenology scheme (incl. fixed leaf life span?)	Prognostic, based on carbon availability and allocation. Relatively fixed leaf lifespan (some variation in stress-deciduous PFTs)	Temperature based phenology with drought response, no fixed leaf life span for deciduous trees	Daily leaf phenology based on temperature and water availability	Carbon-gain-based dynamic phenology. The leaf onset starts when the environmental conditions are advantageous for the plant to produce leaves and to start carbon assimilation; and the leaf offset starts when the plant experiences unfavourable environmental conditions resulting in C loss, such as cold temperature, photoperiod (shorter day length), and soil moisture stress	Daily leaf phenology based on temperature and water availability for grasses and cold-deciduous woody PFTs (deciduous needle-leaf and deciduous broad-leaf) Evergreen needle-leaf trees have fixed leaf turnover rate.	Monthly leaf phenology based on temperature and water availability[1]	Satellite-constrained daily leaf phenology, no fixed leaf life span	Phenology scheme based on logistic growth with two different forest phenology types in the considered study area: evergreen and summergreen. Both types have temperature-based phenologies with different leaf shedding rates.	Leaf onset for temperate and boreal deciduous trees is driven by the accumulation of warm temperatures (growing degree day) and the number of chilling days. Leaf senescence is based on air temperature for deciduous trees and on both temperature and moisture for grasses. For all PFTs, another leaf turnover is triggered if leaf age exceeds a PFT-specific parameter of critical leaf age, and thus leaf biomass gradually decrease with time.	Daily, based on thermal time above 5C. No drought phenology or fixed leaf age
Permafrost (yes/no?)	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes	No
Dynamic vegetation or static/prescribed land cover?	Prescribed land cover	Dynamic vegetation	Dynamic vegetation	Prescribed land cover	Prescribed land-cover	Prescribed land cover	Prescribed land cover	Prescribed land cover	Prescribed land cover	Dynamic vegetation
Explicit representation of decid. needleleaf forests?	Yes	Yes	Yes	Yes, model explicitly account for primary and secondary decid. Needleleaf forests.	Yes	Yes (as boreal deciduous needle-leaved forest)	Yes	no (in this simulation only one PFT was used to represent temperate deciduous forests)	yes	No
Carbon allocation scheme (how and how often updated?	Allocation to individual living tissue components updated at every model timestep (30 minutes)	Annual allocation of carbon to leaf, sapwood and root compartments, according to PFT dependent allometric rules.	annual carbon allocation from individual NPP to living tissue compartments[3]	The allocation of carbon to leaf, stem, root, and grain pools is a dynamic process based on temperature, water availability, light, and nutrients to alter the carbon allocation fractions dynamically at each model time step (1h)	Daily carbon allocation. Static fraction to roots, except during green-up of deciduous Pfts, when carbon is preferentially allocated above-ground. Relative allocation to leaves and wood is constrained by leaf-area to sapwood-area ratio.	Allocation for leaf first, then residual carbon is allocated to stem and root with a constant ratio, in order to realize the biome-specific growth form at each time step (monthly).	Daily carbon allocation from NPP to individual tissue	NPP enters the vegetation C pools daily according to fixed PFT specific proportions, adapted to adhere to structural limits.	Daily allocation from NPP to individual compartments (leaf, root, sapwood, fruit, carbohydrate reserve), according to PFT-specific fractions and regulated by light limitation, soil moisture and soil temperature.	10 daily allocation between growth and spreading. Growth allocation is split between leaves, roots and stem. Spreading allocation to increasing fractional coverage of pools.

PFT, plant functional type; NPP, net primary productivity.