Extratropical forests increasingly at risk due to lightning fires

Thomas A. J. Janssen [1] ✉, Matthew W. Jones [2], Declan Finney [3], Guido R. van der Werf [1,4], Dave van Wees [1,5], Wenxuan Xu [6] & Sander Veraverbeke [1]

Fires can be ignited by people or by natural causes, which are almost exclusively lightning strikes. Discriminating between lightning and anthropogenic fires is paramount when estimating impacts of changing socioeconomic and climatological conditions on fire activity. Here we use reference data of fire ignition locations, cause and burned area from seven world regions in a machine-learning approach to obtain a global attribution of lightning and anthropogenic ignitions as dominant fire ignition sources. We show that 77% (uncertainty expressed as one standard deviation = 8%) of the burned area in extratropical intact forests currently stems from lightning and that these areas will probably experience 11 to 31% more lightning per degree warming. Extratropical forests are of global importance for carbon storage. They currently experience high fire-related forest losses and have, per unit area, among the largest fire emissions on Earth. Future increases in lightning in intact forest may therefore compound the positive feedback loop between climate change and extratropical wildfires.

Global fire activity has decreased since the early 2000s, driven mainly by the decline of human-caused fires in savannahs, linked to a mixture of agricultural intensification, cropland expansion and shifting rainfall patterns[1,2]. By contrast, fire activity has increased in many forested extratropical regions, including the western United States[3] and Southeast Australia[4], eastern Siberia[5] and western Canada[6]. Tropical savannah and grassland fires account for approximately 67% of the global burned area (Fig. 1a), while extratropical forest fires account for less than 5% of the global burned area[7]. Because of the abundance of tropical savannah and grassland fires, changes in these fires have dominated the declining global burned-area trend[1,7]. Nevertheless, extratropical forests are of global importance because of their considerable carbon (C) storage and biodiversity[8]. While changes in extratropical forest fire activity may be masked in global trends in burned area[9,10], the C emitted from these fires has offset the reduced C emissions from fires in tropical savannahs[9]. Therefore, increased fire activity in extratropical forests is showing potential to disproportionately perturb the global C cycle[11].

Fires can have an anthropogenic or natural ignition source. Anthropogenic fires vary widely in their drivers and include, among others, planned management fires, deforestation fires, accidents and arson[12]. Natural fires originate almost exclusively from lightning[13]. Lightning and anthropogenic fires have different characteristics[14,15]. In the United States, for example, lightning fires are on average larger, more intense and more strictly constrained to remote areas and periods of extreme fuel dryness than anthropogenic fires[14]. In general, lightning fires prevail in remote landscapes and usually occur during the relatively short seasonal peak in lightning occurrence, which often co-occurs with a seasonal low in fuel moisture[16–18]. Conversely, anthropogenic fires occur throughout the year, and human accessibility is a key determinant of anthropogenic ignitions[19].

Due to their different drivers and characteristics, lightning and anthropogenic fires might show diverging future trends in different regions. While, for example, anthropogenic savannah fires might show a continued decline in the future with agricultural intensification[1] and

[1]Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [2]Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich, UK. [3]School of Earth and Environment, University of Leeds, Leeds, UK. [4]Meteorology and Air Quality Group, Wageningen University and Research, Wageningen, the Netherlands. [5]BeZero Carbon Ltd, London, UK. [6]Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences, Nanjing, China. ✉e-mail: t.a.j.janssen@vu.nl
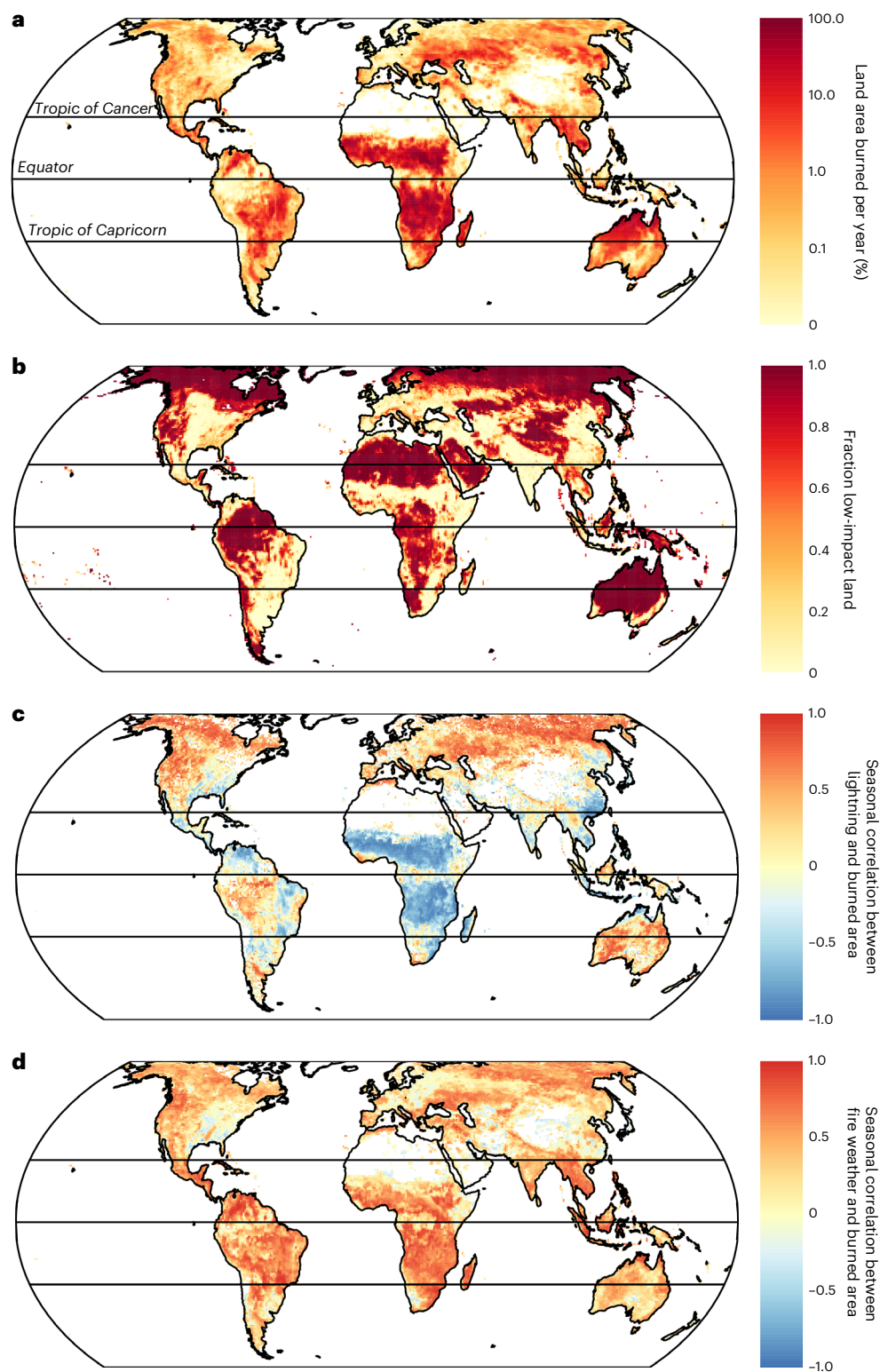
Fig. 1 | Global burned area and fire-cause predictors. a, The global burned land area frequency, here calculated as the mean annual area burned between 2001 and 2020[37] divided by the land area per 0.5° grid cell. **b**, The percentage low-impact land per 0.5° grid cell[34]. **c**, The correlation between the climatological mean monthly lightning density[19] and burned area. **d**, The correlation between the climatological mean monthly FWI and burned area.

continued limitations to vegetation productivity through decline in rainfall[2], lightning fire activity in boreal regions may increase due to interacting trends induced by climate change, including increases in dry fuel availability and lightning activity[13,17]. Differentiating between anthropogenic and lightning fires is essential to improving our understanding of trends in global fire activity, burned area, intensity and behaviour and for enabling accurate representation of fires in global land-surface models. In this Article, we identify the differing temporal and spatial niches of anthropogenic and lightning fires and construct statistical models to attribute the dominance of
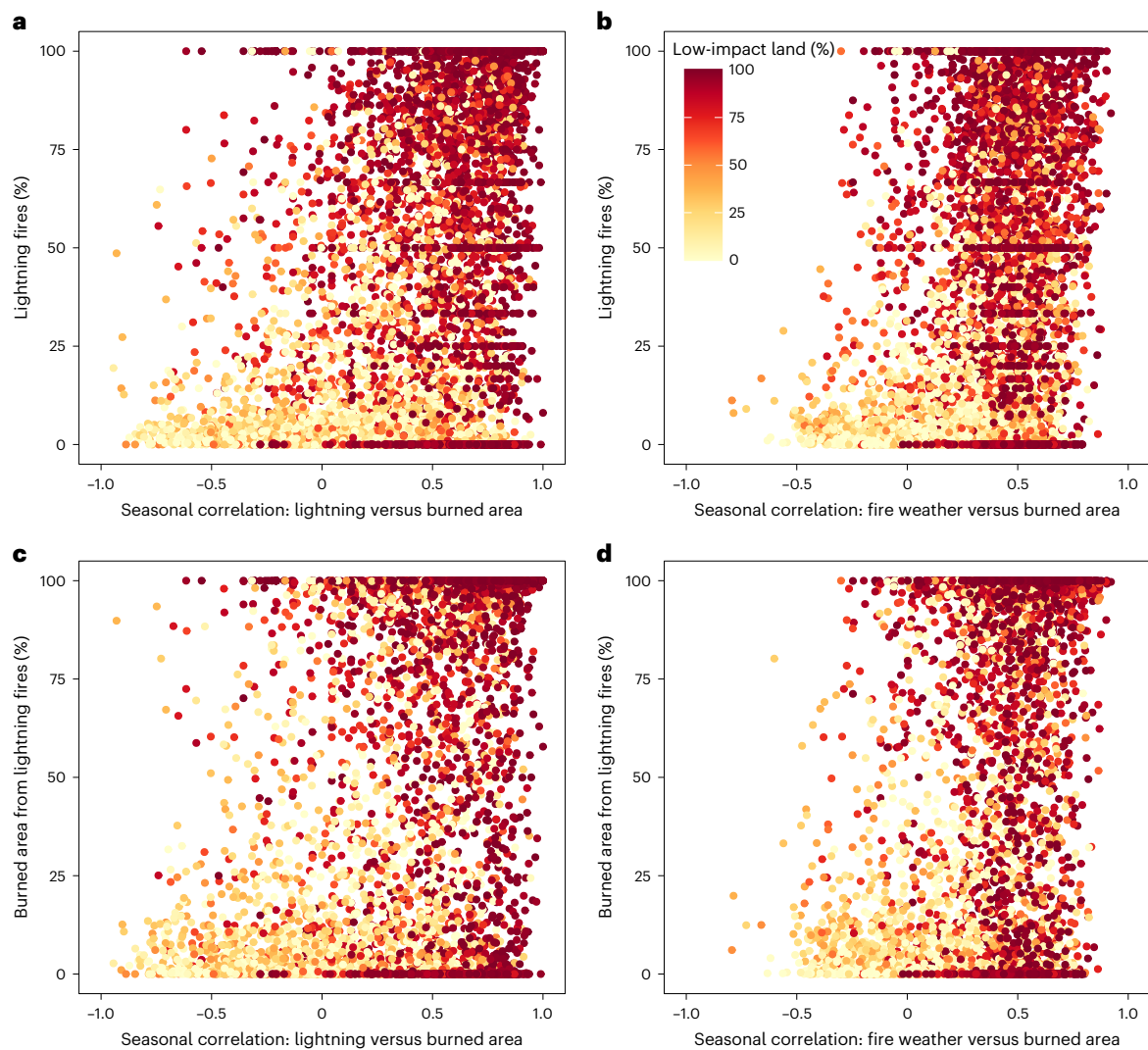
Fig. 2 | Relationships between the predictors and causes of fire and burned area. a–d, Lightning fires (**a**,**b**) and the burned area attributed to lightning (**c**,**d**) tend to be dominant in areas with low human impact. Furthermore, lightning fires (**a**) and the burned area from lightning (**c**) also tend to have a high seasonal correlation between lightning and burned area. Finally, human-dominated fire regimes are found in regions with a low seasonal correlation between fire weather and burned area (**b**,**d**). Data points represent 0.5° geospatial grid cells obtained from reference data from seven world regions (Extended Data Fig. 1).

lightning or anthropogenic ignition sources evident in global patterns of fire activity and burned area. Specifically, we used the XGBoost machine-learning algorithm to predict the fractions of fires related to lightning and anthropogenic ignition sources at 0.5° resolution and separately to predict the area burned by those fires (Methods). The predictors used in the model were the fractional cover of low-impact land (a measure of remoteness), the correlation between monthly lightning and burned-area climatologies (a measure of temporal coincidence between potential lightning ignition sources and fire) and the correlation between monthly fire-weather index (FWI) and burned-area climatologies (a measure of the temporal coincidence between fuel dryness and fire; Methods). For model training, we assembled a dataset of more than 1.2 million records of individual fires and their causes spanning seven extratropical regions, using a weighted sampling to ensure that samples from North America did not dominate the training dataset (Methods, Extended Data Fig. 1 and Supplementary Table 1). We used our model to identify the regions most at risk of increased lightning fire activity with climate change and evaluated the sensitivity of lightning to climate change using projections from climate models.

Lightning fires dominate in extratropical forests

Our model reproduced large portions of the spatial variability present in the reference data (Figs. 2 and 3). Indeed, the two statistical models explained on average 53% of the spatial variability in fraction of lightning fires and 39% of the spatial variability in the fraction of burned area from lightning (Extended Data Fig. 4). Furthermore, the models performed well in discriminating lightning- and human-dominated fire regimes in the independent test data and were robust against different splits of model training and validation datasets (Extended Data Fig. 1, Supplementary Text 1 and Supplementary Table 2). The fraction of low-impact land showed the highest feature importance in both Extreme Gradient Boosting (XGBoost) models. The seasonal correlation between lightning and burned area had the second-highest feature importance, while the correlation between fire weather and burned area had the lowest feature importance yet still positively contributed to the overall model performance (Supplementary Text 1 and Extended Data Fig. 3).

Our results show that, on a global scale, lightning is the primary ignition source of fires in temperate and boreal forests and confirm that anthropogenic fires dominate savannahs and agricultural lands
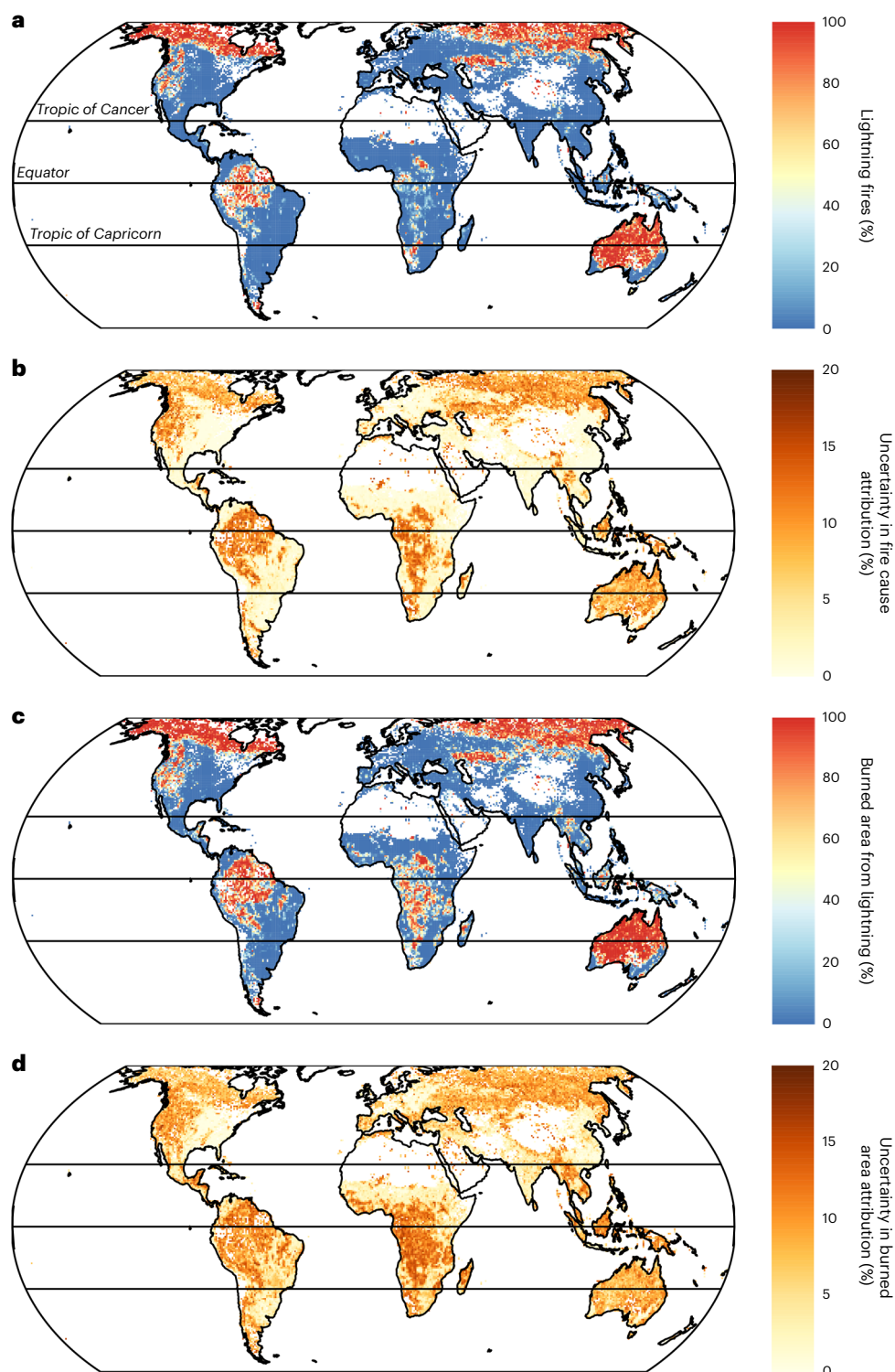
Fig. 3 | Global attribution of lightning and anthropogenic fires, their burned area and uncertainties. a, The estimated fraction of lightning fires from our model. **b**, The uncertainty, expressed as the standard deviation of the model prediction, in the fraction of lightning fires. **c**, The estimated fraction of burned area attributed to lightning fires. **d**, The uncertainty, expressed as the standard deviation of the model prediction, in fraction of burned area attributed to lightning fires.

(Fig. 3). Here we focus on intact-forest landscapes, defined as 0.5° spatial grid cells with at least 1% intact-forest[8] cover (Fig. 4a). We estimate that in extratropical intact-forest landscapes, 77% (spatial standard deviation weighted by burned area = 34%, standard deviation of uncertainty = 8%) of the burned area stems from lightning, much larger than the estimates for tropical intact forests or global estimates (Fig. 5a). Extratropical intact

forests currently experience a gross fire-related forest loss rate of 0.25% of their extent per year (spatial standard deviation weighted by land area = 0.45% per year), which is almost triple that of the fire-related forest loss rates observed in intact tropical forests or the global average (Fig. 5b).

The temperate and boreal forests contain the largest area of remaining intact forest on Earth (~53%, or 6.8 million km²), larger
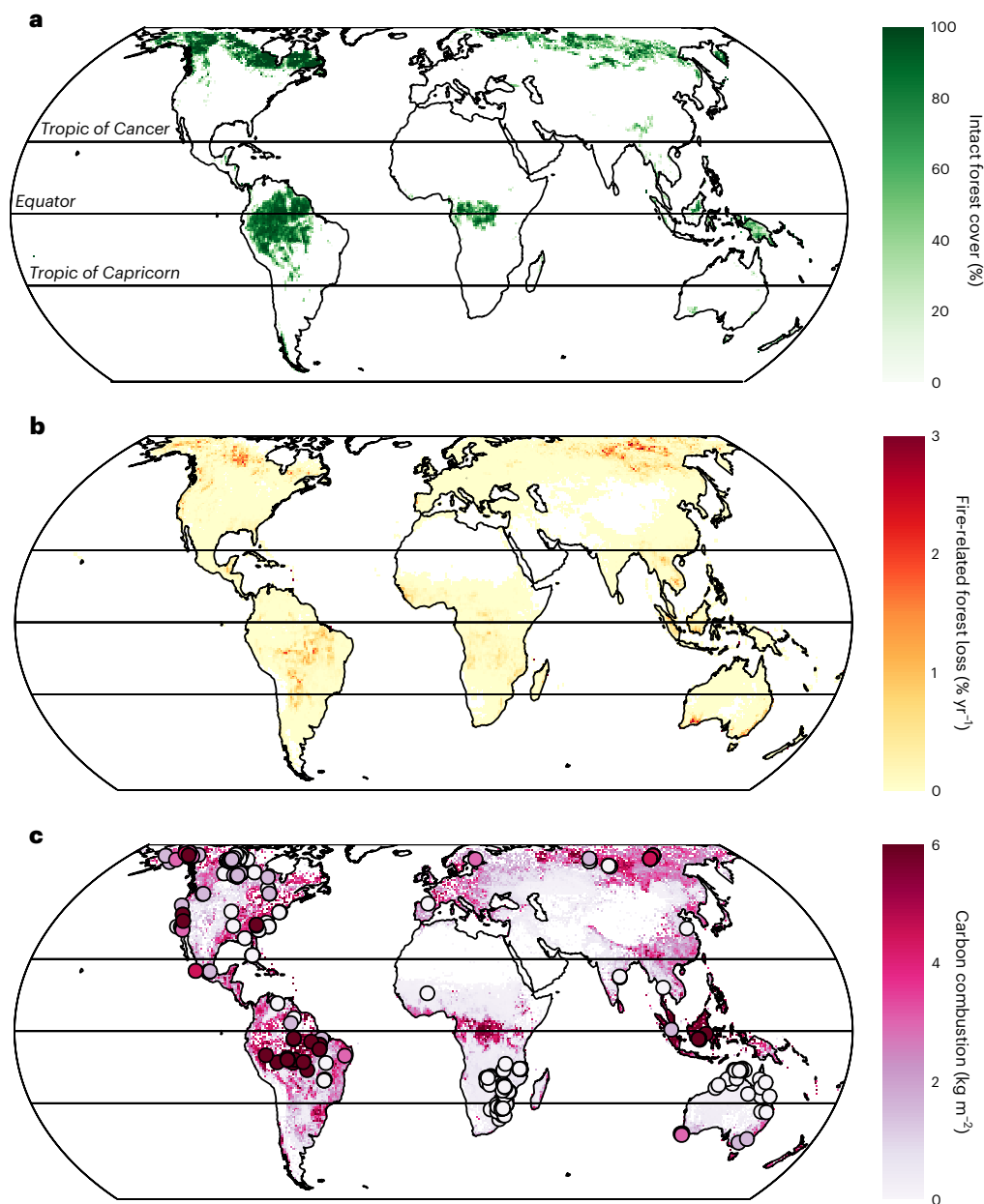
Fig. 4 | Lightning fires coincide geographically with areas of extratropical intact forests, high fire-related forest loss rates and high carbon combustion.
a, The distribution of intact forests[8]. **b**, Fire-related forest loss rates[20]. **c**, Carbon combustion from fire[21]. The locations shown in **c** show the sites included in the carbon combustion database of ref. 21.

than the area covered by intact tropical forest (6.1 million km²; Fig. 4a)[8]. In addition, the temperate and boreal forests are currently also experiencing high fire-related forest loss rates (Fig. 4b)[20] and are characterized by some of the largest carbon combustion rates per unit area on Earth (Fig. 4c)[21]. Extratropical intact-forest landscapes accounted for 1.2% of the global burned area between 2002 and 2020 yet emitted 8.5% of the global carbon emissions from fires[21]. The geographical overlap among lightning-dominated fire regimes, high C combustion and prevalence of intact forest demonstrates the vulnerability of extratropical intact forests to increases in lightning fires (Fig. 5).

On the basis of a global state-of-the-art model of fire fuel combustion and emissions[21], we calculated that fires in extratropical intact forests combust on average 2.87 kgC m⁻² (spatial standard deviation weighted by burned area = 1.31 kgC m⁻²), which is almost double the

average combustion rate of fires in tropical forests (1.63 ± 1.67 kgC m⁻²) and more than five times the global average (0.56 ± 0.78 kgC m⁻²; Fig. 5c). By comparison, field-derived C combustion measurements in extratropical intact forests have a mean combustion rate of 4.95 kgC m⁻² (standard deviation: 7.03 kgC m⁻²) (Extended Data Fig. 5), suggesting that the global model might underestimate the actual C combustion from these extratropical forest ecosystems.

Our model predicted with high certainty that humans are the dominant source of fire ignitions in open savannah regions (Fig. 3), in line with expectations for these regions[1]. By contrast, the more remote and densely vegetated savannahs and tropical forests of Central Africa, the Amazon basin, Borneo and New Guinea occasionally show lightning-dominated or mixed fire regimes, although with considerable uncertainties (Fig. 3). Fires in these inner tropical rain forests are extremely rare (Fig. 1a), and although it has been assumed that lightning

plays a limited role in the fire regime, there has been little direct quantitative evidence with which to test this assumption[22]. While our model already attaches a large uncertainty to the attribution of dominant fire cause in tropical forests, we reiterate that our training dataset did not include records from these regions (Extended Data Fig. 1), which further weakens our confidence in the attribution of lightning as ignition source in tropical forests. Therefore, we interpret these results as an indicator that lightning ignitions may be more important than previously thought in tropical forests and that regionally focused work involving the study of individual fires is required to clarify the relevance of lightning ignitions in these regions.

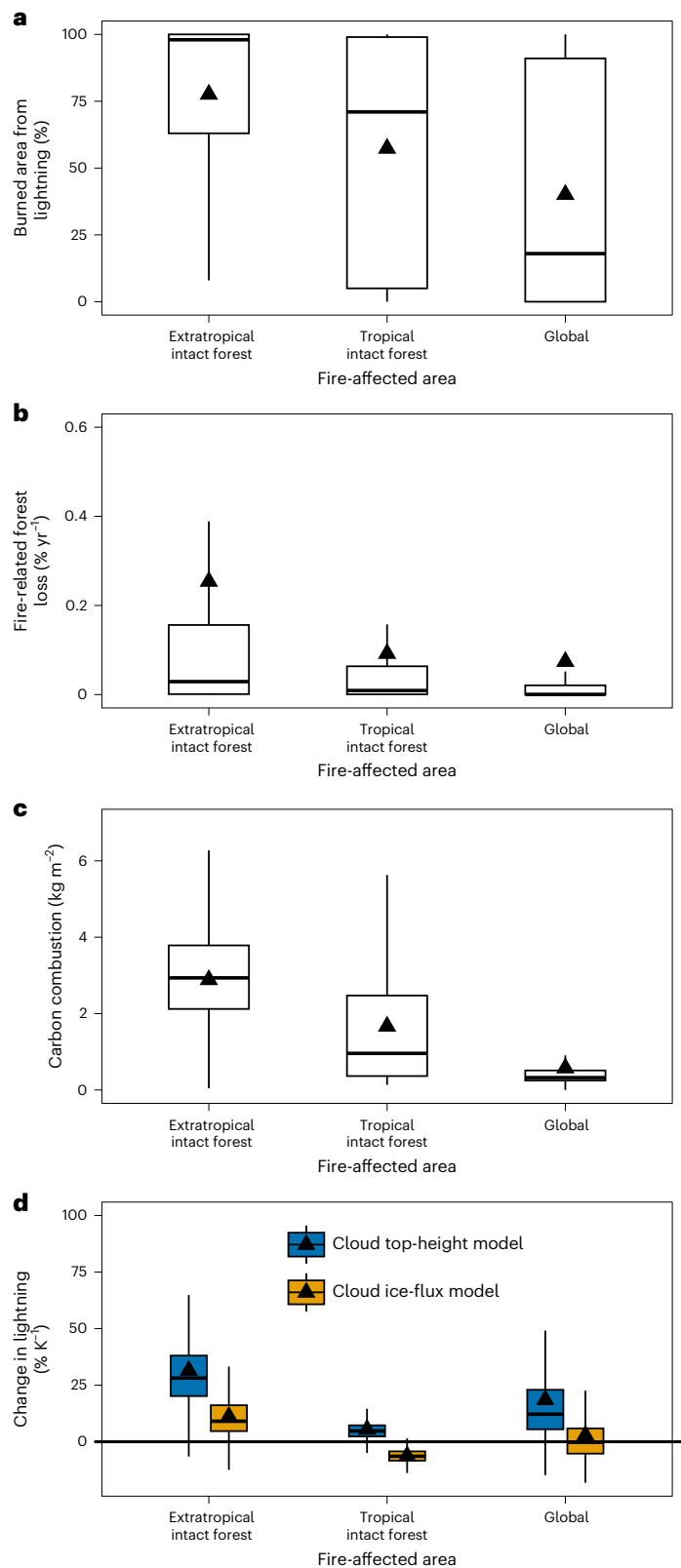## Increases in lightning in extratropical forests

The temperate and boreal forests, where lightning fires dominate, coincide geographically with expected increases in lightning because of climate change. We explored the risk of increases in lightning using two lightning modelling approaches, one based on the cloud-top height (Fig. 6a) and another based on the cloud ice flux (Fig. 6b). Per degree warming, the lightning model based on the cloud-top height approach estimates a 31.1% K^{-1} (spatial standard deviation weighted by grid cell area = 17.2% K^{-1}) increase in lightning in extratropical intact-forest landscapes, whereas the model based on the cloud ice-flux approach projects a more modest increase of 10.8% K^{-1} (spatial standard deviation weighted by grid cell area = 10.2% K^{-1}). For the extratropical intact forests north of 55° N, the multimodel mean estimated increase in lightning per degree warming from ref. 23 of 27.8% K^{-1} falls between the estimates from the cloud ice-flux approach (14.6% K^{-1}) and the cloud-top height approach (34.8% K^{-1}). For both approaches, the lightning increases over extratropical intact forests are larger than the global average, whereas over tropical intact forests, future lightning projections diverge in sign (Fig. 5d). These results suggest that in extratropical intact-forest landscapes, lightning fires not only are most prevalent but also are expected to become more frequent in the future, with a potential increase in fire-induced C emissions from these ecosystems.

## Climate–lightning–fire feedbacks in extratropical forests

Intact extratropical forest landscapes store large amounts of C in tree biomass and below-ground[8]. An intensification of the lightning fire regime in extratropical intact forests has the potential to accelerate forest degradation and forest loss. By reducing forest cover, an increase in lightning fires results in a direct decline of the terrestrial C stock. Since the mid-1970s, lightning fire activity has increased markedly in Alaska and parts of northwestern Canada[13]. In boreal regions, a portion of the C lost to increased fire is legacy C[24] that has escaped the previous fire cycle. In other words, legacy C is C that had accumulated in organic soils before the last fire cycle. In this way, legacy C emissions represent an accelerating (warming) carbon cycle–climate feedback. Furthermore, approximately 91% of the areas where extratropical intact

forests prevail in the Northern Hemisphere are underlain by permafrost. Natali et al.[25] estimated that fires may amplify emissions of greenhouse gases from permafrost thaw by 30% by the end of the century following a moderate emissions scenario. Greenhouse gas emissions after fires in permafrost peatlands may thus represent a range of positive warming feedbacks that results from increases in lightning fires.

The two lightning climate proxies capture key physical components of lightning generation and represent the dominant

Fig. 5 | Extratropical intact forests, unlike other parts of the world, experience large burned area from lightning, are characterized by high fire-related forest loss rates and carbon combustion and are expected to see an increase in lightning with climate warming. a, Burned area from lightning in extratropical intact forests is considerably higher than in the rest of the world. **b,c**, The fire-related forest loss rate (**b**) and carbon combustion (**c**) are also higher in extratropical intact forests compared with both tropical intact forests and the global average. **d**, Both lightning models predict a clear increase in lightning over extratropical intact forest, in contrast to tropical intact forests and the global average. All data have been weighted by burned area (**a,c**) or land area (**b,d**). The black triangles represent the weighted average. The black line within each box denotes the median, the limits of the box represent the 25th (Q_1) and 75th (Q_3) percentiles and the lines extending above and below denote Q_1 and Q_3 minus or plus 1.5 times the interquartile range ($Q_3 - Q_1$), respectively. The sample size (*n*) in each geographical region is greater than 5,000.
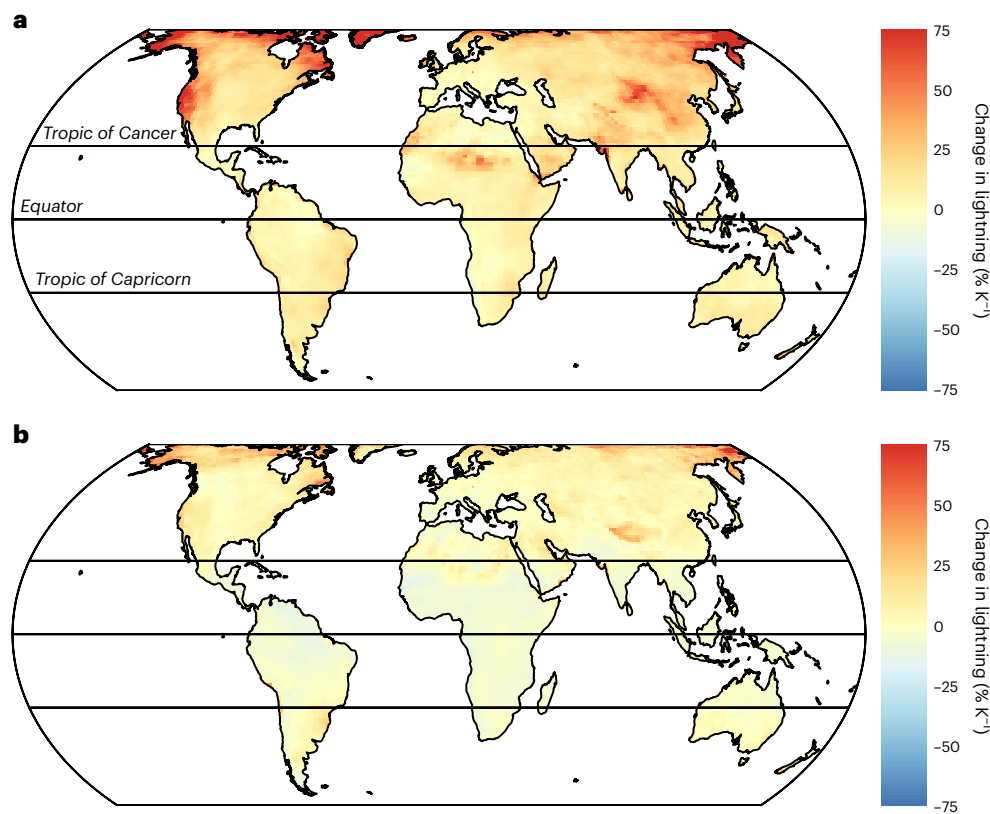
Fig. 6 | Lightning fires coincide geographically with areas that show expected increases in lightning with climate warming. a,b, Change in lightning per degree warming following the cloud-top height approach (**a**) and following the cloud ice-flux approach[26] (**b**).

parameterization-based uncertainty in future lightning projections. Their consistency in projection in boreal regions provides robust evidence for increasing lightning in extratropical forests with climate change. Previously, the cloud-top height and cloud ice-flux models have been used to demonstrate uncertainty in whether global and tropical lightning will increase or decrease in response to climate change[26]. However, there is less uncertainty for projections over the northern extratropics, where increases are consistently projected in both models (Figs. 5d and 6). For the northern land areas, the projections presented here (Figs. 5d and 6) are consistent with projected increases in lightning based on another climatological proxy, the multiplication of precipitation and convective available potential energy[23]. The increase of lightning occurrence in boreal intact forests, especially in combination with fuel drying[17], will probably result in a higher potential for lightning fires in boreal forests and associated increases of fire-induced C emissions.

Our approach does not account for the effect that changing socioeconomic drivers might have on the potential climate-driven increase of lightning fires in the future. Improvements in communication technology, early-warning systems and accessibility might aid suppression of lightning fires in extratropical forests[27]. Conversely, expected increases in land abandonment in rural areas might lead to a build-up of fuel conducive to lightning and anthropogenic ignitions and fire spread, which is, for example, already observed in Mediterranean mountain regions[28,29]. Furthermore, improved accessibility to extratropical intact forests could also increase their vulnerability to anthropogenic ignitions[20].

Towards a fire-cause attribution of individual fires

In this study, we attributed global fires and their burned-area patterns to anthropogenic and lightning causes on a relatively coarse spatial scale of 0.5° by using reference data and statistical models. While our approach provides a coarse global fire-cause attribution, it does not resolve ignition causes in a spatially and temporally explicit manner at the scale of individual fires. Recent efforts have attributed individual ignitions to anthropogenic or lightning causes by combining high-resolution fire ignition data from remote sensing with lightning strike observations and infrastructure data[17,18,30–33]. These studies have matched individual lightning strikes to fire ignition points and demonstrated the importance of dry lightning for ignition in boreal forest ecosystems[17,18].

Recent studies have also sought to identify the ignition sources of individual fires in the tropics from observations of individual lightning and fire ignition points[31,32]. While these studies have shown that most ignitions in savannah and tropical forests are caused by people, they also indicate that lightning regularly causes fire during the transition from the dry to the wet season in some of Brazil's savannah and tropical forest ecosystems[31,32]. Our models in part corroborate the findings of regional work by predicting a mixture of lightning and anthropogenic fires within Brazilian biomes and the global tropical forests more generally (Fig. 3). Nonetheless, the high uncertainty in our model prediction of lightning fire in tropical forests highlights the possibility that the contribution of lightning to fire activity and burned area is overpredicted in our model. A possible explanation for this could be that the data layer that we used to quantify human impact on the land[34] may be too coarse to accurately capture indigenous land use in tropical forests, while the alignment of the annual peaks in lightning activity and fire weather with the timing of fire ignitions by people may lead to the misidentification of human fires as lightning (and vice versa). Further developments in spatially explicit mapping of fire cause for the entire Earth surface will become possible when high-resolution lightning and ignition data become globally available.

Relevance for global fire models

We show a sensitivity of extratropical intact forests to potential increases in lightning fires, which would have far-reaching consequences for terrestrial C storage and biodiversity. Our global fire-cause attribution between lightning and anthropogenic fires contrasts human-dominated fire regimes in savannahs and agricultural lands with lightning-dominated fire regimes in remote temperate and boreal forests. While changes in human ignitions, landscape fragmentation and increases in wet climatic conditions largely explain the decreasing trend in fire activity in savannahs[1,2], changes in lightning occurrence have the potential to increase lightning ignitions in C-rich temperate and boreal forests[9]. In these forests, increases in lightning occurrence are likely to further exacerbate the effects of increased flammability on fire occurrence and burned area[17,35]. This will be especially the case as peak lightning activity typically coincides with warm and dry conditions in summer months in extratropical regions[16]. Our global fire-cause attribution partly explains the contrasting observations of widespread declines in savannah fires and reported increases in fire activity in many temperate and boreal ecosystems[6,13,36]. We highlight the importance of discriminating between different causes and drivers of fire in Earth system models when projecting future trends of regional and global fire activity. This exercise is paramount to advancing our capabilities to model fire occurrence under changing socioeconomic and climatological conditions.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41561-023-01322-z.

References

1. Andela, N. et al. A human-driven decline in global burned area. *Science* **356**, 1356–1362 (2017).
2. Zubkova, M., Boschetti, L., Abatzoglou, J. T. & Giglio, L. Changes in fire activity in Africa from 2002 to 2016 and their potential drivers. *Geophys. Res. Lett.* **46**, 7643–7653 (2019).
3. Parks, S. A. & Abatzoglou, J. T. Warmer and drier fire seasons contribute to increases in area burned at high severity in western US forests from 1985 to 2017. *Geophys. Res. Lett.* **47**, e2020GL089858 (2020).
4. Canadell, J. G. et al. Multi-decadal increase of forest burned area in Australia is linked to climate change. *Nat. Commun.* **12**, 6921 (2021).
5. Talucci, A. C., Loranty, M. M. & Alexander, H. D. Siberian taiga and tundra fire regimes from 2001–2020. *Environ. Res. Lett.* **17**, 025001 (2022).
6. Hanes, C. C. et al. Fire-regime changes in Canada over the last half century. *Can. J. For. Res.* **49**, 256–269 (2019).
7. Chen, Y. et al. Multi-decadal trends and variability in burned area from the 5th version of the Global Fire Emissions Database (GFED5). Preprint at *Copernicus* https://doi.org/10.5194/essd-2023-182 (2023).
8. Potapov, P. et al. The last frontiers of wilderness: tracking loss of intact forest landscapes from 2000 to 2013. *Sci. Adv.* **3**, e1600821 (2017).
9. Zheng, B. et al. Increasing forest fire emissions despite the decline in global burned area. *Sci. Adv.* **7**, eabh2646 (2021).
10. Jones, M. W. et al. Global and regional trends and drivers of fire under climate change. *Rev. Geophys.* **60**, e2020RG000726 (2022).
11. Zheng, B. et al. Record-high CO_2 emissions from boreal fires in 2021. *Science* **379**, 912–917 (2023).
12. Short, K. C. A spatial database of wildfires in the United States, 1992–2011. *Earth Syst. Sci. Data* https://doi.org/10.5194/essd-6-1-2014 (2014).
13. Veraverbeke, S. et al. Lightning as a major driver of recent large fire years in North American boreal forests. *Nat. Clim. Change* **7**, 529–534 (2017).
14. Cattau, M. E., Wessman, C., Mahood, A. & Balch, J. K. Anthropogenic and lightning-started fires are becoming larger and more frequent over a longer season length in the USA. *Glob. Ecol. Biogeogr.* **29**, 668–681 (2020).
15. Hantson, S., Andela, N., Goulden, M. L. & Randerson, J. T. Human-ignited fires result in more extreme fire behavior and ecosystem impacts. *Nat. Commun.* **13**, 2717 (2022).
16. Abatzoglou, J. T., Kolden, C. A., Balch, J. K. & Bradley, B. A. Controls on interannual variability in lightning-caused fire activity in the western US. *Environ. Res. Lett.* **11**, 045005 (2016).
17. Hessilt, T. D. et al. Future increases in lightning ignition efficiency and wildfire occurrence expected from drier fuels in boreal forest ecosystems of western North America. *Environ. Res. Lett.* **17**, 054008 (2022).
18. Xu, W., Scholten, R. C., Hessilt, T. D., Liu, Y. & Veraverbeke, S. Overwintering fires rising in eastern Siberia. *Environ. Res. Lett.* **17**, 045005 (2022).
19. Balch, J. K. et al. Human-started wildfires expand the fire niche across the United States. *Proc. Natl Acad. Sci. USA* **114**, 2946–2951 (2017).
20. van Wees, D. et al. The role of fire in global forest loss dynamics. *Glob. Change Biol.* **27**, 2377–2391 (2021).
21. Van Wees, D. et al. Global biomass burning fuel consumption and emissions at 500 m spatial resolution based on the Global Fire Emissions Database (GFED). *Geosci. Model Dev.* **15**, 8411–8437 (2022).
22. van Marle, M. J. E. et al. Fire and deforestation dynamics in Amazonia (1973–2014). *Glob. Biogeochem. Cycles* **31**, 24–38 (2017).
23. Chen, Y. et al. Future increases in Arctic lightning and fire risk for permafrost carbon. *Nat. Clim. Change* **11**, 404–410 (2021). 2021 115.
24. Walker, X. J. et al. Increasing wildfires threaten historic carbon sink of boreal forest soils. *Nature* **572**, 520–523 (2019).
25. Natali, S. M. et al. Permafrost carbon feedbacks threaten global climate goals. *Proc. Natl Acad. Sci. USA* **118**, 2100163118 (2021).
26. Finney, D. L. et al. A projected decrease in lightning under climate change. *Nat. Clim. Change* **8**, 210–213 (2018). 2018 83.
27. Barmpoutis, P., Papaioannou, P., Dimitropoulos, K. & Grammalidis, N. A review on early forest fire detection systems using optical remote sensing. *Sensors* **20**, 6442 20 (2020).
28. Azevedo, J. C., Moreira, C., Castro, J. P. & Loureiro, C. in *Landscape Ecology in Forest Management and Conserv*ation (eds Li, C. et al.) 329–351 (Springer, 2011); https://doi.org/10.1007/978-3-642-12754-0_14
29. Mantero, G. et al. The influence of land abandonment on forest disturbance regimes: a global review. *Landsc. Ecol.* **35**, 2723–2744 (2020).
30. Scholten, R. C., Jandt, R., Miller, E. A., Rogers, B. M. & Veraverbeke, S. Overwintering fires in boreal forests. *Nature* **593**, 399–404 (2021).
31. Menezes, L. S. et al. Lightning patterns in the Pantanal: untangling natural and anthropogenic-induced wildfires. *Sci. Total Environ.* **820**, 153021 (2022).
32. Schumacher, V. et al. Characteristics of lightning-caused wildfires in central Brazil in relation to cloud–ground and dry lightning. *Agric. For. Meteorol.* **312**, 108723 (2022).
33. Xu, W. et al. Active fire dynamics in the Amazon: new perspectives from high-resolution satellite observations. *Geophys. Res. Lett.* https://doi.org/10.1029/2021gl093789 (2021).

34. Jacobson, A. P., Riggio, J., M. Tait, A. & Baillie, E. M. J. Global areas of low human impact ('low impact areas') and fragmentation of the natural world. *Sci. Rep.* **9**, 14179 (2019). 2019 91.

35. Abatzoglou, J. T., Williams, A. P. & Barbero, R. Global emergence of anthropogenic climate change in fire weather indices. *Geophys. Res. Lett.* **46**, 326–336 (2019).

36. Bowman, D. M. J. S. et al. Vegetation fires in the Anthropocene. *Nat. Rev. Earth Environ.* **1**, 500–515 (2020).

37. Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L. & Justice, C. O. The Collection 6 MODIS burned area mapping algorithm and product. *Remote Sens. Environ.* **217**, 72–85 (2018).

Methods

Fire-cause reference data

We acquired fire-cause reference data for seven different world regions: the United States, Canada, Portugal, Southern France, Victoria (Australia), Tasmania (Australia) and Yakutia (Russia) (Supplementary Table 1). These datasets contained data on the location, timing, size and cause of individual fires. We used data starting from 2001 onwards in correspondence with the Moderate-Resolution Imaging Spectroradiometer (MODIS) burned-area data used in our analysis. We aggregated all anthropogenic and lightning fires per 0.5° grid cell and calculated the fractions of anthropogenic and lightning fires and the fractions of burned area from anthropogenic and lightning fires (Extended Data Fig. 1).

For the United States, we downloaded fire-point data from the interagency Fire Program Analysis Fire-Occurrence Database version 4[12], which contains government records between 1992 and 2015. After removing fires with unknown causes, we retained 1,091,678 records between 2001 and 2015 for all American states excluding Alaska. For Alaska, we opted for the Alaska fire-history points database maintained by the Alaska Interagency Coordination Center because of its high-quality record in the satellite era[38]. After removing fires of unknown cause, we retained 8,810 records for Alaska between 2001 and 2019.

For Canada, we acquired fire-point data from the Canadian National Fire Database[39]. After removing fires with unknown causes, we retained 117,947 records between 2001 and 2018. For Portugal, we received the governmental fire-point data record from the Instituto da Conservação da Natureza[40]. After removing all fires with unknown cause, we retained 9,018 records between 2001 and 2020. For Southern France, we downloaded the government records from the Prometheus fire database[41]. After removing fires with unknown cause, we retained 13,559 records between 2001 and 2020. For Victoria, Australia, we received the government records from the Forest Fire Management Victoria agency. After removing fires with unknown cause, we retained 16,593 records between 2001 and 2020. For Tasmania, Australia, we downloaded fire-history data from Tasmania Fire Service and Parks and Wildlife Service[42]. After removing fires with unknown cause, we retained 3,959 records between 2001 and 2020.

For Yakutia, Russia, individual fire starts and fires between 2012 and 2020 were mapped by combining the MODIS MCD64A1 burned-area product and the Visible Infrared Imaging Radiometer Suite VNP14IMG active-fire product[18]. Fire starts were attributed to lightning, human, overwintering and unknown causes using a set of spatial and temporal rules[18]. After removing fire starts from an unknown cause and overwintering cause, which represented less than 5% of the fire starts, we retained 2,525 records between 2012 and 2020.

Lightning data

We used the contemporary lightning climatology at 0.5° from the Optical Transient Detector (OTD) onboard Orbview-1 and the Lightning Imaging Sensor (LIS) onboard the Tropical Rainfall Measuring Mission version 2.3.2015[43]. This lightning climatology product includes data from the OTD sensor between 1995 and 2000 acquired from a near-polar orbit, while the climatology included LIS data for the regions between 38° N and 38° S between 1998 and 2014. The overlap between the LIS and OTD datasets between 38° N and 38° S makes the merged product more robust for these areas[43]. The OTD and LIS sensors recorded lightning flashes, which include both cloud-to-cloud and cloud-to-ground lightning strikes. Flash rates from OTD correlated strongly with cloud-to-ground lightning strike data from the Alaskan Lightning Detection Network[23].

We used two different global projections of future lightning that are based on two different approaches to simulate lightning: the cloud-top height and cloud ice-flux approaches[26]. The cloud-top height approach is widely used and rooted in the positive associations

among convective updraughts, thunderstorms and lightning[44]. The cloud ice-flux approach, however, accounts for cloud ice microphysics since the formation of ice and graupel particles is important for thunderstorm charging and associated lightning activity[26]. The cloud-top height and cloud ice-flux approaches lead to diverging future projections of lightning, especially in the low latitudes[26]. The cloud-top height approach predicts increases in lightning over most of the global land area, whereas the cloud ice-flux approach predicts decreases over most of the tropical land area and no noteworthy changes to modest increases in lightning over extratropical land. Simulations of both approaches were performed using the UK Chemistry and Aerosols model coupled to the atmosphere-only version of the UK Met Office Unified Model version 8.4. The atmosphere component is the Global Atmosphere 4.0. More details about the model set-up can be found in ref. 26. For both approaches, simulations were performed representing the years 2000 and 2100 under the representative concentration pathway 8.5. For the land areas north of 55° N, we further compared the results from the cloud-top height and cloud ice-flux approaches with a third approach that uses the multiplication of the convective available potential energy (CAPE) and precipitation as a proxy for lightning[23]. We compared the changes in lightning from the cloud-top height and cloud ice-flux approaches with the mean change in lightning from the CAPE × precipitation approach calculated from 15 global climate models. For all three approaches, we expressed the change in lightning per degree warming by assuming a linear relationship[45] (Fig. 6).

Burned-area data

We used burned-area data from the MODIS MCD64A1 Collection 6 burned-area product at 500 m resolution[37] between 2001 and 2020. The MCD64A1 algorithm classifies burned area on the basis of fire-induced reflectance changes and thermal anomalies. We aggregated the burned area in 0.5° grid cells and calculated the monthly burned-area climatology for the years between 2001 and 2020.

Low-impact land data

We used the global areas of low human impact at 1 km resolution dataset[34] to characterize low-impact land. In this dataset, 1 km² pixels were classified into pixels with and without human impact by assimilating contemporary datasets on human population, livestock density, forest-cover change, land cover and night-time lights[34]. We calculated the fraction of low-impact land per 0.5° grid cell as the ratio between the number of land pixels without human impact and the total number of land pixels (Fig. 1b).

Intact-forest data

We used the delineation of intact-forest landscapes for the year 2000[8]. Intact-forest landscapes are mosaics of forest and naturally treeless ecosystems with no remotely detected signs of human activity. The intact-forest landscape polygons are based on the vectorization of contiguous pixels with Landsat-derived tree cover higher than 20% and including water bodies and naturally treeless ecosystems. In addition, intact-forest landscape polygons have a minimum size of 500 km², a minimum width of 10 km and a minimum corridor width of 2 km (ref. 8). In areas with intact-forest polygons of 500 km², smaller patches of intact forest often exist, yet these are not included in the dataset. We calculated the fraction of intact forest per 0.5° grid cell as the ratio between the area of the intact-forest landscapes and the land area per grid cell (Fig. 4a). We used the Northern Hemisphere map from ref. 46 to discriminate intact forest underlain by permafrost in the northern extratropics. We thereto converted the vectors to a 0.5° resolution raster on the basis of areal class majority.

Fire-related forest loss data

We used fire-related forest loss estimates between 2003 and 2020 from ref. 20. This methodology estimates gross fire-related forest losses

as it does not account for post-fire recovery. We derived fire-related forest losses at 500 m resolution by combining forest loss data from the Global Forest Change version 1.8 dataset[47] with the MODIS MCD64A1 burned-area[37] and MCD14ML active-fires[48] products. In short, after regridding the forest loss data onto a 500 m grid, fire-related forest losses were defined as spatiotemporal matches between forest loss and the occurrence of burned area or active fires. To account for delayed post-fire tree mortality, forest losses one year after the spatial match with a burned area or active-fire detection were also included in the fire-related forest loss estimate. We developed two scenarios for estimating fire-related forest loss. The first scenario assumed that all forest losses within a burned-area pixel or active-fire footprint were fire-related forest losses. A second, minimum probability, scenario accounted for the geolocational uncertainty of fire within active-fire pixels and the active-fire detection confidence. We then calculated the average of the two scenarios as our best estimate of fire-related forest loss, after validation with higher-resolution products. The fire-related forest losses were aggregated in 0.5° grid cells over the period between 2003 and 2020 (Fig. 4b).

Carbon combustion data
We used C combustion, or area-normalized C emissions (in kgC m^{-2} burned), between 2002 and 2020 from the state-of-the-art 500 m global biomass burning fuel consumption database, which will also be the basis for the Global Fire Emissions Database version 5[21]. These estimates are based on a simplified version of the Carnegie–Ames–Stanford Approach model and assimilate several meteorological and remote-sensing datasets to estimate C combustion. The modelling framework is optimized and constrained using field-measured C combustion estimates[21]. We aggregated the C combustion data in 0.5° grid cells and calculated the mean over the period between 2002 and 2020 and weighted for the burned area in the grid cell over the same period (Fig. 4c). The field measurements of carbon combustion from extratropical intact forests were also retrieved from the synthesis field measurement database of ref. 21 (Extended Data Fig. 5).

FWI data
We obtained the global 0.25° daily FWI from the fire-danger indices historical data from the Copernicus Emergency Management Service dataset[49]. The FWI is a unitless index derived from temperature, relative humidity, precipitation and wind speed and is considered a good indicator of dry and hot weather conditions favourable for fire ignition and fire spread. The data were spatially aggregated to 0.5° and temporally averaged to monthly values for the period 2001–2020 to match the spatial and temporal resolution of the burned-area data in our study.

Fire-cause attribution models
The collated fire-cause attribution datasets consisted of 7,766 0.5° grid cells (Extended Data Fig. 1) with values ranging between 0%, indicating all fires and their associated burned area in that grid cell are ignited by humans, and 100%, indicating all fires are ignited by lightning. We used these two datasets, the fraction of lightning fires and fraction of burned area from lightning, in conjunction with three geospatial predictors for which we had global coverage (Fig. 1b–d), to estimate anthropogenic and lightning fires globally. The first predictor variable was the fraction of low-impact land. The rationale behind using the fraction of low-impact land as a fire-cause predictor is that in areas with low human impact, lightning fires are more likely, while in areas with high human impact, anthropogenic fires are more likely (Extended Data Fig. 3c). The second predictor was the Spearman correlation between the monthly climatologies of lightning and burned area. We used the non-parametric Spearman correlation to account for nonlinear relationships that occurred between the lightning and burned-area climatologies. The rationale behind using the seasonal correlation between lightning and burned area as a fire-cause predictor

is that fires within grid cells with a high correlation are more likely to originate from lightning, while anthropogenic fires are more likely when there is a strong anti-correlation between the monthly lightning and burned-area climatologies. The third predictor was the Spearman correlation between the monthly FWI and burned area. The reasoning behind including the co-occurrence of dry weather conditions, favourable for fire ignition and spread, and burned area is that a strong correlation implies more natural fire regimes with a higher probability of lightning fires while the lack of a correlation or even anti-correlation implies a human-dominated fire regime, in which fires are more easily ignited, independent of weather conditions.

We built two separate XGBoost model routines[50]. One model predicted the fraction of lightning fires in each grid cell, and the other model predicted the fraction of burned area from lightning fires in each grid cell. Our reference data stemmed primarily from North America, and we therefore implemented a region-specific training data selection to geographically balance the model. To do this, we calculated for each region the number of samples that should be taken to have a sample proportional to the burned area of that region between 2001 and 2020 (Supplementary Table 1). Subsequently, we selected from the reference data of each region 60% of this sample number as training and validation data, with replacement, and used all remaining data as test data. This resulted, for example, in 11% less training data from the North American region and 6% more training data from Yakutia, compared with a random 60% draw, accounting for the disparity in data availability between those regions. We tested the robustness of the two models using different training–test data splits (20%, 40%, 60% and 80%) and selected the 60% for training as it showed a good balance between internal model validation accuracy and high performance on the test dataset (Supplementary Table 2).

Our reference dataset showed a strong bimodal distribution, with most grid cells showing either a completely human-dominated fire regime or a completely lightning-dominated fire regime (Extended Data Fig. 2). When using regression objectives in the XGBoost model, or any other machine-learning algorithm, this specific bimodal distribution cannot be reproduced. As the variability in the data is never fully captured by the model, the model aims to minimize the error by predicting the more uncertain cases close to the overall mean (Extended Data Fig. 2). This results in a normal frequency distribution of the predicted grid cells around the mean (~56% of burned area from lightning and ~52% lightning fires), which does not resemble the frequency distribution in the reference data. To account for this, we classified the training data in human-dominated (<10% lightning fires) and lightning-dominated (>90%) fire regimes and omitted all other training data. These selections included 61% of the data for fire counts and 79% of the data for burned area. Subsequently, 100 individual XGBoost models were trained with a logistic regression objective using random subsets of 80% of this reclassified training data, and the remaining 20% was used for validation. The predicted continuous outcomes between 0 and 100% of all models were individually assessed using the validation data (Supplementary Table 2 and Extended Data Fig. 4). Hereafter, each of the 100 continuous global model predictions were classified using a threshold into lightning fires (≥50%) or human fires (<50%). The average of these 100 binary model predictions showed a very similar frequency distribution as the reference data (Extended Data Fig. 2) and a good performance on the test data (Supplementary Table 2) and was therefore used as the final model output (Fig. 3a,c). The standard deviation of the 100 continuous model predictions was used as the model uncertainty (Fig. 3b,d).

Geospatial analysis
We discriminated between extratropical intact forest, tropical intact forest and all global land with burned area. Grid cells with intact forest were selected as soon as their fraction of intact forest exceeded zero. We then calculated the area-weighted mean and standard deviation

for the fraction of burned area from lightning, fire-related forest loss, C combustion and change in lightning for extratropical intact forest, tropical intact forest and globally (Fig. 5). The fraction of burned area and C combustion were weighted by burned area. Fire-related forest loss and the change in lightning were weighted by land area.

Data availability

All data used and produced in this paper are hosted at the Pangaea public repository at https://doi.org/10.1594/PANGAEA.939352.

Code availability

The code for this analysis is available from the corresponding author upon request.

References

38. Kasischke, E. S., Williams, D. & Barry, D. Analysis of the patterns of large fires in the boreal forest region of Alaska. *Int. J. Wildl. Fire* **11**, 131–144 (2002).

39. Stocks, B. J. et al. Large forest fires in Canada, 1959–1997. *J. Geophys. Res. Atmos*. https://doi.org/10.1029/2001JD000484 (2002).

40. Fernandes, P. M., Santos, J. A., Castedo-dorado, F. & Almeida, R. Fire from the sky in the Anthropocene. *Fire* **4**, 13 (2021).

41. Ganteaume, A. & Guerra, F. Explaining the spatio-seasonal variation of fires by their causes: the case of southeastern France. *Appl. Geogr.* **90**, 69–81 (2018).

42. Nampak, H. et al. Characterizing spatial and temporal variability of lightning activity associated with wildfire over Tasmania, Australia. *Fire* **4**, 10 (2021).

43. Cecil, D. J., Buechler, D. E. & Blakeslee, R. J. Gridded lightning climatology from TRMM-LIS and OTD: dataset description. *Atmos. Res.* **135–136**, 404–414 (2014).

44. Price, C. & Rind, D. A simple lightning parameterization for calculating global lightning distributions. *J. Geophys. Res. Atmos.* **97**, 9919–9933 (1992).

45. Finney, D. L., Doherty, R. M., Wild, O., Young, P. J. & Butler, A. Response of lightning NO_x emissions and ozone production to climate change: insights from the Atmospheric Chemistry and Climate Model Intercomparison Project. *Geophys. Res. Lett.* **43**, 5492–5500 (2016).

46. Obu, J. et al. Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km^2 scale. *Earth Sci. Rev.* **193**, 299–316 (2019).

47. Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).

48. Giglio, L., Schroeder, W. & Justice, C. O. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sens. Environ.* **178**, 31–41 (2016).

49. Vitolo, C. et al. ERA5-based global meteorological wildfire danger maps. *Sci. Data* **7**, 216 (2020).

50. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Vol. 13–17) 785–794 (ACM, 2016).

Author contributions

S.V., D.F. and M.W.J. designed the project. S.V. and T.A.J.J carried out the formal analysis, performed the validation and visualization of the results and wrote the first draft. All authors contributed to the methodology development, investigating of data sources and reviewing and editing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41561-023-01322-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41561-023-01322-z.

**Correspondence and requests for materials** should be addressed to Thomas A. J. Janssen.

Peer review information *Nature Geoscience* thanks Jatan Buch, Marco Turco and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Xujia Jiang, in collaboration with the *Nature Geoscience* team.

**Reprints and permissions information** is available at www.nature.com/reprints.
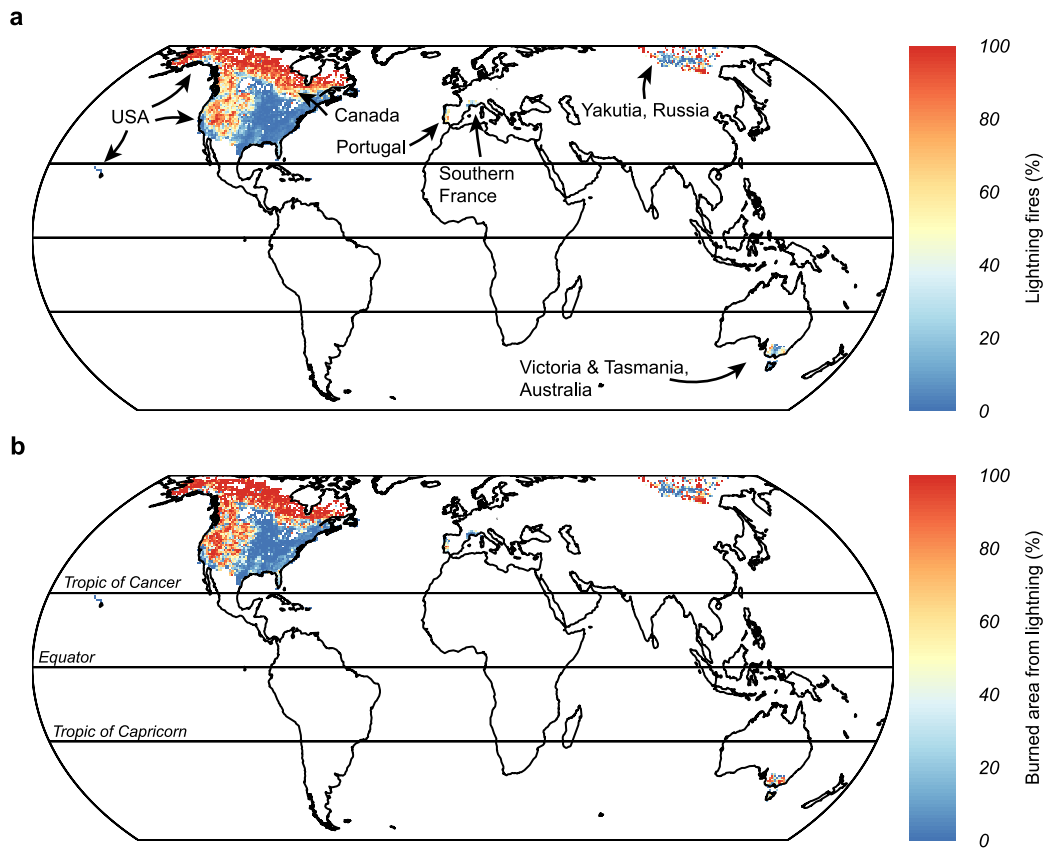
Extended Data Fig. 1 | Reference data for fire cause attribution. Reference data of a, the fraction of lightning fires of the total fire count, and b, the fraction of burned area from lightning of the total burned area. Data was available for seven different world regions: the United States of America (USA) including Alaska and Hawaii, Canada, Portugal, Southern France, Republic of Sakha (Yakutia) in Russia and the states of Victoria and Tasmania in Australia.
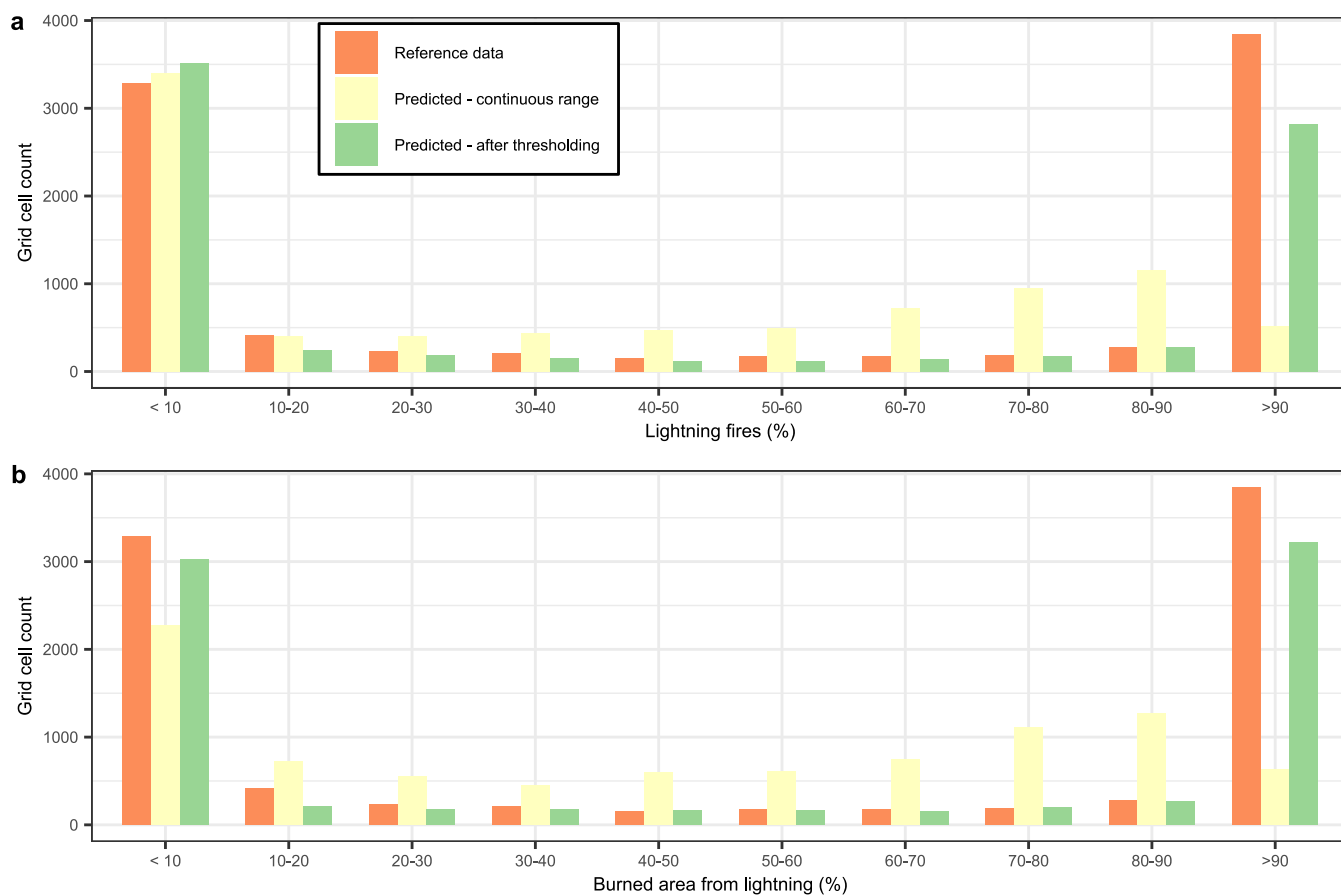
Extended Data Fig. 2 | Frequency distribution of lightning fires in the gridded reference data and model predictions. Both the fraction of lightning fires (**a**) as well as the fraction of burned area attributed to lightning fires (**b**) show a clear bimodal distribution in the reference data. Most grid cells in the reference data show either a lightning or human-dominated fire regime. When using regression objectives in the XGBoost model, such as the logistic regression objective used in this study, this specific bimodal distribution cannot be reproduced. To account for this, the predicted continuous range from the logistic regression was classified using a threshold into lightning fires (≥ 50%) or anthropogenic fires (< 50%) for the 100 model iterations. The frequency distribution from the average of these 100 predictions captured the observed frequency distribution in the reference data much better.
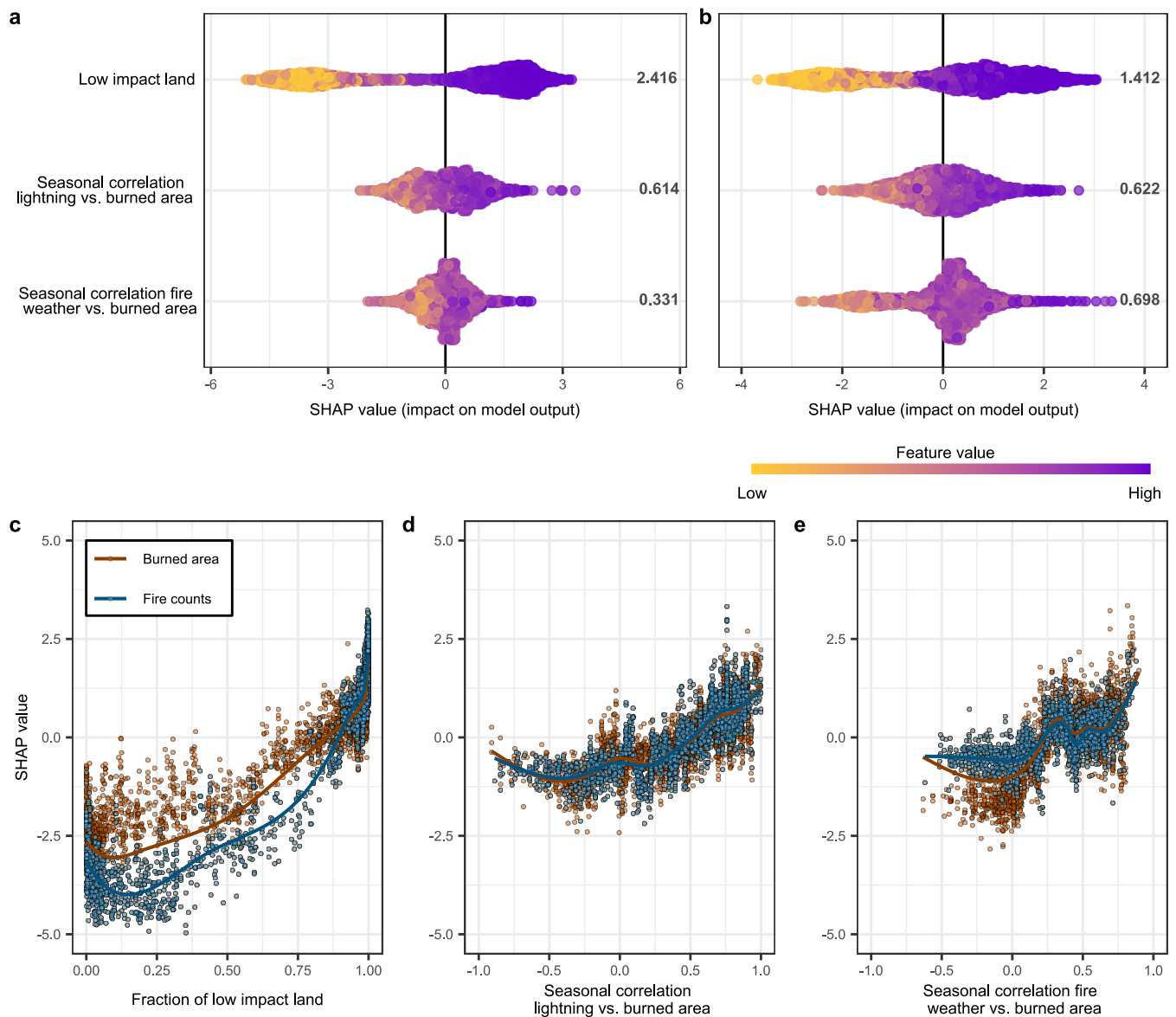
Extended Data Fig. 3 | Feature importance and the effect of the individual features on the model output. The summary plots for the SHapley Additive exPlanations (SHAP) analysis derived from the XGBoost models predicting (**a**) percentage of lightning fires and (**b**) the percentage of burned area attributed to lightning. A positive SHAP value indicates a positive impact on the model output (more lightning fires) while a negative SHAP value indicates a negative impact on the model output (more anthropogenic fires). The mean SHAP value indicates the feature importance, with a high SHAP value indicating a relatively large impact on the model output and a low value a relatively small impact on the model output. With increasing fraction of low impact land (**c**), increasing seasonal correlation between lightning and burned area (**d**) and increasing seasonal correlation between fire weather and burned area (**e**) the models predict a relatively higher fraction of lightning fires and burned area from lightning.
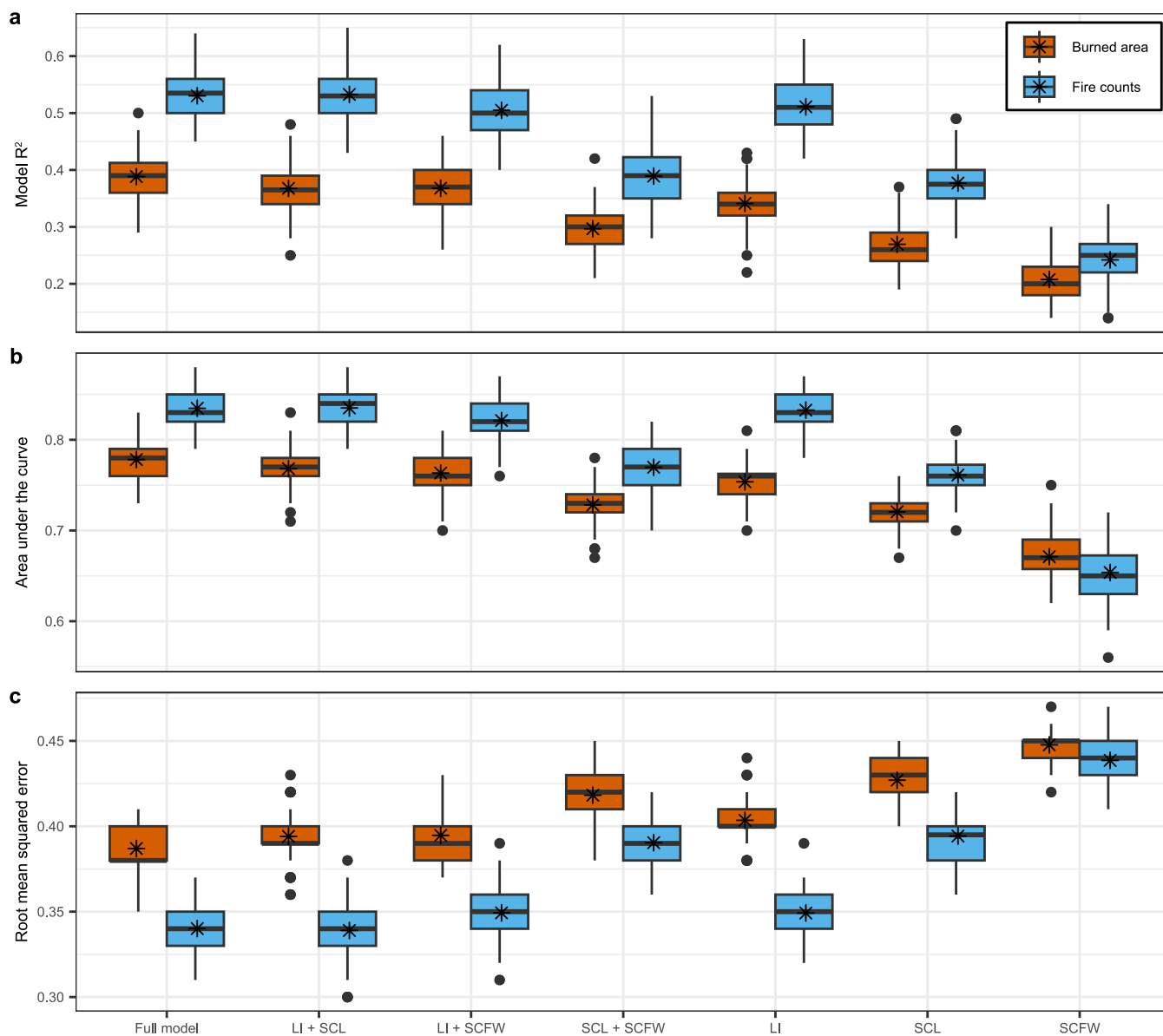
Extended Data Fig. 4 | Statistical robustness of the model routine using different feature combinations. The model R^2 (**a**), area under the receiver operating characteristic (ROC) curve (**b**) and root mean squared error (RMSE, **c**) derived from the internal model validation using 100 XGBoost training and validation iterations with different feature combinations predicting the fraction of burned area or fire count attributed to lightning. LI = low impact land, SCL = seasonal correlation between lightning and burned area, SCFW = seasonal correlation between fire weather and burned area. The logistic regression objective in XGBoost predicts a range (0 to 1) of values which were used to validate the continuous values of lightning attribution, obtaining the R^2 and RMSE. After applying a threshold at 0.5 to the continuous range, we also assessed the binary values and obtain the area under the ROC curve. The black asterisks show the mean. The black line within each box denotes the median, the limits of the box represent the 25th (Q1) and 75th (Q3) percentiles and the lines extending above and below denote Q1 and Q3 minus or plus 1.5 times the interquartile range (Q3 - Q1), respectively. The sample size (n) for each variable combination is exactly 100.
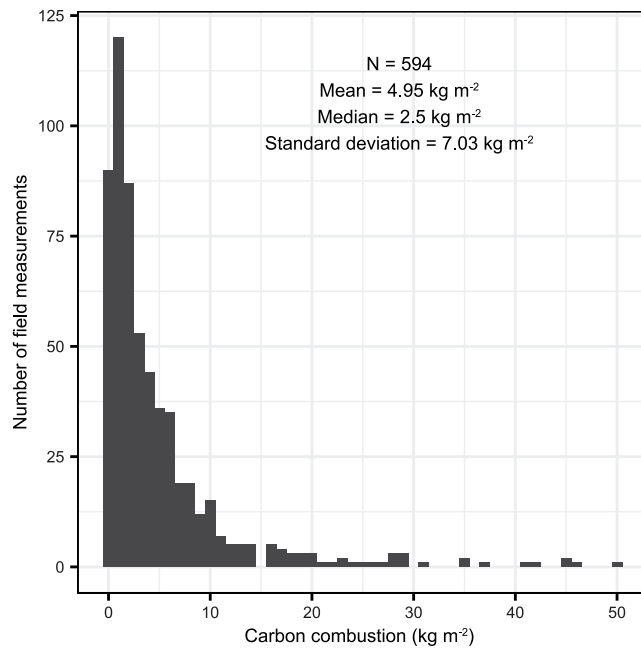
Extended Data Fig. 5 | Field-measured carbon combustion in extratropical intact forest landscapes. The carbon combustion field measurements were retrieved from the dataset of van Wees et al.[20] and only include locations within extratropical intact forests.