nature portfolio

https://doi.org/10.1038/s41586-023-06444-3

Supplementary information

Climate warming increases extreme daily wildfire growth risk in California

In the format provided by the authors and unedited

Supplementary Information for

Climate Warming Increases Extreme Daily Wildfire Growth Risk in California

Patrick T. Brown[1,2,3]*, Holt Hanley[2,4,5], Ankur Mahesh[6,7], Colorado Reed[8], Scott J. Strenfel[9], Steven J. Davis[10], Adam K. Kochanski[2,4], Craig B. Clements[2,4]

[1]Climate and Energy Team, The Breakthrough Institute, Berkeley, CA, USA

[2]Wildfire Interdisciplinary Research Center (WIRC), San José State University, San Jose, CA, USA

[3]Energy Policy and Climate Program, Johns Hopkins University, Baltimore, MD, USA

[4]Department of Meteorology and Climate Science, San José State University, San Jose, CA, USA

[5]KSBW News, Salinas, CA, USA

[6]Climate and Ecosystems Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

[7]Earth and Planetary Science, University of California, Berkeley, Berkeley, CA, USA

[8]Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA, USA

[9]Pacific Gas and Electric Company, Oakland, CA, USA

[10]Department of Earth System Science, University of California, Irvine, Irvine, CA, USA

*Corresponding author. Email: Patrick@thebreakthrough.org

S1. Overview and Concept

The goal of the study is to assess anthropogenic warming's influence on the risk of extreme wildfire behavior in California using an empirical approach where the relationship between temperature and the risk of extreme wildfire behavior is learned from 17,910 geographically dispersed fire-days in California from 2003-2020. We also seek to make attribution statements at the level of individual fire-days. See a video explanation of the method here: https://youtu.be/lHztGWzghRI

The extreme wildfire behavior that we focus on is extreme daily growth, specifically 10,000 acres or more in a day. This threshold was chosen because it's extreme enough to embody the outsized impact we are interested in, but it is not too extreme such that there are only a handful of events to study. Since we are specifically interested in the occurrence of this class of events taking place or not, we convert our wildfire growth data into a binary response variable which indicates if growth above 10,000 acres occurred.

We also characterize the occurrence of the events probabilistically as we treat their occurrence as containing fundamental uncertainty due to incomplete information as well as atmospheric chaos.

We are specifically interested in how a fire's environment (including ambient temperature) influences the risk of extreme daily growth. We, therefore, use regression models to understand the relationship between the risk of extreme daily growth and a number of predictor variables established to be important influences on wildfire behavior (Fig. 1, Table S2). These predictor variables represent fundamental components of the canonical daily-timescale wildfire behavior triangle[1] which has edges of topography, fuels, and weather. We did not prescreen predictors for their predictive skill because we were not trying to optimize for predictive skill over physical understanding. Rather, the predictors were selected so that the models would be able to make predictions about the influence of temperature changes on extreme daily wildfire growth risk conditional on a set of fundamental attributes, like slope and vegetation type.

Many traditional regression methods assume that the influence of any predictor variable is independent of the influence of the other predictor variables and that their influences are monotonic, if not linear. However, it is well known that the influence of any one of our predictor variables will be highly conditional on the state of other predictor variables. Because of this, there are not strong linear relationships between any one of our predictors and daily fire growth.

Thus, rather than use traditional regression methods, we build on recent work in the field[2-5] and implement machine learning models (specifically neural networks and random forests) in order to estimate the associations between the environmental conditions (or predictors) and the risk of extreme daily wildfire growth. These models are able to account for non-linear, non-monotonic, and interactive relationships between the predictors without the researcher having to presuppose the functional forms of such relationships.

We confirm that these predictors do constrain the risk of extreme wildfire growth using leave-3-years-out cross-validation where each 3-year chunk (2003-2005, 2006-2008,…,2018-2020) is held out in sequence, and the remaining data is used to train the machine learning models. Then the machine learning models make probabilistic predictions of extreme growth on the held-out data, and their performance on that data is assessed using four different scoring metrics (log-loss, Brier, reliability diagram, and the area under the Receiver Operating Characteristic curve, ROC-AUC, Fig. S3A, and Fig. S3B).

We test 1,000 neural network configurations and 1,000 random forest configurations by randomly varying hyperparameters for each (individual dots in Fig. S3A and S3B). The scoring metrics are standardized against a naïve model that always predicts the baseline probability of extreme growth (anything over 0 in the columns labeled "skill score" in Fig. S3A and S3B indicated better performance than the naïve model). Our main reported results throughout the manuscript are results that are averaged over the model configurations that performed in the top 10% in the leave-3-years-out cross-validation test. These top 10% models are indicated with small black dots in Fig. S3A and S3B. We also test the models' ability to make predictions outside of their parameter space in the Train on Cool, Test on Warm Experiment discussed below.

The learned relationships take advantage of the large amount of data (which allows for a large exploration of the parameter space) and the large variance in the feature variables that are dominated by geographic, seasonal, and weather variability as opposed to long-term trends. This means that factors like the long-term increase in fuel buildup or the long-term expansion of human population into the wildland-urban interface do not co-vary with temperature in this dataset and thus are not confounding factors in the way they are on, e.g., long-term annual area-burned trends.

Because the learned relationships are associational, they cannot strictly be considered causal, though knowledge of the physical world, for example, that fuels burn more readily when they are drier, can be used to supplement the associations found in the data to make reasonably confident inferences about causality.

Broadly speaking, our approach can be summarized in two steps:

1) Learn relationships between environmental conditions and the risks of extreme daily fire growth for historical fires from 2003 to 2020

2) Recalculate the risks under altered background climatological temperatures associated with anthropogenic warming

We compare the probabilities of extreme daily growth calculated with the original predictor values to those calculated with the altered predictor values using a probability ratio[6],

$$Probability\ Ratio = \frac{P(extreme\ daily\ growth\mid warmed\ temperatures)}{P(extreme\ daily\ growth\mid preindustria\ temperatures)}. \qquad \text{S1}$$

A Probability ratio of 1.5, for example, would mean that the climatological temperature change being considered made the risk of extreme daily growth 50% more likely.

Also, for historical fire-days that did see extreme daily growth, we can calculate the fraction of the risk of that event occurring that was due to climatological temperature change – the Fraction of Attributable Risk [7-9]

$$Fraction\ of\ Attributable\ Risk\ = 1 - \frac{1}{Probability\ Ratio\ for\ an\ extreme\ growth\ day}. \qquad \text{S2}$$

The idea is to quantify the proportion of the risk of extreme growth that is due to the climatological *change* in temperature (See section S4 for how the climatological change in temperature is calculated).

If the change in temperature doubles the risk (the probability ratio is 2), then half of the risk can be attributed to temperature change. If it triples the risk, then 2/3^{rds} of the risk can be attributed to temperature change and so on.

The altered background climatological temperatures come from multi-decadal and multi-model means produced by global climate models (Table S1, [10]). These climatological changes in temperature are a function of space and the month of the year.

It is already well-established that anthropogenic warming's influence on wildfire behavior is not primarily via temperature per se but rather through temperature's influence on aridity. Thus, when attempting to assess anthropogenic warming's influence on the risk of extreme wildfire behavior, we must also consider temperature's influence on aridity predictors that have a direct relationship with temperature. Accordingly, we also propagate changes in temperature into the daily mean vapor pressure deficit (which influences the aridity of fine fuels), 100 hour dead fuel moisture, and 1,000 hour dead fuel moisture (which represent fuel moistures for fuels of different sizes and thus different response times). Both of these dead fuel moisture variables experience the background climatological change in temperature over a period of 125 days prior to the fire day (Eqs S5-S10).

We test the influence (on risk change) of propagating temperature into every possible predictor combination and find that 100 hour dead fuel moisture has the largest influence while temperature itself has the smallest influence (Fig. 3A and 3B).

Every other predictor variable other than temperature and the three aridity metrics is held constant in our procedure, including topography, vegetation characteristics, precipitation, wind, and absolute moisture content of the atmosphere (vapor pressure). Conceptually, the question being asked is:

> "What if everything in terms of weather and ignitions over 2003-2020 was the same, except the background temperature was changed to be like the 1850-1900 average, or the 2041-2060 average, or the 2081-2100 average under different emissions scenarios?"

We believe this experiment design results in conservative estimates of the change in risk as there is evidence that many of the predictors that are held constant are likely being pushed in a direction that would exacerbate rather than attenuate the risk of extreme wildfire growth.

Our approach is similar to the established "Pseudo-Global Warming" or "Storyline" approaches in the extreme event attribution literature [11,12], which also hold almost everything about a given weather event constant except for a small handful of variables of interest.

The difference, however, is that in the "Pseudo-Global Warming" or "Storyline" approaches, the influence of the independent variable (temperature, for example) on the extreme weather event is typically quantified through a physical or mechanistic model that calculates interactions through a conglomeration of physical equations and semi-empirical parameterizations. Our approach is similar in principle, but the models that quantify the influence of the independent variable on the extreme event are machine learning models that have learned the empirical relationships directly from the data. To reiterate, the advantage of using machine learning models is that the relationships can be quantified more accurately than in some mechanistic models, but the disadvantage is that relationships are associational by definition and much care needs to be taken before they can be interpreted as causal.

An advantage of holding everything about the weather during fire-days constant and thus isolating the influence of temperature alone (as well as temperature’s influence on aridity) is that our confidence in long-time-mean temperature changes simulated by global climate models is higher than our confidence in any other predictor variable output from global climate models and higher than our confidence in simulated changes in the aggregate statistics of daily weather extremes. Climate change is often broken down into a thermodynamic and a dynamic component where thermodynamic changes refer to the influence of warming alone and dynamic changes take into account all shifts in circulation patterns that may arise from warming (e.g., change in the relative frequency of El-Ninos, change in meridional temperature gradients, change in the strength of location of the descending branch of the Hadley cell, etc.). Thermodynamic changes are much more certain and constrained than dynamic changes. Thus, it is a reasonable representation of the state of knowledge of future climate change to focus on the impact of thermodynamic changes while holding dynamic changes constant.

Using anthropogenic warming calculated from global climate models, the trained machine learning models can then make altered predictions of the probability of extreme daily growth on the same historical fire-days but under different background climatological temperatures.

In our procedure, the out-of-training-sample predictions and shifts in future probability are calculated by the same model fit on the same training data in the cross-validation procedure, so the models never have access to predictor information for any fire-day that they are assigning a probability to, either using original or altered predictor values.

Due to the output of our method being fundamentally probabilistic, it also has similarities to the "Probability-based approach" of the extreme event attribution literature [12]. Thus, with its conceptual similarity to the "Probability-based approach" as well as the "storyline approach", we refer to our method as the "Probabilistic-Storyline Approach" for extreme event attribution.

Further details on the predictors, response, machine learning models, methods, and sensitivity tests are provided below. See also a video explanation of method: https://youtu.be/lHztGWzghRI

S2. Response Data (Daily Fire Growth)

Geolocated daily fire growth from 2003 to 2020 within California state lines was calculated from raw fire detect data obtained from the MODIS system on NASA's Terra and Aqua satellites by a team at Sonoma Technology (http://www.sonomatech.com/). Analysis was restricted to the state of California because the dataset was created for use by the Pacific Gas and Electric Company (PG&E) whose territory is contained within California.

We only investigate fire-days that occurred in nine land categories defined in the Weather Research and Forecasting (WRF) model. We aggregated these nine categories to three categories in our predictor dataset representing either Forest (Category 1), Shrub (Category 2), or Savanna/Grassland (Category 3). We also restricted the analysis to locations with at least 20% vegetation fraction. After this initial filtering, there are 17,910 fire-days in our dataset and 380 instances of extreme daily growth (10,000 acres in a day), which we treat as an "event". The set-up of the binary classification problem is shown in Table S2.

S3. Predictor Data

The 11 predictor variables (Table S2) used as input to the machine learning models were obtained from a high resolution (2km×2km, hourly, with 50 vertical levels) reanalysis produced from the National Center for Atmospheric Research's (NCAR) Weather Research and Forecasting (WRF) model (version 4.1.2). The WRF model is open source and can be downloaded at https://github.com/wrf-model/WRF/releases. The National Centers for Environmental Prediction's (NCEP) Climate Forecast System Version Reanalysis (*30*) provided initial and boundary conditions (every 6 hours) for the high-resolution WRF reanalysis. CFSR was used prior to 2011, and CFSv2 was used after 2011.

The WRF reanalysis was produced to complement PG&E's Operational Mesoscale Modelling System (POMMS). The WRF model was tested and validated by teams at DTN (https://www.dtn.com/) and Atmospheric Data Solutions (http://www.atmosphericdatasolutions.com/). Roughly 20 model configurations were tested against a network of hundreds of weather stations in California (from ASOS, PG&E, and RAWS), and the configuration below was deemed to be optimal for PG&E's needs in operational forecasting with a particular emphasis on fire-weather forecasting.

Nested grid resolutions: 18, 6, 2km; Model top pressure level: 20mb; Land use: MODIS30s with lakes (modis_landuse_20class_30s_with_lakes); Land Surface Model: NoahMP; Radiation Schemes: RRTMG; Microphysics scheme: Thompson; Planetary Boundary layer scheme: MYNN2.5; Surface layer scheme: MYNN; Cumulus scheme for outer domain: Kain-Fritsch; Topo shading for innermost domain; Slope-dependent radiation for innermost domain; No nudging.

The WRF reanalysis data was originally at the hourly temporal resolution, but all variables were averaged to the daily temporal resolution to match the temporal resolution of the fire growth data. Values of predictors were obtained for each fire-day in the dataset by using the value of the grid box closest to the latitude and longitude of fire ignition. The 2km resolution portion of the domain excluded the southeastern portion of California (innermost box in Fig. S1).

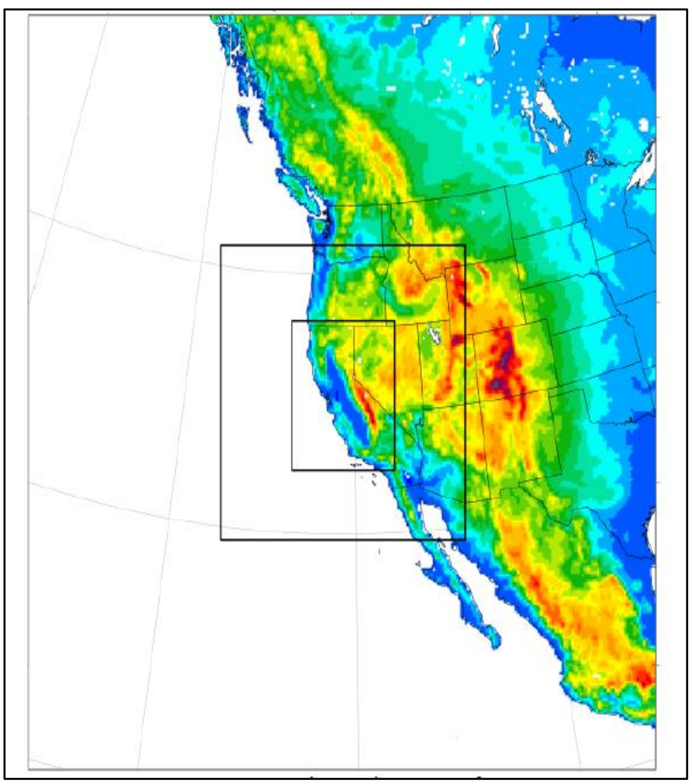

Fig. S1. WRF reanalysis domain configuration (used for predictor values) with elevation shaded. The outer-most domain has a grid spacing of 18km, the middle domain has a grid spacing of 6km, and the inner-most domain (the domain of this study) has a grid spacing of 2km. Data and figure were produced from a partnership between DTN, ADS, and PG&E.

S4. Calculations of Background Anthropogenic Temperature Change

Background anthropogenic warming was obtained from 34 Global Climate Models (GCMs) that participated in the Coupled Model Intercomparison Project – Phase 6 (CMIP6, Table S1). The following procedure was implemented.

1) The mean temperature was calculated across three time periods: 1850-1900 in the historical experiment, 2041-2060, and 2081-2100 in the SSP1-2.6, SSP2-4.5, and SSP5-8.5 emissions scenarios. This was obtained from the IPCC WGI Interactive Atlas (https://interactive-atlas.ipcc.ch/regional-information). Data were downloaded separately for each month of the year and as a function of space on a 1º×1º grid over our domain.

2) Data was bilinearly interpolated to our fire-day locations.

3) A mean temperature was calculated for the current period 2003-2020 by interpolating in time between the 1995-2014 value and the 2041-2060 value (for each month of the year and grid point over our domain). Given that the two endpoints were 20-year means, the interpolated value was also a 20-year mean but centered on the 2003-2020 period.

4) Temperature changes were calculated all relative to the 2003-2020 period, so no climatological mean biases would be allowed to affect the analysis (Fig. S2).

5) Temperature changes were incorporated into predictors for all the fire-days.

We note that the SSP5-8.5 emissions scenario is thought to be a very high and thus unlikely scenario[13]. It is inconsistent with current projections of energy systems over the remainder of the century[14,15]. Nevertheless, even if SSP5-8.5 emissions levels are now outside of projections associated with economic growth and energy systems, greenhouse gas concentrations could approach those associated with the SSP5-8.5 scenario if carbon cycle feedbacks are on the positive end of their uncertainty range. Furthermore, including SSP5-8.5 in our analysis allows us to calculate probability ratios associated with temperature changes that are very much still on the table, even if they were to occur sometime in the 22nd or 23rd centuries rather than in the period described here (2081-2100).

Table S1. CMIP6 Models used in this study. Model availability and use in this study are denoted with an X.

					Projections 2041-206 and 2081-2100			Res		ESGF versions	Ensemble
	Model	Historical (1850-1900)	Historical (1995-2014)	Interpolated 2001-2020	SSP1-2.6	SSP3-4.5	SSP5-8.5	lat	lon		
1	ACCESS-CM2:	X	X	X	X	X	X	1.88º,	1.25º	v20191108	https://doi.org/10.22033/ESGF/CMIP6.4332
2	ACCESS-ESM1-5:	X	X	X	X	X	X	1.88º,	1.25º	v20191115	https://doi.org/10.22033/ESGF/CMIP6.4333
3	AWI-CM-1-1-MR:	X	X	X	X	X	X	0.94º,	0.93º	v20190529	https://doi.org/10.22033/ESGF/CMIP6.2817
4	BCC-CSM2-MR:	X	X	X	X	X	X	1.12º,	1.11º	v20190318	https://doi.org/10.22033/ESGF/CMIP6.3050
5	CAMS-CSM1-0:	X	X	X	X	X	X	1º,	1º	v20191106	https://doi.org/10.22033/ESGF/CMIP6.11052
6	CANESM5:	X	X	X	X	X	X	2.81º,	2.77º	v20190429	https://doi.org/10.22033/ESGF/CMIP6.3696
7	CESM2:	X	X	X	X	X	X	1.25º,	0.9º	v20200528	https://doi.org/10.22033/ESGF/CMIP6.7768
8	CESM2-WACCM:	X	X	X	X	X	X	1.25º,	0.94º	v20200702	https://doi.org/10.22033/ESGF/CMIP6.10115
9	CMCC-CM2-SR5:	X	X	X	X	X	X	1º,	1º	v20200622	http://doi.org/10.22033/ESGF/CMIP6.3896

10	CNRM-CM6-1:	X	X	X	X	X	X	1.41º,	1.39º	v20190219	https://doi.org/10.22033/ESGF/CMIP6.4224
11	CNRM-CM6-1-HR:	X	X	X	X	X	X	0.5º,	0.5º	v20191202	https://doi.org/10.22033/ESGF/CMIP6.4225
12	CNRM-ESM2-1:	X	X	X	X	X	X	1.41º,	1.39º	v20191021	https://doi.org/10.22033/ESGF/CMIP6.4226
13	EC-EARTH3:	X	X	X	X	X	X	0.7º,	0.7º	v20200310	https://doi.org/10.22033/ESGF/CMIP6.4912
14	EC-EARTH3-Veg:	X	X	X	X	X	X	0.7º,	0.7º	v20200225	https://doi.org/10.22033/ESGF/CMIP6.4914
15	FGOALS-g3:	X	X	X	X	X	X	2º,	5.18º	v20190819	https://doi.org/10.22033/ESGF/CMIP6.3503
16	GFDL-CM4:	X	X	O	O	X	X	1.25º,	1º	v20180701	https://doi.org/10.22033/ESGF/CMIP6.9268
17	GFDL-ESM4:	X	X	X	X	X	X	1.25º,	1º	v20180701	https://doi.org/10.22033/ESGF/CMIP6.8706
18	HADGEM3-GC31-LL:	X	X	X	X	X	X	1.88º,	1.25º	v20200114	https://doi.org/10.22033/ESGF/CMIP6.10901
19	IITM-ESM:	X	X	X	X	X	X	2º,	2º	v20200915	https://doi.org/10.22033/ESGF/CMIP6.44
20	INM-CM4-8:	X	X	X	X	X	X	2º,	1.5º	v20190603	https://doi.org/10.22033/ESGF/CMIP6.12337
21	INM-CM5-0:	X	X	X	X	X	X	2º,	1.5º	v20190724	https://doi.org/10.22033/ESGF/CMIP6.12338
22	IPSL-CM6A-LR:	X	X	X	X	X	X	2.5º,	1.27º	v20190903	https://doi.org/10.22033/ESGF/CMIP6.5271
23	KACE-1-0-G:	X	X	X	X	X	X	1.88º,	1.25º	v20200317	https://doi.org/10.22033/ESGF/CMIP6.8456
24	KIOST-ESM:	X	X	O	O	O	O	1º,	1º		
25	MIROC-ES2L:	X	X	X	X	X	X	2.81º,	2.77º	v20200318	https://doi.org/10.22033/ESGF/CMIP6.5770
26	MIROC6:	X	X	X	X	X	X	1.41º,	1.39º	v20191016	https://doi.org/10.22033/ESGF/CMIP6.5771
27	MPI-ESM1-2-HR:	X	X	X	X	X	X	0.93º,	0.93º	v20190710	https://doi.org/10.22033/ESGF/CMIP6.4403
28	MPI-ESM1-2-LR:	X	X	X	X	X	X	1.88º,	1.85º	v20190710	https://doi.org/10.22033/ESGF/CMIP6.6705
29	MRI-ESM2-0:	X	X	X	X	X	X	1.12º,	1.11º	v20190608	https://doi.org/10.22033/ESGF/CMIP6.6929
30	NESM3:	X	X	X	X	X	X	1.88º,	1.85º	v20190811	https://doi.org/10.22033/ESGF/CMIP6.8790
31	NorESM2-LM:	X	X	X	X	X	X	2.5º,	1.89º	v20191111	https://doi.org/10.22033/ESGF/CMIP6.8319
32	NorESM2-MM:	X	X	X	X	X	X	1.25º,	0.94º	v20191115	https://doi.org/10.22033/ESGF/CMIP6.8321
33	TaiESM1:	X	X	O	O	X	X	1.25º,	0.9º	v20200902	https://doi.org/10.22033/ESGF/CMIP6.9823
34	UKESM1-0-LL:	X	X	X	X	X	X	1.88º,	1.25º	v20190726	https://doi.org/10.22033/ESGF/CMIP6.6405

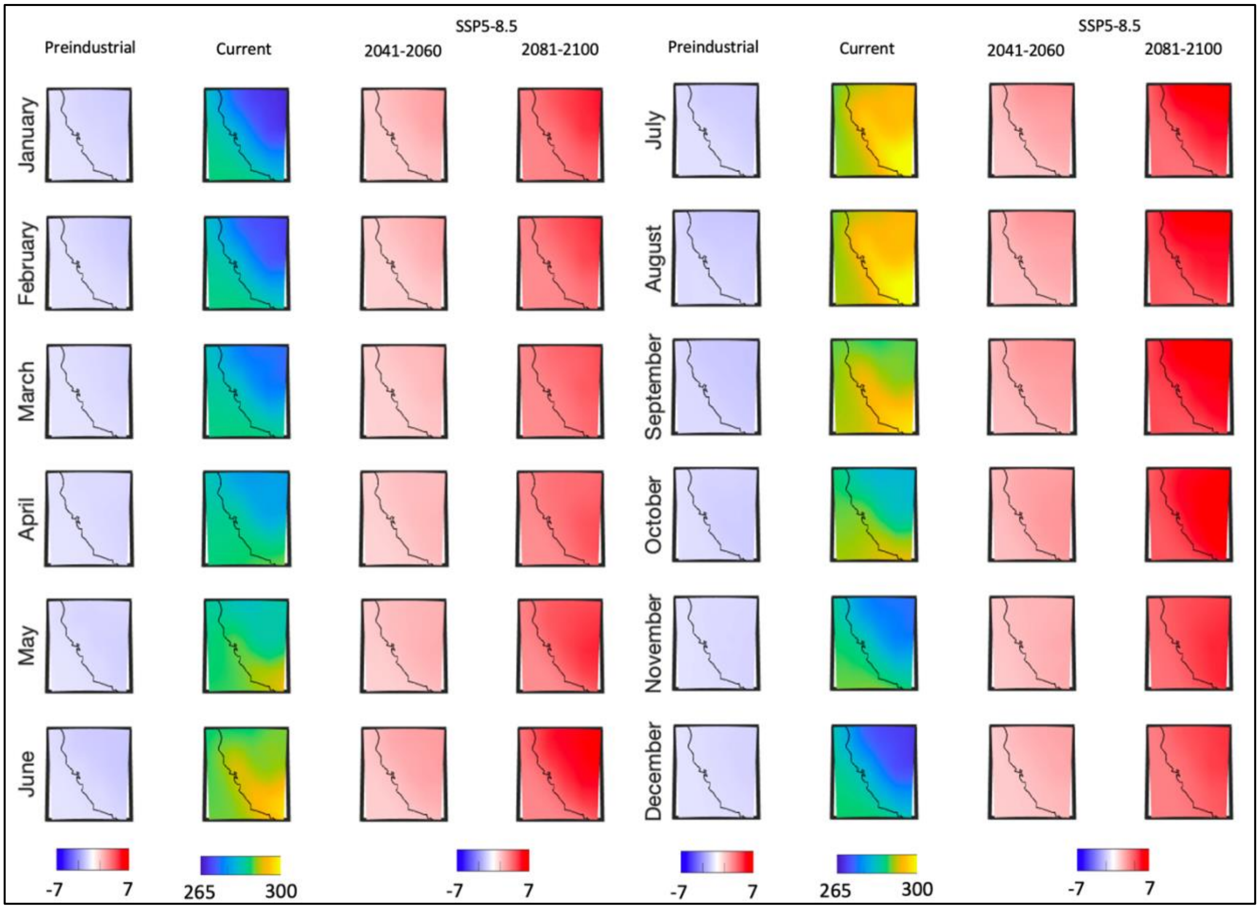

Fig. S2. Structure of background climatological temperature changes imposed on the historical fires from 2003 to 2020. This figure uses SSP5-8.5 as an example. These temperature changes are calculated from the CMIP6 multi-model mean (Table S1) over decade+ time periods in order to 'average out' unforced variability. Their structure is latitude by longitude by month-of-year.

S5. Propagation of Background Temperature Change into Temperature-Dependent Predictors

GCM-calculated changes in climatological temperature were propagated into the three aridity-related predictors that have a direct relationship with temperature (i.e., it is possible to related them to temperature with a well-known equation or small set of equations, Eqs. S3-S10, Fig. 1C, 1D). These variables are Vapor Pressure Deficit (VPD) and the two dead-fuel moisture variables (100 hour and 1,000 hour). Although temperature inevitably affects precipitation and wind speed, these relationships are less local, less direct, and less certain, so we hold those two variables constant.

Changes in temperature are propagated into the VPD predictor via Saturation Vapor Pressure (SVP),

$$VPD(T) = SVP(T) - VP, \qquad \text{S3}$$

which is exponentially related to temperature. We use a common approximation of this relationship, with T in ºC:

$$SVP(T) = 611 \cdot e^{\frac{17.27 \cdot T}{237.3+T}}. \qquad \text{S4}$$

Vapor pressure deficit on the day of the fire is associated with fine fuel moisture (see section S27 for the use of 10 hour dead fuel moisture instead of vapor pressure deficit). We also propagate temperature directly into two fuel moisture estimates associated with coarser fuels: 100 hour and 1,000 hour dead fuel moisture estimates. This is done at all locations regardless of the actual fuel characteristics there (the models must use the other features like land use category and vegetation fraction in combination with calculated dead fuel moisture to learn the circumstances under which dead fuel moisture characteristics are important). The 100 hour and 1,000 hour time-lags (t_L) represent the time it takes a fuel particle to progress 63% of the difference between the current moisture content (m, in kg of water per kg of wood) and the moisture content it would have if it were perpetually under the same environmental conditions, referred to as the equilibrium moisture content (E). 100 hour fuels correspond roughly with particles of 4cm diameter, and 1,000 hour fuels correspond roughly to particles of 7.62cm diameter.

We create estimates of 100 hour and 1,000 hour dead fuel moisture for each fire-day in the dataset via the estimation made in ref.[16] by integrating the differential equation

$$\frac{dm}{dt} = \frac{E-m}{t_L} \quad \text{S5}$$

over 3,000 hours (125 days) prior to the fire-day. Three thousand hours was chosen since it represents the period over which an 1,000 hour particle would reach ~95% of its equilibrium value. The equilibrium value is modified slightly depending on if the moisture content starts above or below the equilibrium value,

$$\frac{dm}{dt} = \begin{cases} \frac{E_d-m}{t_L}\ if\ m > E_d \\ 0\ if\ E_d \geq m \geq E_w, \\ \frac{E_w-m}{t_L}\ if\ m < E_w \end{cases} \quad \text{S6}$$

where E_d is the drying equilibrium and E_w is the wetting equilibrium,

$$E_d(T) = 0.924 \cdot RH(T)^{0.679} + 0.000499 \cdot e^{0.1 \cdot RH(T)} + 0.18 \cdot (21.1 + 273.15 - T)\left(1 - e^{-0.115 \cdot RH(T)}\right) \quad \text{S7}$$

$$E_w(T) = 0.618 \cdot RH(T)^{0.753} + 0.000454 \cdot e^{0.1 \cdot RH(T)} + 0.18 \cdot (21.1 + 273.15 - T)\left(1 - e^{-0.115 \cdot RH(T)}\right) \quad \text{S8}$$

Also, when it rains (r, mm/h), the equilibrium moisture is replaced by the saturation moisture contents S, and the fuel moisture equation is modified to achieve the rain-wetting lag time t_r,

$$\frac{dm}{dt} = \frac{S-m}{t_r}\left(1 - e^{-\frac{r-r_0}{r_s}}\right), if\ r > r_0. \quad \text{S9}$$

We propagate changes in temperature into the 100 hour and 1,000 hour dead fuel moisture features by adjusting temperature directly in the equations above as well as adjusting relative humidity in accordance with SVP's response to the temperature change,

$$RH(T) = \frac{VP}{SVP(T)}. \quad \text{S10}$$

See ref.[16] for further discussion. Overall the 100 hour and 1,000 hour dead fuel moisture predictors provide information on the aridity experienced by a location over the antecedent weeks and months, and our results indicate that this information is more important for predicting extreme daily growth than temperature itself.

To reiterate points from S1, this study in on the influence of anthropogenic *temperature* change on the risk of extreme daily wildfire growth and thus we only propagate *temperature* changes into our aridity predictor variables and hold vapor pressure constant (equations S10, S3).

We believe that studying the influence of warming in isolation is valuable because temperature is the variable in the wildfire behavior triangle (Fig 1A) that is by far the most directly related to increasing greenhouse gas concentrations and, thus, the most well-constrained in future projections. Furthermore, we have confidence that changes in relative humidity (S10) and vapor pressure deficit (S3) are dominated by changes in temperature (and thus saturation vapor pressure)[17]. It is unclear, however, how we should expect vapor pressure to change in the long term over our region as there is a disagreement between historical observed decreases in vapor pressure[18,19] and climate model projected increases[20]. This uncertainty in the expected direction of change in vapor pressure justifies holding it constant.

Table S2. Example of data fed into the modeling framework

Fire Information not used by the models						**Response**		**Predictors**										
Fire_ID	Date	Ignition lat	Ignition lon		daily fire growth acres	Growth above 10,000 acres?		daily mean temperature	daily mean VPD	precip mean past 125 days	Daily mean wind speed	daily mean dead fuel moisture 100 hr	daily mean dead fuel moisture 1000 hr	slope	aspect	elevation	land use category	vegetation fraction
2003_1	1/2/03 0:00	35.30	-118.41		193.99	0		282.80	0.81	0.07	5.80	21.03	15.54	2.44	'4'	970.2042	'2'	27.77
2003_5	1/6/03 0:00	33.94	-117.62		4274.56	0		291.17	1.62	0.03	9.11	12.17	14.19	0.37	'3'	170.9051	'3'	33.16
2003_7	1/7/03 0:00	34.05	-118.87		1990.06	0		293.39	1.78	0.04	8.92	10.94	15.81	4.88	'3'	209.7267	'2'	33.83
2003_8	1/8/03 0:00	41.36	-123.85		193.99	0		288.70	1.31	0.28	2.40	49.53	26.55	10.06	'3'	489.9456	'1'	90.48
2003_9	1/8/03 0:00	40.60	-122.50		387.97	0		284.58	0.70	0.30	2.08	41.78	24.45	1.17	'1'	400.6298	'1'	75.01
2003_10	1/8/03 0:00	39.47	-123.68		193.99	0		289.88	1.28	0.22	2.33	46.04	27.61	1.27	'3'	176.5472	'1'	89.86
2003_13	1/13/03 0:00	34.42	-118.79		785.53	0		288.22	0.85	0.05	3.78	15.63	13.64	2.00	'2'	212.4662	'2'	29.76
2003_14	1/13/03 0:00	34.38	-118.86		193.99	0		287.73	0.81	0.05	3.87	16.54	14.17	4.14	'4'	291.4915	'2'	29.36
2003_15	1/13/03 0:00	34.33	-119.17		383.43	0		287.17	0.61	0.07	1.53	17.61	15.99	2.77	'2'	235.0125	'2'	27.08
2003_18	1/17/03 0:00	38.54	-122.48		193.99	0		286.70	0.86	0.18	2.16	28.58	22.25	3.34	'3'	160.9028	'1'	61.65
2003_19	1/17/03 0:00	41.36	-120.84		193.99	0		279.64	0.62	0.11	2.78	23.38	19.00	4.12	'1'	1671.546	'1'	64.89
2003_20	1/18/03 0:00	38.34	-120.57		193.99	0		286.42	1.01	0.14	2.44	15.95	17.44	1.08	'3'	832.7508	'1'	59.80
2003_21	1/19/03 0:00	40.37	-122.42		788.28	0		283.97	0.23	0.13	0.96	26.43	19.14	0.44	'1'	167.6649	'3'	36.93
2003_22	1/20/03 0:00	38.52	-120.86		575.91	0		283.43	0.36	0.11	2.12	18.86	17.37	1.92	'4'	358.6583	'3'	45.65
2003_23	1/20/03 0:00	39.38	-121.27		976.78	0		284.84	0.64	0.20	1.62	20.44	19.72	1.56	'2'	493.6429	'3'	44.84
2003_24	1/21/03 0:00	40.71	-123.87		193.99	0		284.12	0.32	0.33	5.42	23.75	26.79	5.00	'1'	482.8854	'1'	92.37
2003_25	1/21/03 0:00	36.50	-120.91		383.23	0		281.81	0.19	0.10	2.54	14.94	17.19	0.78	'2'	922.6371	'1'	33.85

2003_26	1/21/03 0:00	35.99	-120.67		385.44	0		283.13	0.17	0.06	1.48	15.03	16.10	1.80	'2'	573.1996	'1'	32.53
2003_28	1/21/03 0:00	41.14	-124.11		955.67	0		284.81	0.25	0.28	3.59	33.09	30.41	2.70	'1'	117.8349	'1'	89.66
2003_29	1/21/03 0:00	34.67	-118.98		384.56	0		279.49	0.39	0.03	2.20	11.16	13.55	8.35	'4'	1820.822	'1'	41.30

S6. Statistical Machine Learning Models

We use neural network and random forest machine learning models to quantify the relationship between our predictor variables and the risk of extreme growth. We also include simpler logistic regressing models.

S7. Neural Network Models

We use a feed-forward, fully connected, shallow neural networks produced with the function "fitcnet" built into Matlab (https://www.mathworks.com/help/stats/fitcnet.html).

The output from the neural network is a classification score computed using the softmax activation function that follows the final fully connected layer in the network,

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{2} e^{x_j}}. \quad \text{S11}$$

Since we treat this as a binomial classification problem, the final fully connected layer has two outputs, x_1 and x_2. The scores output from the above equation can be interpreted as the posterior probabilities of extreme daily growth on that day, given an active fire and given the values of the features fed into the neural network.

S8. Random Forest Models

We use a bootstrap aggregated (bagged) forest of decision trees produced with the function "treebagger" built into Matlab (https://www.mathworks.com/help/stats/treebagger.html).

The probability of observing extreme daily growth is taken as the fractional number of extreme daily growth days in a tree leaf from the training dataset averaged over all the trees in the ensemble (100 trees per model).

S9. Logistic Regression Models

In addition to the neural network and random forest models, we use a logistic regression model produced with the Matlab function "fitglm" (https://www.mathworks.com/help/stats/fitglm.html). We make use of the function to create derived predictors that are non-linear and/or combinations of multiple predictors. Below are the four separate logistic regression models that result as well as their functional forms.

<u>Linear:</u> The model contains an intercept and linear term for each predictor.

$$p(y_i) = \frac{1}{1+e^{(-b_0-b_1x_1-\cdots)}} \quad \text{S12}$$

<u>Pure Quadratic:</u> The model contains an intercept term and linear and squared terms for each predictor.

$$p(y_i) = \frac{1}{1+e^{(-b_0-b_1x_1-b_2x_1^2\cdots)}} \quad \text{S13}$$

<u>Interactions:</u> The model contains an intercept, linear term for each predictor, and all products of pairs of distinct predictors (no squared terms).

$$p(y_i) = \frac{1}{1+e^{(-b_0-b_1x_1-b_2x_1\cdot x_2-b_3x_1\cdot x_3\cdots)}} \quad \text{S14}$$

Quadratic: The model contains an intercept term, linear and squared terms for each predictor, and all products of pairs of distinct predictors.

$$p(y_i) = \frac{1}{1+e^{(-b_0-b_1x_1-b_1x_1^2-b_2x_1\cdot x_2-b_3x_1\cdot x_3 \ldots)}} \quad \text{S15}$$

where (x_1, x_2, …x_{11}) are the feature variables, and the model learns the coefficients b_0, b_1, …b_{11} to predict the probability of extreme growth $p(y_i)$

S10. Varying Neural Network and Random Forest Hyperparameters

In order to identify the most reliable model types and hyperparameter configurations, we created an ensemble of 1,000 neural network and 1,000 random forest models (See supplementary dataset for all 2,000 configurations).

The neural networks varied in i) the number of layers in the network, ii) the number of neurons in each layer, iii) the activation functions for the fully connected layers, and the iv) value of a regularization term that penalizes large weights in the neural network by adding a ridge (L2) penalty term to the cost function. The random forest models varied in i) their minimum leaf size, ii) the maximum number of splits, and iii) the split criteria (either deviance or Gini's diversity index).

Random samples of these hyperparameters were generated with a "random search" function built into Matlab.

S11. Machine Learning Model Performance and Relationship Between Performance and Calculated Shifts in Risk

We used four scoring methods to assess the performance of the 2,004 models (1,000 neural network, 1,000 random forest, and 4 logistic regression). These scoring methods were the log-loss score, the Brier Score, a Reliability Diagram Score, and an area under the Receiver Operating Characteristic curve, ROC-AUC. Ultimately, we use the log-loss score as our authoritative measure of model performance as it is the only scoring metric that satisfies the desirable conditions of additivity, "locality", and strictly proper behavior [21]. However, we show results for the other three scores to assess the sensitivity of results to the chosen performance metric.

Leave-3-years-out cross-validation was used where predictions were made on one 3-year block of data at a time, using the remaining data to train the models (see video explanation of method: https://youtu.be/lHztGWzghRI

S12. Log Loss (Binary Cross-Entropy) Skill Score

The log-loss score is calculated as,

$$Log\ loss\ score = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \cdot log\big(p(y_i)\big) + (1-y_i)\cdot log\ (1-p(y_i)\big)\right). \quad \text{S16}$$

Above, y_i is the binary indicator (1=extreme growth, 0=not extreme growth), and $p(y_i)$ is the model's predicted probability of extreme growth. A perfect model has a log loss score of zero.

S13. Brier Score

The Brier Score[22] is analogous to the mean square error and is calculated as,

$$Brier\ score = \frac{1}{N}\sum_{i=1}^{N}(p(y_i)-y_i)^2. \quad \text{S17}$$

A perfect model has a log loss score of zero.

S14. Reliability Diagram Score

We measure how well-calibrated the models' predictions are using reliability diagrams (Fig. S3C, S3D) and a reliability diagram score. We use bin widths that vary such that the number of fire-days within each bin is constant. This number was 1,000 for the leave-3-years-out cross-validated data (Fig. S3C) and 300 for the Train on Cool, Test on Warm Experiment (Fig. S3D, see below). We calculate the reliability diagram score as the mean absolute error over all bins between the mean model-predicted probability of extreme growth within a bin and the actual observed frequency of extreme growth corresponding to those predictions. This is effectively the distance of the reliability diagram lines from the 1:1 line (Fig. S3C, S3D)

S15. ROC-AUC

The Receiver Operating Characteristic Curve plots the true positive rate against the false-positive rate when continuously changing probability thresholds for a positive prediction. The area under this curve is a measure of model performance, with 1 being a perfect model and 0.5 representing a model with no discriminatory capacity. The ROC-AUC is calculated with the built-in Matlab function perfcurve (https://www.mathworks.com/help/stats/perfcurve.html).

S16. Skill Scores

In Fig S3A and Fig. S3B, we convert raw log-loss, Brier, and reliability diagram scores into skill scores to make them more interpretable,

$$Skill\ Score = 1 - \frac{Score_{machine\ learning\ model}}{Score_{naive\ model}}. \qquad \text{S18}$$

The skill score we use compares the models’ scores to a score that would have been achieved if the baseline probability of extreme growth was predicted for every instance (the naïve model). Skill scores greater than 0 indicate that the machine learning model is performing better than the naïve model, and thus the predictors do indeed provide information on the risk of extreme daily growth. As the machine learning model approaches perfection, the skill score approaches 1. Since the ROC-AUC score is already normalized and ranges from 0.5 to 1, we leave that score as a raw score and do not convert it to a skill score.

S17. Representative Ensemble of Machine Learning Models Used for Main Reported Results.

To make inferences about how the risk of extreme daily growth may change in the future, we wish to identify the model hyperparameter configurations that best generalize to out-of-sample data. Simultaneously, we are interested in the relationship between model-calculated change in risk and model performance on out of sample data. This information is shown in Fig. S3. Specifically, the leave-three-years out cross-validated skill of the 2,005 models (using the four different skill metrics) compared to their mean probability ratio (historical vs. preindustrial) over all fire-days is shown in Fig. S3A. Fig. S3B shows the same, but for the mean fraction of attributable risk. Mean probability ratios and fractions of attributable risk are not particularly sensitive to restricting models to higher standards of skill (i.e., they remain relatively stable when the top 2/3^{rds}, 1/3^{rd,} and 1/10th of model configurations are sampled.

We average over the top 10% of model configurations according to the log-loss skill score in our main reported results (black dots in all panels in Fig. S3A and S3B). In other words, once the top 10% of model configurations are identified, in terms of their log-loss score on out-of-sample data, their predicted probabilities for each fire day are averaged together prior for subsequent calculations of the probability ratios. The reasoning is that the model configurations that best predicted out-of-sample extreme growth days should also be the model configurations that we have the most confidence in for assessing shifts in probabilities of extreme growth with shifts in temperature. We also take these model configurations and conduct a Train on Cool, Test on Warm experiment (Section S18) that is analogous to the out-of-parameter space predictions being made under the climatological temperature changes being considered here.

Coincidentally, the top 10% of models (in terms of log-loss skill score) were equally represented by random forests and neural networks (50%-50% split). None of the four logistic regression models were in the top 10%. Table S3 shows the hyperparameter combinations of these top performing models.

A concern regarding the accuracy of our results has to do with whether the 11 predictor variables sufficiently constrain the probability of extreme daily growth given that they do not include obviously important factors like fine detail on fuel loads/breaks and firefighting efforts. However, our study demonstrates that extreme daily growth is indeed predictable even without this information. Specifically, our chosen model configurations are able to make predictions on out-of-training-sample data with a Receiver Operator Characteristic Area Under the Curve (ROC-AUC) of 0.89 (Fig. S3). This score is right on the border of what is typically considered "excellent" (0.8-0.9) and "outstanding" (>0.9)[23].

Similarly, the reliability diagram (Fig. S3C) shows that the predictions on out-of-training-sample data are reliable. In other words, when the models predict, e.g., about a 10% chance of extreme growth, extreme growth does, in fact, occur in about 10% of cases.

This means that the 11 predictor variables provided to the models give them a great deal of information about the likelihood of extreme growth. It would undoubtedly be helpful to include information on the additional attributes listed above and this could be the focus of future work. However, our results indicate that this information is not absolutely necessary in order to make skillful predictions.

Table S3. Hyperparameters of top 10% of Random Forest models (left 3 columns) and top 10% of Neural Network models (right three columns). Top 10% was determined on out-of-sample predictions in the historical period.

Random Forest			Neural Network		
Minimum leaf size	Maximum number of splits	Split Criterion	Activations	Lambda	Layer Sizes
2	114	gdi	sigmoid	5.61E-09	30
5	503	deviance	tanh	0.0001086	[15 20]
50	2	gdi	tanh	0.0003177	[45 3 5]
19	103	deviance	sigmoid	4.81E-06	59
2107	1248	gdi	sigmoid	1.63E-05	[32 1 3]
232	1149	gdi	tanh	0.0008725	[16 1 18]
5	165	gdi	sigmoid	6.88E-05	[8 2 2]
25	3243	deviance	sigmoid	0.0002414	7
2	2838	deviance	sigmoid	7.11E-06	[23 3]
1167	23	deviance	sigmoid	1.00E-05	6
1113	9700	gdi	sigmoid	2.13E-06	[37 1]
14	41	gdi	sigmoid	7.88E-05	[31 48]
2	1455	gdi	tanh	0.0003695	27

125	185	deviance	tanh	0.0023694	[240 129]
2	3802	gdi	sigmoid	5.68E-06	[7 34]
12	75	gdi	sigmoid	4.83E-07	14
1503	12	gdi	sigmoid	0.0002934	[61 110 4]
237	69	gdi	sigmoid	0.0001294	[6 194]
671	8032	deviance	tanh	8.37E-05	[49 15]
487	1	deviance	sigmoid	0.0001678	[40 30 25]
12	1423	gdi	sigmoid	0.0003446	[7 37]
1519	3967	gdi	tanh	0.0002848	[17 72 4]
76	75	deviance	tanh	0.0011261	[28 27]
544	2	gdi	tanh	0.0022476	[103 34]
43	2	deviance	tanh	0.0001614	[25 29]
7	758	deviance	sigmoid	4.74E-08	23
6908	9470	gdi	tanh	0.0003621	[13 71]
5	95	gdi	tanh	0.0031919	178
6936	67	deviance	sigmoid	6.89E-09	36
177	15	gdi	sigmoid	5.36E-05	[22 275]
4030	1163	gdi	sigmoid	0.0001678	[19 3]
24	13431	deviance	tanh	2.10E-05	9
2	229	gdi	sigmoid	1.63E-07	21
379	11140	deviance	tanh	0.0002846	[68 26]
614	90	gdi	sigmoid	5.04E-06	[61 72]
55	1	gdi	sigmoid	0.0011403	175
6	79	gdi	tanh	0.0026058	[41 27]
16	287	gdi	tanh	7.62E-06	17
1	116	deviance	sigmoid	0.0001398	10
1926	16	gdi	sigmoid	1.27E-06	[50 12]
583	1	gdi	sigmoid	2.35E-05	[8 101]
295	1030	deviance	sigmoid	3.12E-08	[51 91 82]
19	41	gdi	tanh	0.0001356	[18 24 1]
1	1260	deviance	sigmoid	3.38E-06	[47 2]
173	2	gdi	tanh	0.0003708	[20 246 71]
3	19	deviance	sigmoid	3.88E-08	[12 61]
1193	373	deviance	sigmoid	1.35E-07	[89 28]
5	222	deviance	sigmoid	2.16E-05	[8 64]
50	37	gdi	sigmoid	3.41E-06	[7 47]
202	4	gdi	sigmoid	4.75E-08	[7 62]
7	782	gdi	tanh	0.0014583	[107 8]
5799	1897	deviance	tanh	0.0003407	[20 293]
5	8628	deviance	sigmoid	4.79E-06	[6 116]
2124	24	gdi	sigmoid	1.40E-06	[52 4]
172	6	gdi	sigmoid	0.0000868	[2 27]

2	880	deviance	sigmoid	2.86E-07	22
5041	2282	gdi	sigmoid	7.62E-06	[3 284 4]
1285	2735	deviance	tanh	0.0007385	[18 82 3]
497	6383	deviance	sigmoid	1.91E-09	41

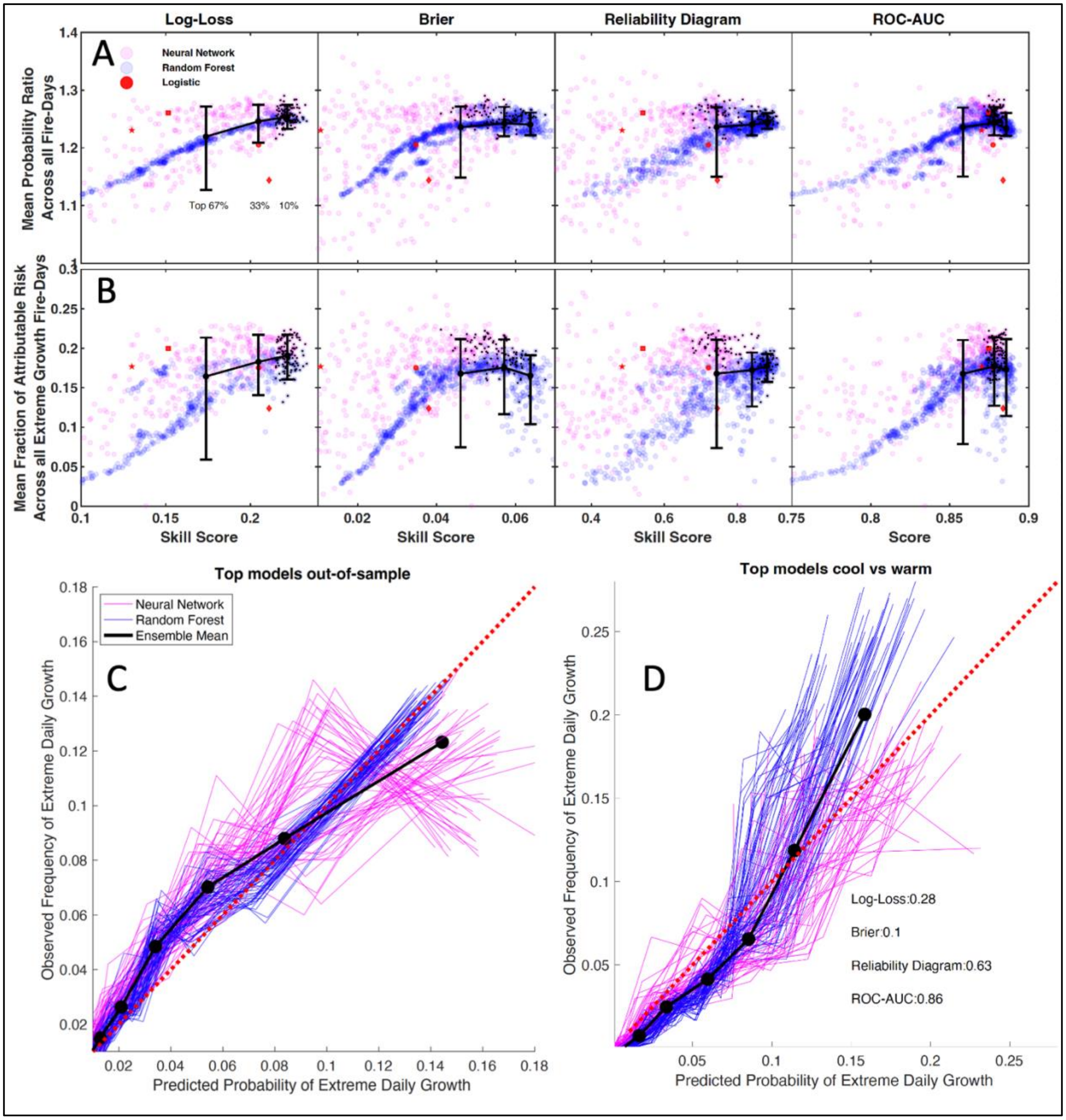

Fig. S3. (A) Relationship between a machine learning models' mean shift in probability (present vs. preindustrial, Y axis) across all fire-days and their out-of-training-sample performance using four score metrics (X axis) across 1,000 variations of neural network (magenta dots) and 1,000 variations of random forest (blue dots) hyperparameter combinations. There are also four versions of logistic regression models represented in red (The red circle corresponds to the Eq. S1, the red square corresponds to Eq. S2, the red diamond corresponds to eq. S3, and the red star corresponds to Eq. S4. The median values across the top 2/3rds, 1/3rd, and

1/10th of models are shown with error bars of 5-95 percentile ranges. Results from the top 10% of models in the log-loss score (small black dots) are averaged together to produce all the main reported results of this study. (B) Same as (A) but for the Fraction of Attributable Risk of the extreme growth days. (C) Reliability diagrams for the top 10% of models (in terms of log-loss) using the leave-3-years out cross-validation. (D) same as (C) but using the train-on-cool, test on warm training-testing split experiment.

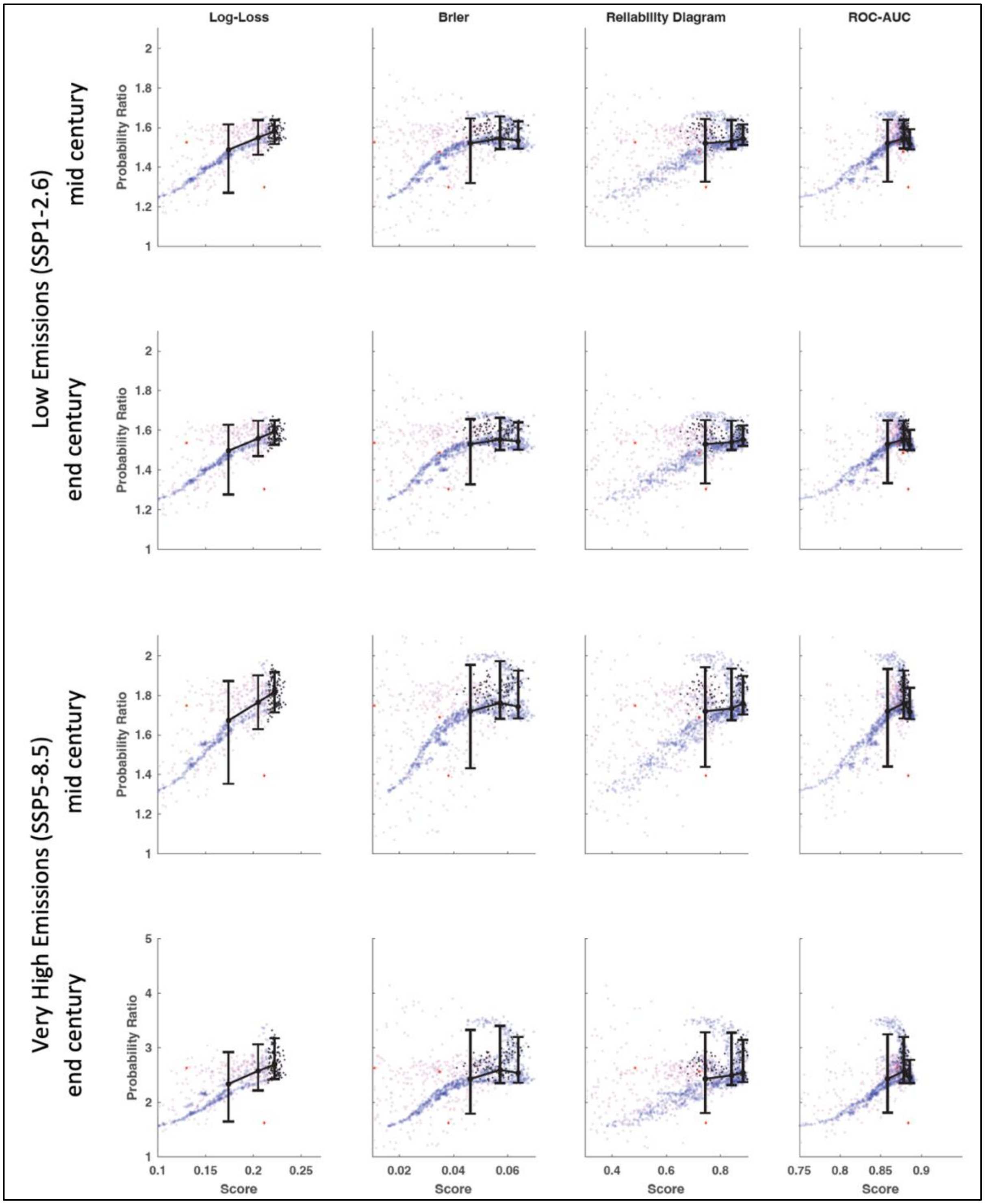

Fig. S4. Same as Fig. S3A but for future projected temperature changes.

S18. Train on Cool, Test on Warm Experiment

Because of their large number of degrees of freedom, machine learning models are prone to overfitting and should always be tested on data that they are not trained on. Our main test of model performance is on out-of-sample data *in time* (using leave-3-years-out cross-validation, https://youtu.be/lHztGWzghRI?t=135). This out-of-sample test is important for a model to perform well on, however, an even more stringent test is one that assess the models' ability to make predictions on fire-days with novel predictors values (i.e., outside of the "parameter space" they were trained on). Since the point of this study is to compare predicted probabilities between current climates and future novel climates, it is relevant to demonstrate that the models can generalize and make reasonable predictions to this new parameter space.

In this "Train on Cool, Test on Warm Experiment", we divide the data such that the training-testing split separates the data in parameter space in a similar way to how the data is separated in parameter space between the historical and end-century SSP5-8.5 climates (the most extreme climate change scenario we consider). We then assess how well the pre-selected models that performed best on the cross-validation *in time* (black dots in Fig. S3A and S3B) performed making predictions using this new training testing split.

The distributions of predictor values across all fire-days between the historical climate and the end-century SSP5-8.5 climate are shown in the top row of Fig. S5.

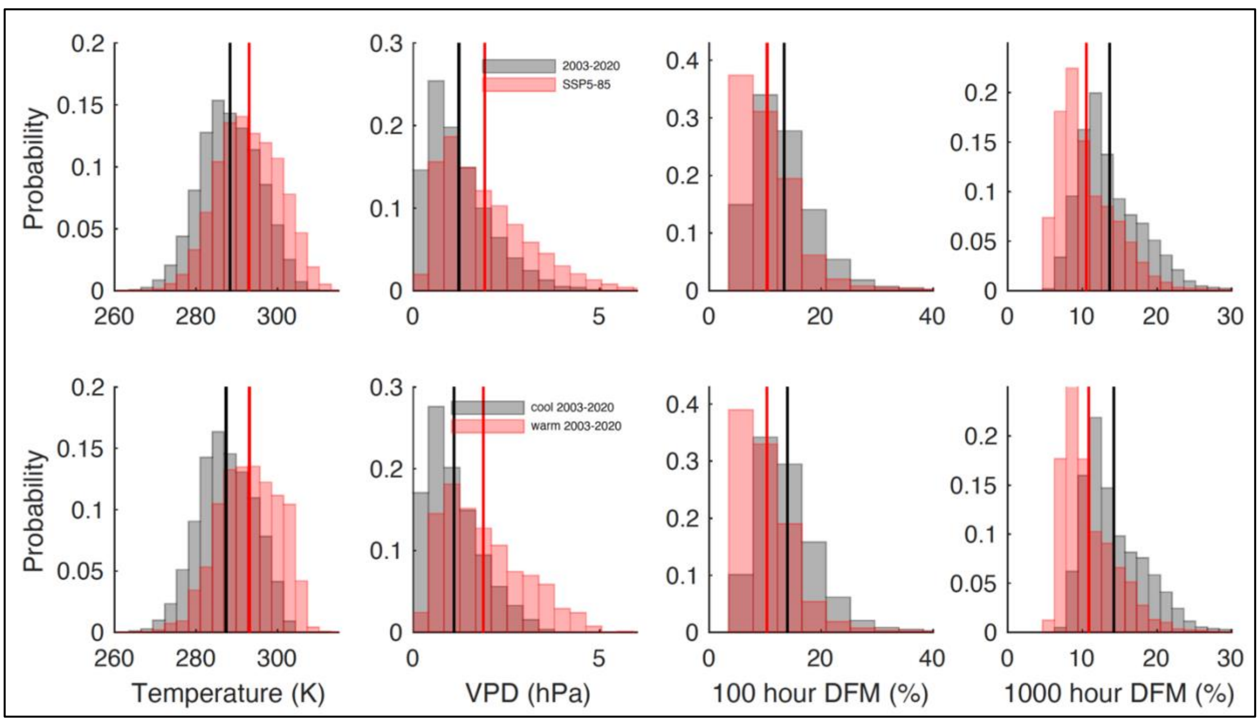

Fig. S5. Top) Distributions of predictor values (for the four that are altered) across all fire-days for the historical climate (grey) and end-century SSP5-8.5 (red). Bottom) Distributions of predictor values for the cool (grey) and warm (red) observations training/testing split. All distributions are normalized to sum to 1.

In order to split the data in a way that is analogous to the distributional differences in the top row of Fig. S5 (while attempting to preserve as many moments of the distributions as possible), we use a nearest-neighbor approach described below. There are four predictors that are shifted with climate change, but we illustrate the method using only two: temperature and 1000-hr dead fuel moisture (Fig. S6).

1. We begin with all 17,910 fire-days (black dots in Fig. S6)
2. We alter all predictor values in accordance with the climate change estimates for SSP5-8.5 (magenta dots in Fig. S6)
3. We select a random sample of 3,000 of these altered values (blue dots in Fig. S6), so we are left with the original 17,910 fire-days and 3,000 fire-days from the altered distribution.
4. We go through each of the 3,000 altered fire-days (blue) and find the fire-day from the original non-altered dataset (black) that is closest to it in parameter space. To find the closest fire-day in parameter space, we normalize all four predictors such that the minimum is zero and the maximum is one. We then find the minimal Pythagorean distance in four-dimensional space. Fire-days from the original dataset are only eligible to be picked once.

This procedure results in the selection of 3,000 fire-days from the original non-altered distribution that are the closest analogs to the 3,000 climate-change altered fire-days (red dots in Fig. S6). So we are left with 17,910-3,000=14,910 “cool” fire-days from the original dataset and 3,000 “warm” fire-days from the original dataset. The distributions of the four predictors for the “cool” and “warm” fire-days are shown in the bottom row of Fig. S5.

We then train the machine learning models on the cool fire-days and test them on the warm fire-days. The scores are shown in Fig. S3D. Of the four scoring metrics, the models performed better in two (log-loss and Brier) and worse in two (Reliability Diagram and ROC-AUC) than they did on the out-of-sample data in time.

In both model tests (out-of-sample in terms of time and out-of-sample in terms of parameter space), we interpret the results as indicating that the predictors do constrain the probability of extreme growth reasonably well, despite the models not having access to a great deal of relevant information (e.g., status of firefighting operations).

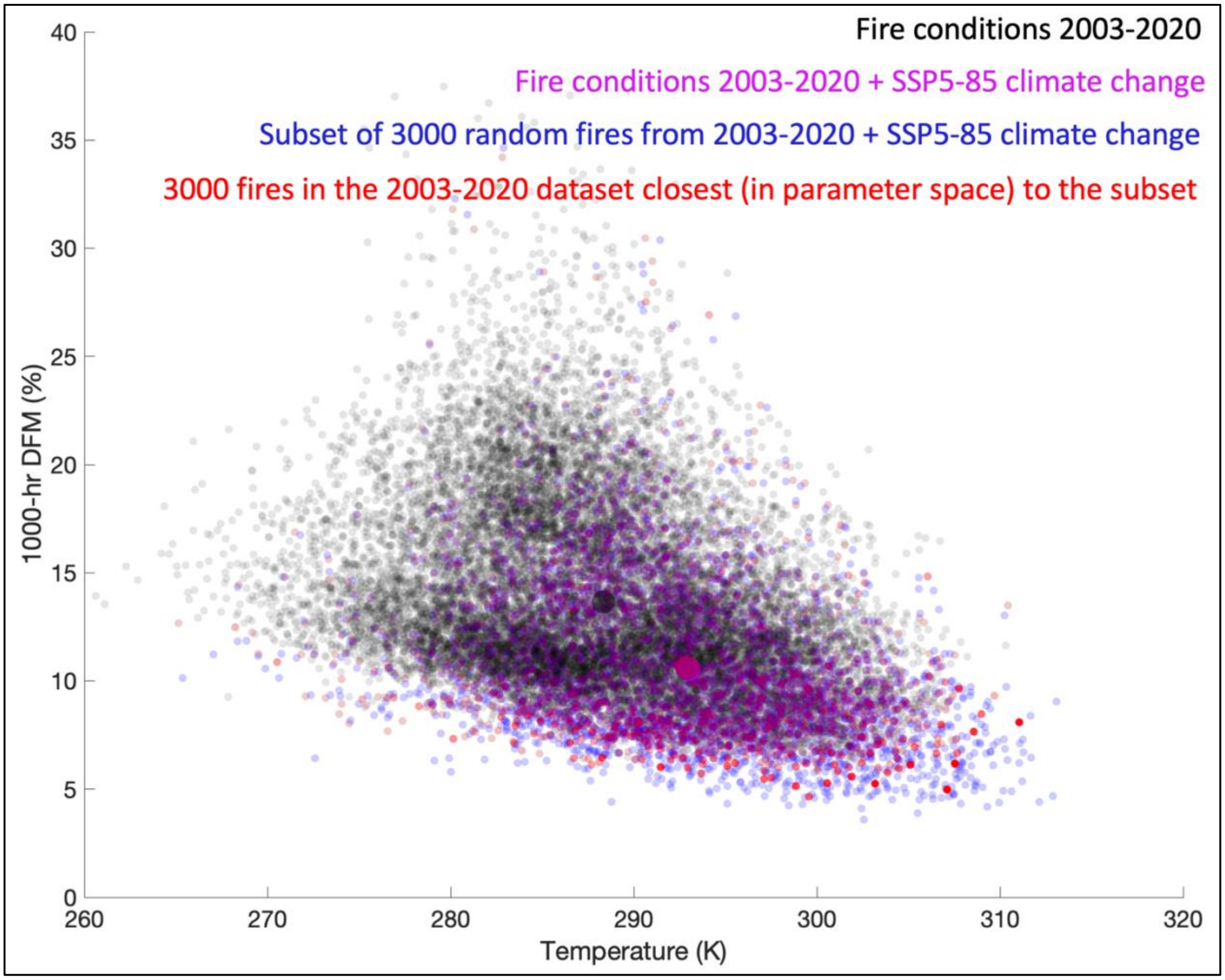

Fig. S6. Illustration of nearest neighbor approach in the train-on-cool, test-on-warm experiment. See text above for explanation.

S19. Expected Frequency of Extreme Daily Growth

The machine learning models assign risk to days based purely on the environmental conditions on that day and contain no information on the antecedent behavior of the fire. Thus, the daily probabilities are not temporally dependent and would receive the same probabilities if their sequences were scrambled. This means that it is not unreasonable to treat the probabilities as being independent of each other and sum daily probabilities of extreme daily growth to get an estimate of the aggregate expected frequency of extreme daily growth. This is how the expected frequencies in Fig. 2G are calculated, with Poisson distributions used to estimate the expected random variability in these frequencies. The finding that the observed frequency of 380 falls reasonably within the Poisson distribution for the cross-validated historical predictions is evidence that summing the probabilities provides a reasonable estimate of aggregate expected frequency even if probabilities are not actually independent of each other.

For the notable destructive fires highlighted in Fig. 2A-F, and Fig. 3C, 3D, probability ratios are calculated across their lifetimes such that probabilities under a given climate condition are averaged before the ratio is taken,

$$Probability\ Ratio = \frac{\frac{1}{n}\sum_{i=1}^{n}(P(extreme\ growth \mid warmed\ T))_i}{\frac{1}{n}\sum_{i=1}^{n}(P(extreme\ growth \mid preindustrial\ T))_i}, \quad \text{S19}$$

where i represents a day in the lifetime of the fire and n is the total number of days.

Calculating the probability ratio this way allows it to be interpreted as the ratio in the expected frequency of extreme growth days across the lifetime of the fire. Using the same calculation across all fire-days results in ratios exactly proportional to the changes in expected aggregate frequency calculated in Fig. 2G. This is a different calculation than the mean probability ratio of the fire-days,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{P(extreme\ growth \mid warmed\ T)}{P(extreme\ growth \mid preindustrial\ T)}\right)_i, \tag{S20}$$

which is what is displayed with vertical lines in the inset in Fig. 2.

S20. Probability Ratios Below One

There are portions of the 11-dimensional parameter space where warming and drying (for vapor pressure deficit, 100 hour and 1,000 hour fuel moisture) results in a decrease in the probability of extreme daily growth and thus probability ratios below 1. Figure S7 compares the geographical occurrence and the historical predictor value distributions between fire-days that display probability ratios above 1 (95.4% of all fire-days, red) and fire-days that display a probability ratio below 1 (4.6% of all fire-days, blue). We speculate here on the possible causes of probability ratios below 1 but leave a detailed investigation to future work (likely requiring physical models to disentangle causality from association).

Fire-days with probability ratios below 1 are spread throughout the domain and not concentrated in any particular geographic location (Fig. S7). Their distributions are bimodal in temperature, vapor pressure deficit, dead fuel moisture, and vegetation fraction. There are local minimums in the occurrence of probability ratios below 1 near the critical thresholds identified in Fig. 3C, 3D and S13.

The bimodal nature of the distributions of predictor values for fire-days with probability ratios below 1 suggests that at least two separate physical explanations might be necessary. Thus, Fig. S8 takes the fire-days with probability ratios below 1 and separates them by temperatures above and below 291K. This shows that these fire-days fit into two rough categories:

1) Those that are hot, dry, low elevation, low vegetation fraction, and disproportionately savanna/grassland (land classification 3)
2) Those that are cool, moist, higher elevation, high vegetation fraction, and disproportionately forest (land classification 1)

Category 1 fire-days seem to represent fires that are in fuel-limited, rather than aridity-limited conditions. Thus, in these situations, the machine learning models have associated increased temperature and aridity with decreased probability of extreme growth because this signals a move to a *more* arid situation that is likely *more* fuel-limited. Given that these are disproportionately savanna/grassland regions, it could indeed be the case that warming/drying (particularly in the 1,000 hour dead fuel moisture variable) reduces fuel loads and thus decreases the risk of extreme daily growth.

Category 2 fire-days are far on the moist side of the critical thresholds identified in Fig. 3C and 3D and thus not particularly prone to extreme daily growth. We speculate that in order to achieve extreme daily growth under these conditions, strong atmospheric instability must be present in order to generate a "plume-dominated" forest crown fires [24-26]. However, warming and drying in these situations may be associated with higher lifting condensation levels and synoptic-scale high pressure, both of which would *increase* atmospheric stability, making plume aided spread less likely. This hypothesis can be tested in future work using dynamical fire-atmosphere coupled models [27].

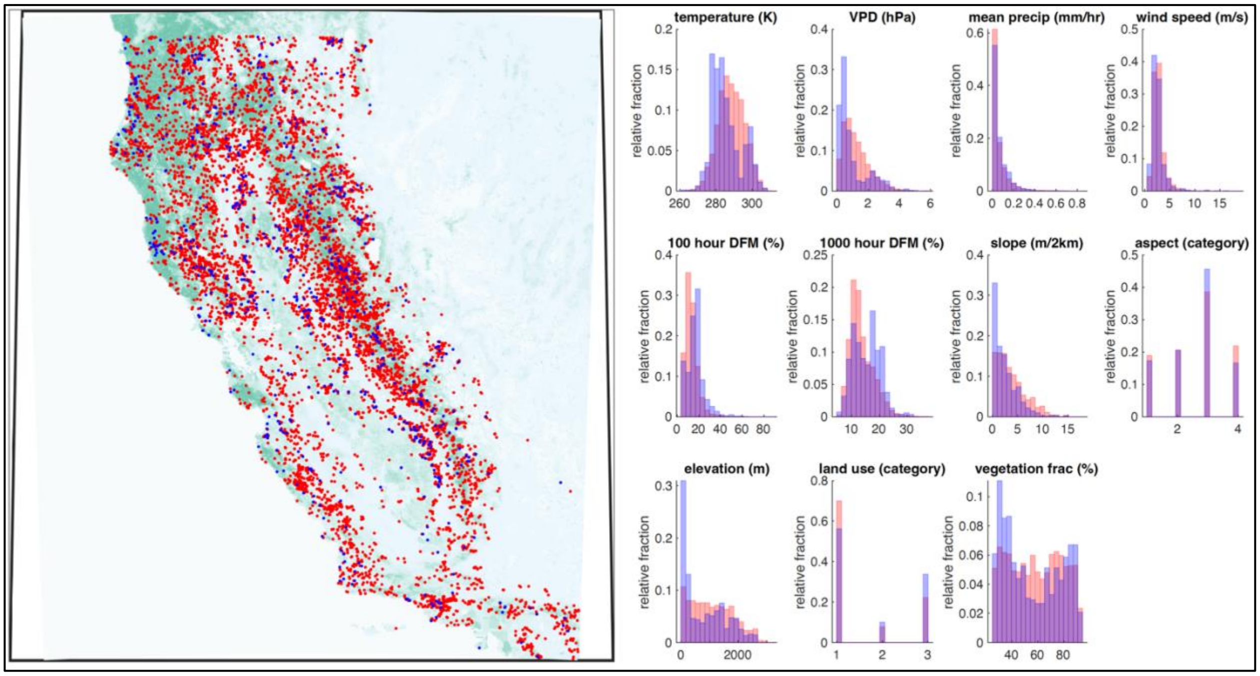

Fig. S7. Comparison of fire-days with historical vs. preindustrial probability ratios above 1 (red) and below 1 (blue).

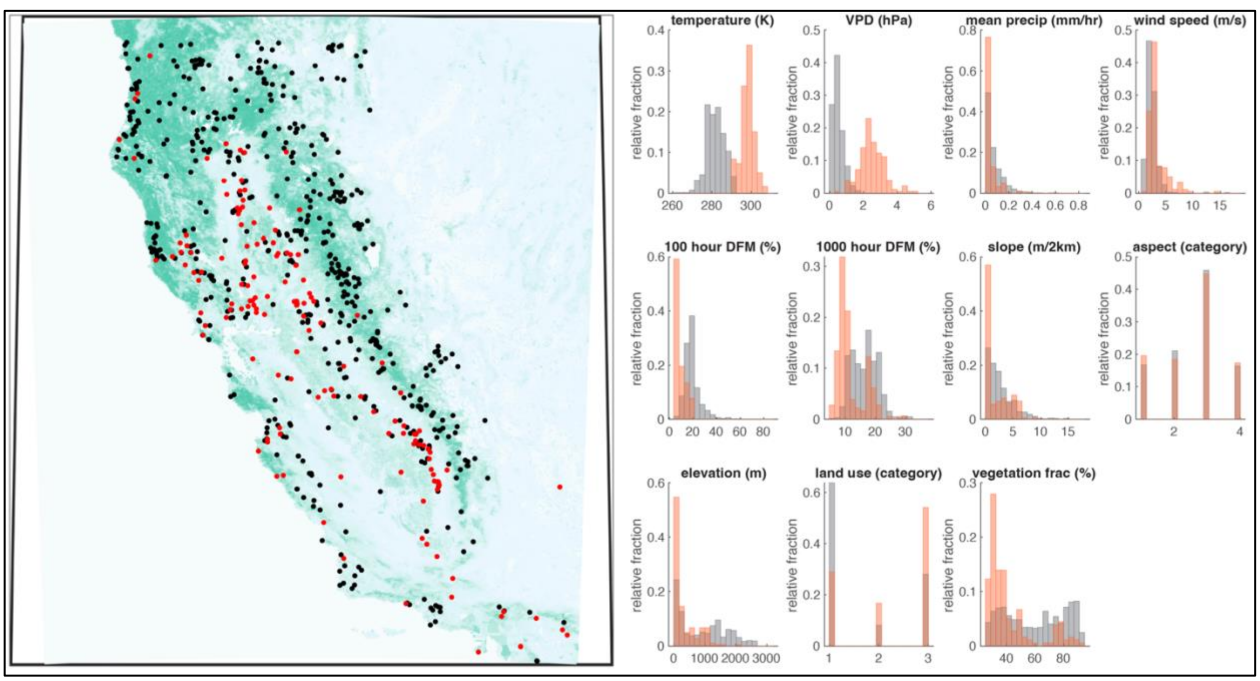

Fig. S8. The below 1 probability ratio fire-days from S7 split into categories of above 291 K (red) and below 291 K (black).

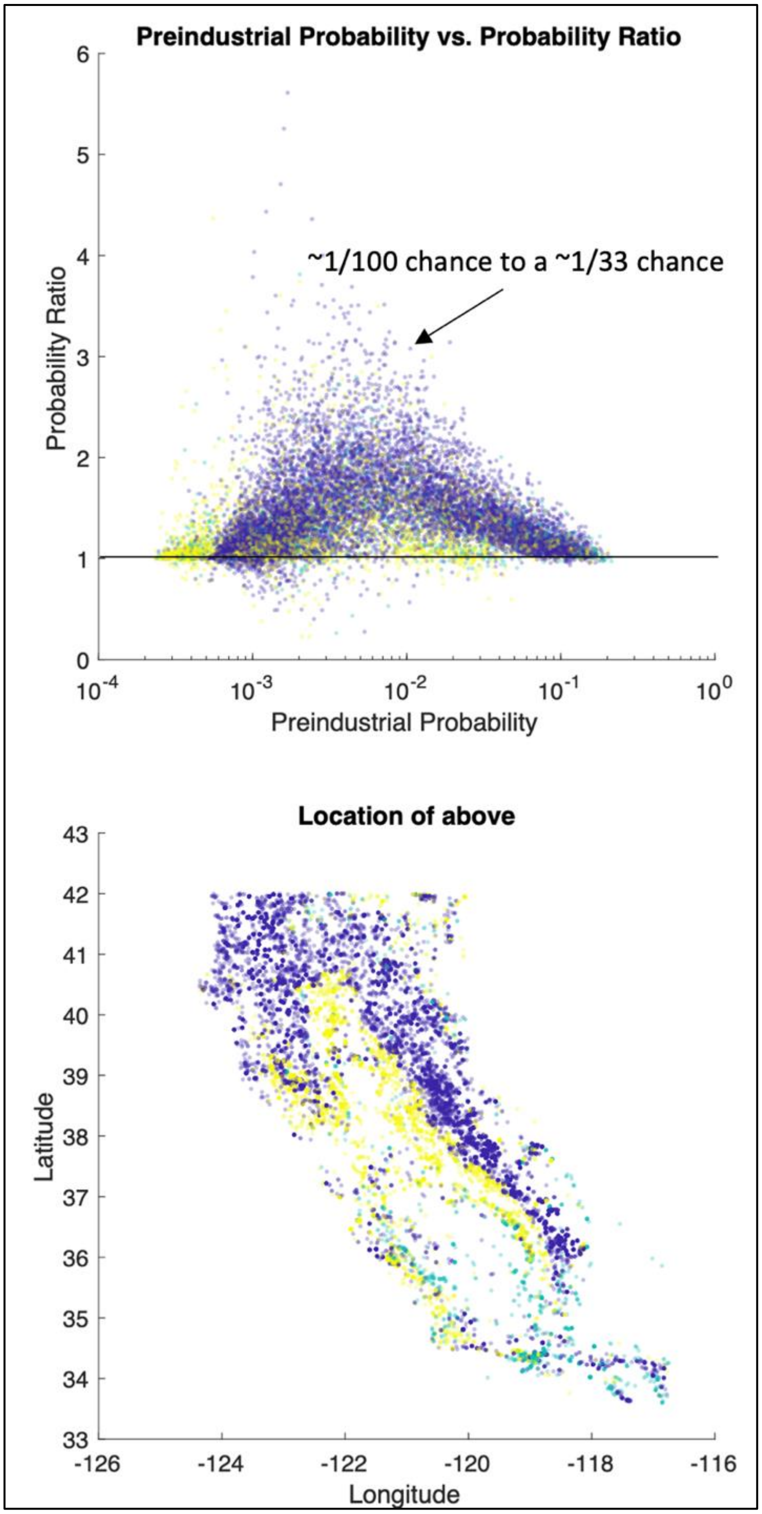

Fig. S9. Probability ratios as a function of preindustrial probability of extreme daily growth. The fire-days with the largest probability ratios tend to have preindustrial probabilities (of extreme daily growth) between 0.1% and 2%. This is roughly the probability of extreme growth for a fire-day just on the moist side of the identified critical thresholds (i.e., it is weather or aridity limited [26]), and thus anthropogenic warming causes large relative shifts in probability.

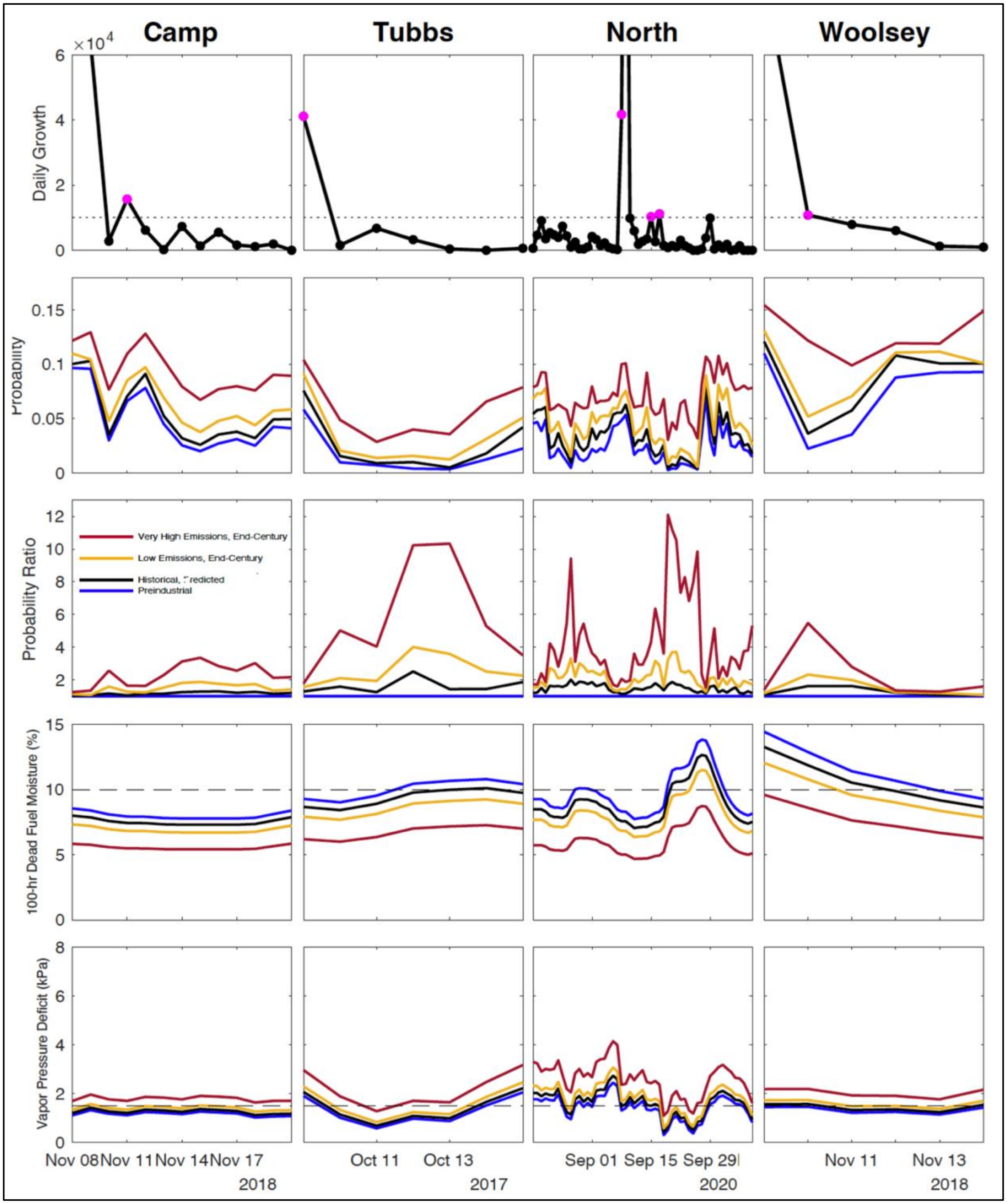

Fig. S10. Same as Fig. 4 but for additional notable fires.

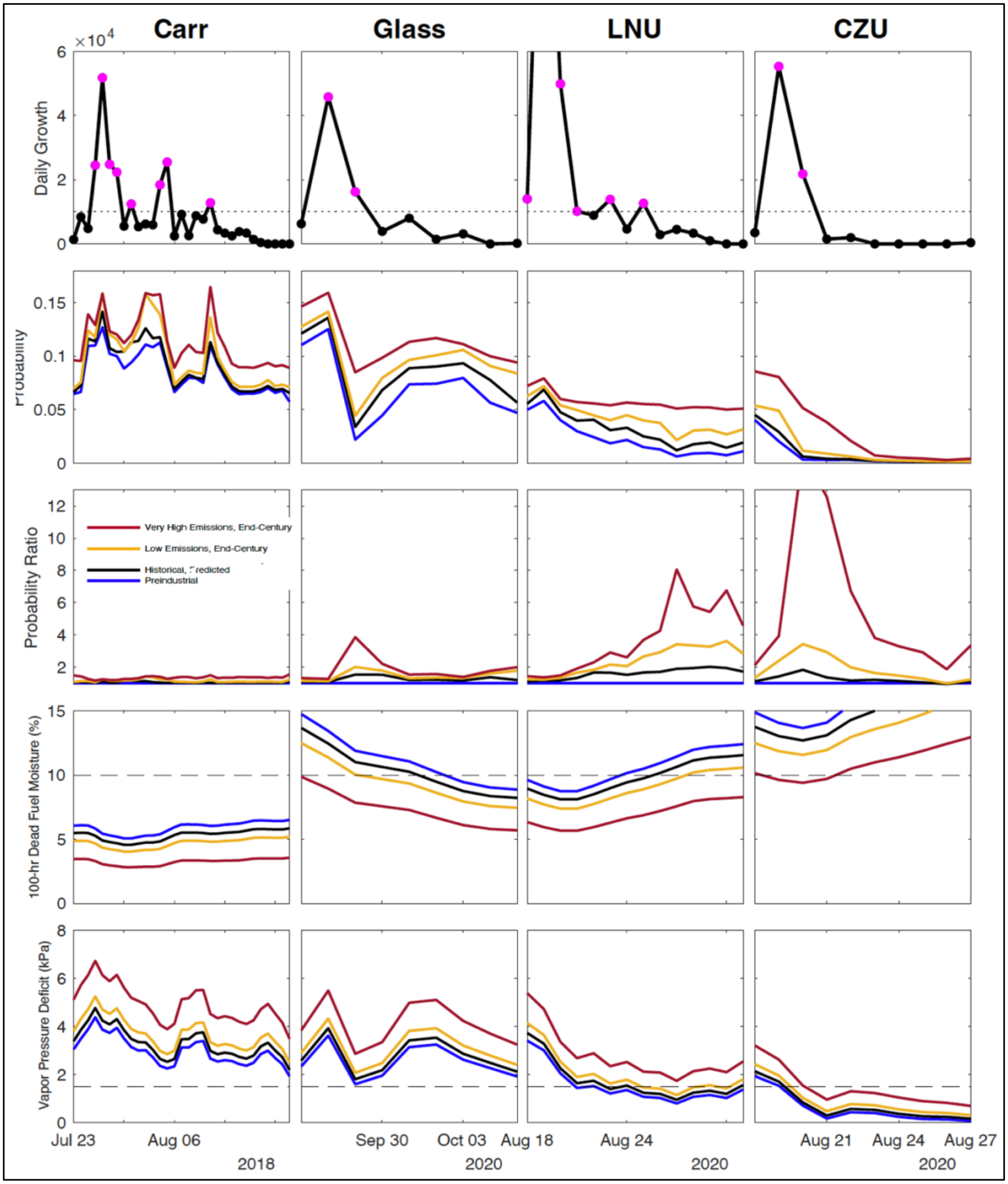

Fig. S11. Same as Fig. 4 but for additional notable fires.

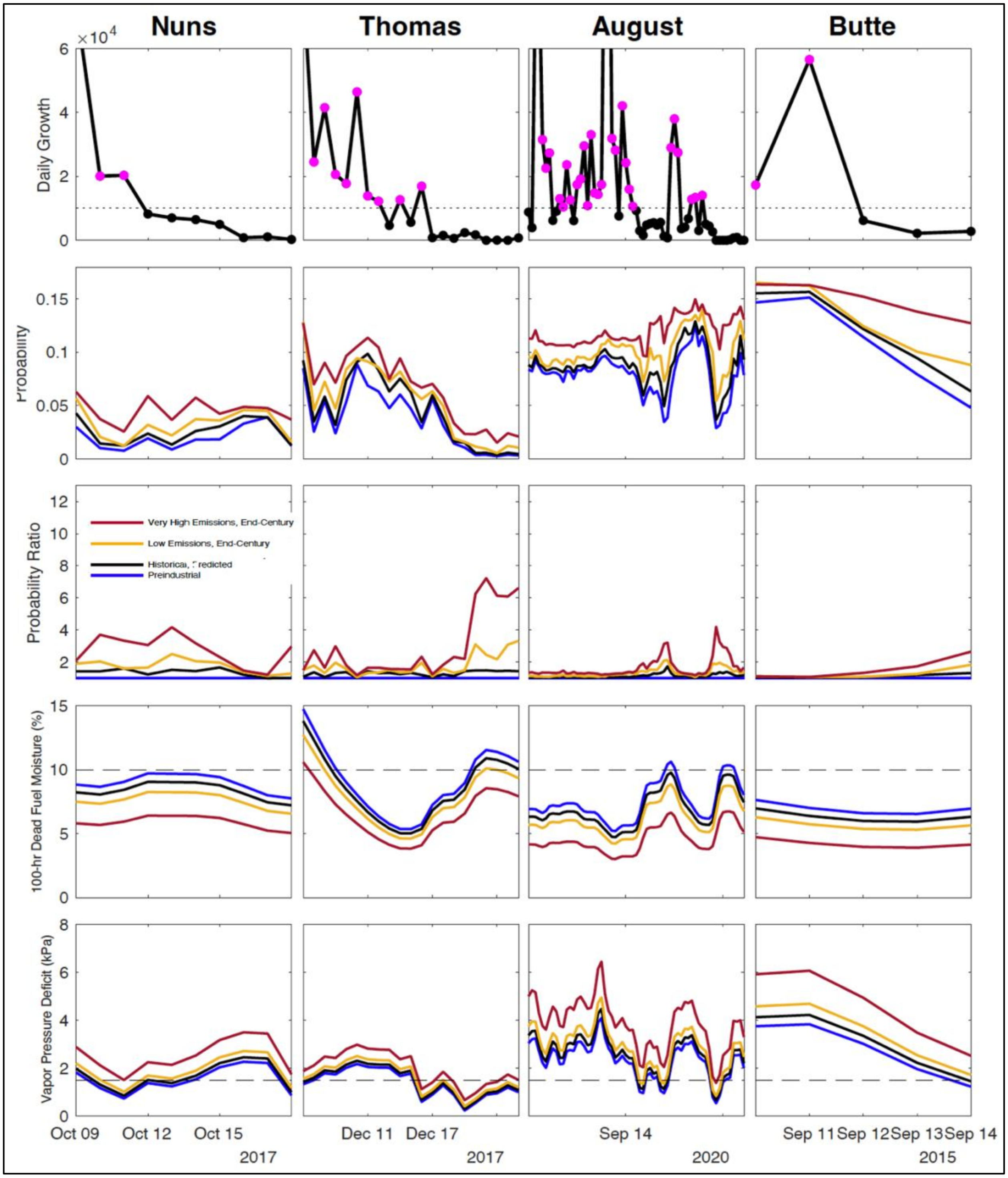

Fig. S12. Same as Fig. 4 but for additional notable fires.

S21. Response Functions

Figure S13 shows 'response functions' for temperature and the three aridity variables that are the focus of this study. Response functions are calculated by holding the values of all predictors constant (at their mean values in the dataset) and plotting how the probability of extreme growth changes as a function of the value of a single predictor.

Figure S13A shows two versions of the response function for temperature; one where temperature is allowed to propagate into the three aridity variables (black) and one where it is not (magenta). This is another way of visualizing the finding that it is temperature's effect on aridity, not temperature per se, that is the dominant influence on the risk of extreme daily growth.

The response functions are particularly steep around 1.5 kPa of vapor pressure deficit (Fig. S13B) and around 10% for 100 hour and 1,000 hour dead fuel moisture (Fig. S13C and S13D). Each one of these values is approximately where the response function crosses the 1-in-200 chance of extreme growth. These response functions are thus a better visualization of the nonlinearity or critical thresholds inferred from Fig. 2C, Fig. 2D and Fig. 4. though they are not directly comparable because the values of the other predictors are not necessarily at their mean value in Fig. 2C, Fig. 2D and Fig. 4 (as they are for the response functions). These critical threshold values are also very close to the thresholds identified in a more formal manner (using independent data and methodologies)[28]. That study identified 1.5 kPa for vapor pressure deficit, 9.4% for 100 hour dead fuel moisture, 12.5% for 1000 hour dead fuel moisture as thresholds for large growth in Southern California.

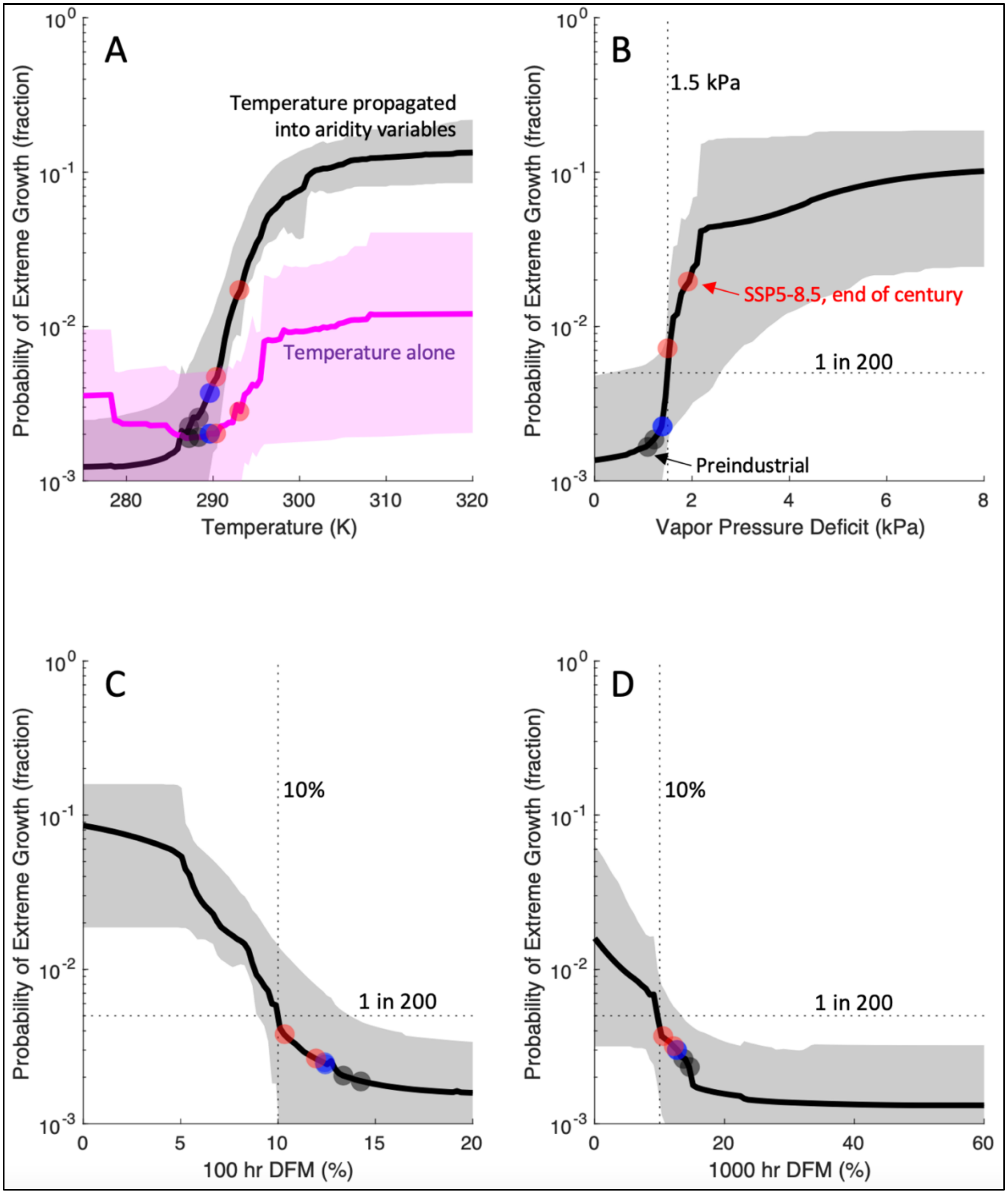

Fig. S13. Response functions for the risk of extreme daily growth (>10,000 acres). In each case, 10 predictor variables are held constant at their mean values (across all fire-days) and only the predictor value on the x-axis is varied. Solid lines are means and the shaded range is the 2σ range across the top 10% of machine learning models (in terms of predictive performance on out-of-training-sample data, Section S1). Circles represent the average value of the variable across all fire-days for the different emissions scenarios and time periods.

S22. Results in the Context of Process-Based Physical Models

Process-based physical modelling has advantages over statistical modelling in that the underlying equations are meant to describe the physics governing the system and are not purely associational. This is particularly advantages when causality is being assessed.

High resolution fire models like WRF SFIRE[16] (meter spatial scale and minute timescale) explicitly simulate much of the physics relevant to assessing the causality of extreme wildfire behavior, including extreme daily growth rates. However, it is practically difficult to use such models in climatological studies because simulating each individual fire-day is computationally expensive in its own right and here we are studying 17,910 fire-days under eight different background climate conditions.

Climate change studies are possible using the fire modules in global dynamic vegetation models. However, it is difficult to use these models to directly study the risk of extreme daily growth rates of individual historical fires (the topic of our study). For example, LPJ-SPITFIRE[29] is driven by weather variables aggregated to the monthly timescale and 0.5°×0.5° spatial scale and it outputs monthly area-burned within grid cells (though internal calculations are at the daily timescale).

Other similar models are included in the Fire Model Intercomparison Project (FireMIP[30]). Though their results are at a lower spatial resolution (Fig. S14) and temporal resolution (Fig. S15 and Fig. S16, also Table 1 in ref.[31]) than ours, and thus are not directly comparable, we briefly analyze relevant output over our domain as a first order check to see if there are any obvious inconsistencies with our results.

This analyses shows area-burned under a standardized set of temporally varying historical drivers consisting of atmospheric CO_2, climate, lightning, land use and population density (See Table A1 of ref.[32] for references for each land/vegetation and fire model).[32]

Even when forced by a standardized set of drivers, the process-based models do not agree on the climatological value of area-burned or the long term direction of the change in area-burned over our domain (Fig. S15). Similarly, there is a wide spread in the average fire size and the magnitude of change in the average fire size over time (Fig. S16). However, when the influence of climate change is isolated from other drivers, area burned is consistently projected to increase over our domain, though the magnitude of increase spans a broad range (Figure A7 in ref.[32]).

Overall, this brief analysis provides a few pieces of information relevant to our study:

1) The current generation of process-based fire modules included in global dynamic vegetation models calculate fire characteristics at a coarser spatiotemporal scale than we focus on in our study and thus do not provide a direct point of comparison. This is a motivation to use alternative approaches like our machine learning approach to address our research question.

2) The coarser spatiotemporal resolution of these models entails parameterizations of important subgridscale processes and the lack of agreement on the climatological characteristics and sign of change of e.g., area-burned between these models indicates that parameterization and structural model uncertainty are large. This further suggests that there is significant room for statistical models like those used here to provide valuable information, a conclusion that has been reached elsewhere[33].

3) The fact that these process-based models generally indicate that historical climate change should have increased area-burned (Figure A7 in ref.[32]) is consistent with our results of increased probability of extreme growth, especially considering that area-burned is disproportionally impacted by very large fires that experience this extreme growth[34].

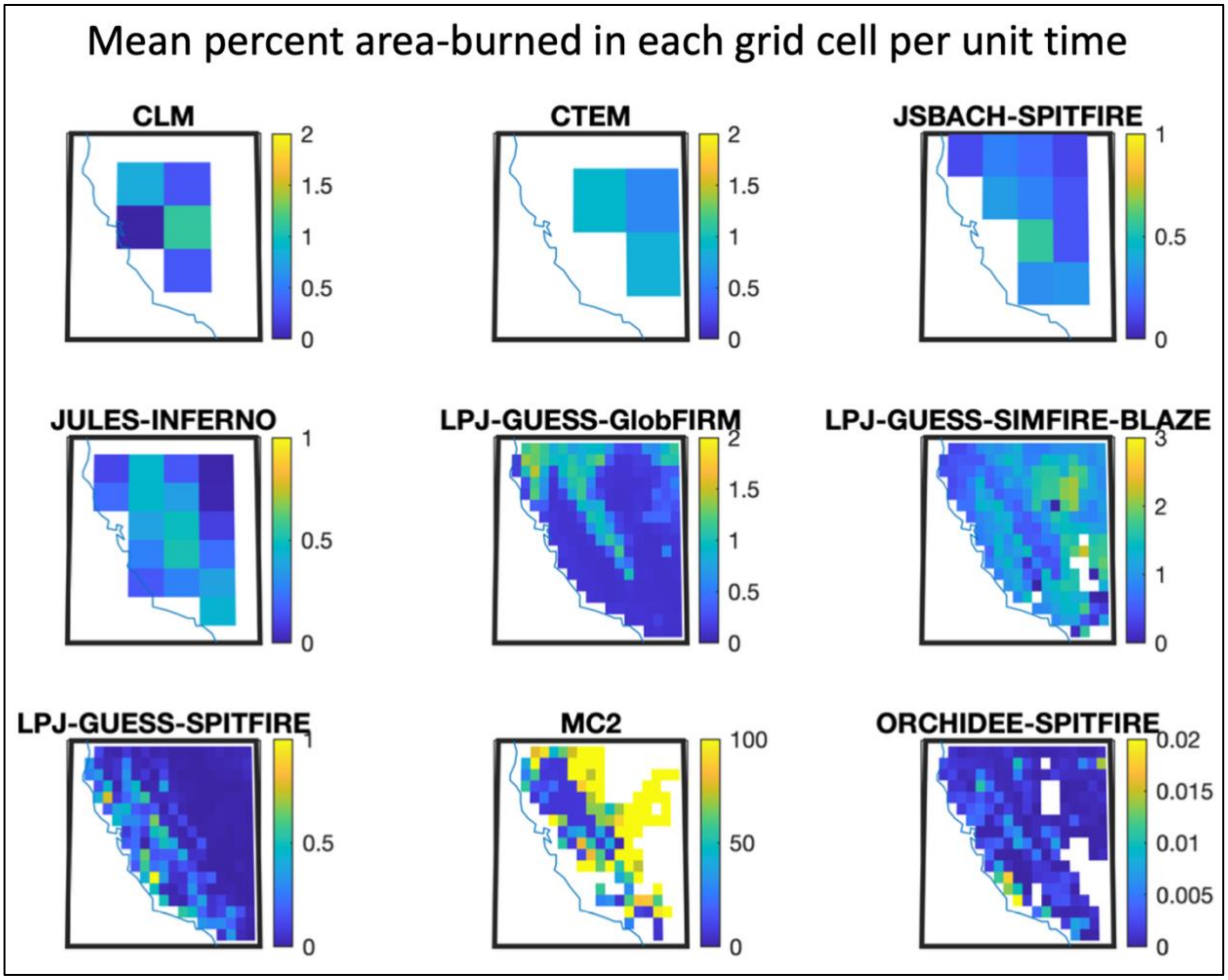

Fig. S14. Mean percent area burned in each grid cell per unit time (time is years for LPJ-GUESS-GlobFIRM and MC2 and months for all others) over the time periods shown in Fig. S15.

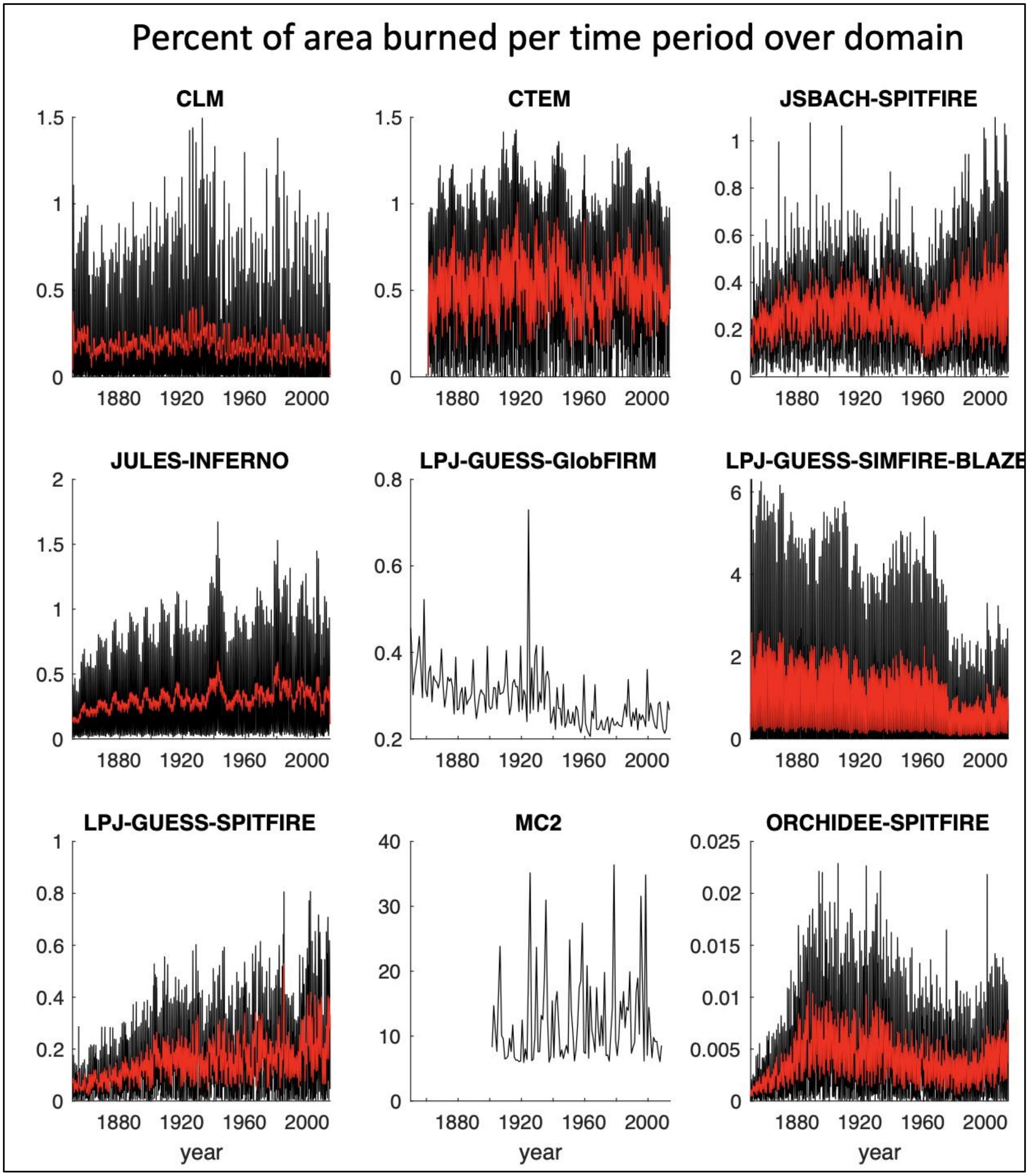

Fig. S15. Percent of domain burned per time period (time period is years for LPJ-GUESS-GlobFIRM and MC2 and months for all others). A 12-month running mean is added to the monthly time series and shown in red.

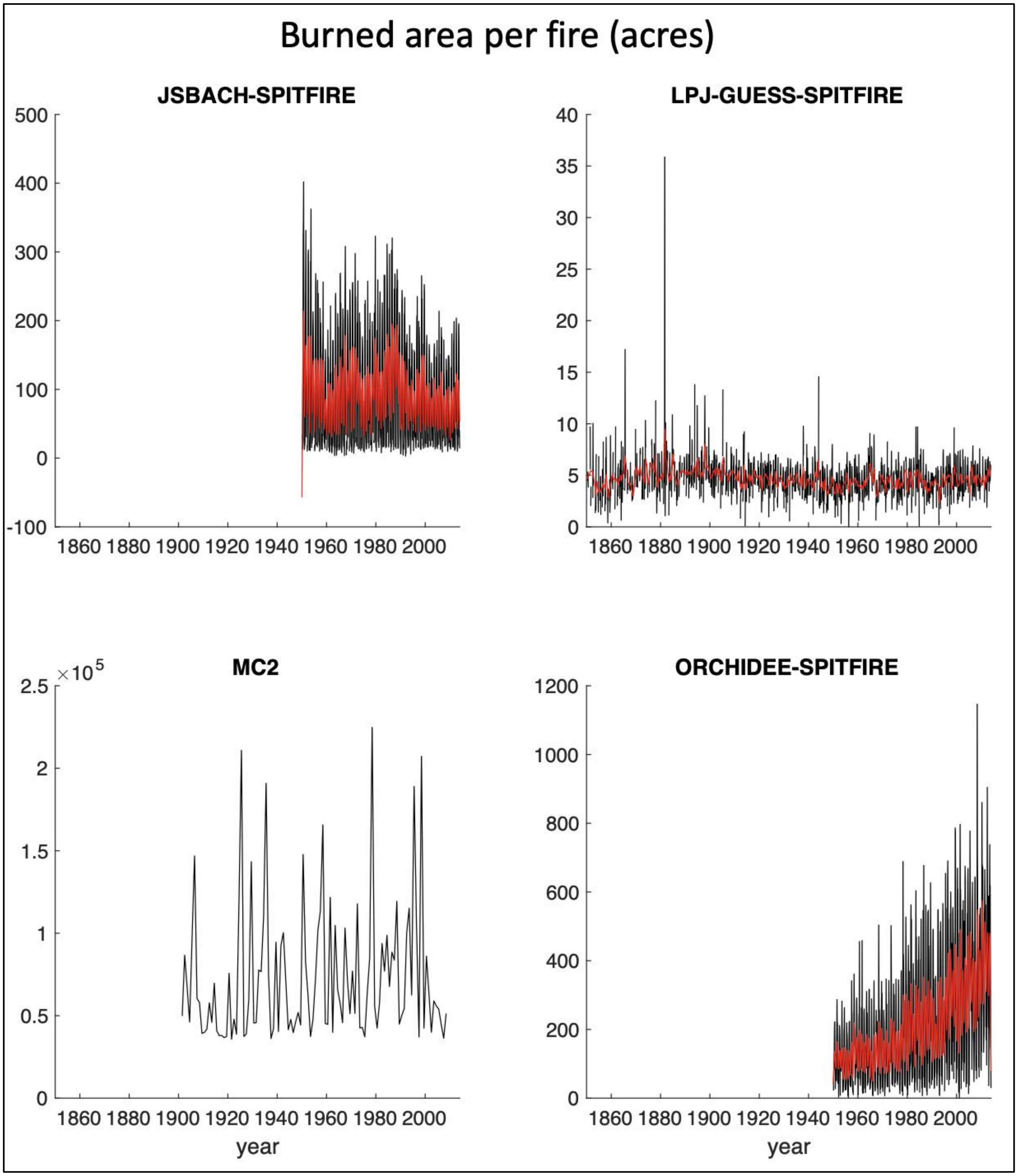

Fig. S16. Burned area per fire for four models in which this calculation was output.

S23. Results in the Context of Changes in Conditions Conducive to Wildfires and Fire Weather Indices

Because of the current limitations of process-based models, the impact of increasing greenhouse gas concentrations on wildfire activity is often studied in terms of environmental conditions conducive to wildfires[35]. For example, Ref.[36] found historical increases in warm season maximum temperature, vapor pressure deficit, and decreases in 1000 hour dead fuel moisture in our domain. Also, ref.[37] found that anthropogenic climate change enhanced eight metrics related to fuel aridity over our domain including potential evapotranspiration, vapor pressure deficit, climatic water deficit, and Palmer Drought Severity Index. Looked at another way, ref.[38] found an increase in the annual time period in California for which conditions are conducive to wildfires (wildfire season length) of roughly 35 days between 1979 to 2015. This was due mostly to increases in daily maximum temperature, and decreases in minimum relative humidity with small contributions from increases in rain free days and increases in maximum windspeed. Studies similar to our own, where changes in environmental conditions are translated into changes in wildfire activity via a statical model, also show substantial increases in fire activity under future warming scenarios. For example, ref.[39] found that environmental changes associated with the RCP8.5 scenario would amount to a 50% increase in the frequency of fires growing to size very large (24,710 acres).

Individual environmental variables are often combined to create fire weather or fire danger indices whose changes are then studied. Over the US West, there have been positive trends in the Canadian Forest Fire Danger Rating System’s Fire Weather Index, The US National Fire Danger Rating System’s Energy Release Component, the McArthur Forest Fire Danger Index, and the Keetch–Byram Drought Index[37]. Studies generally show increases in mean fire weather index values as well as the frequency of extreme daily fire weather index values over our region[19,35,40-42].

Days of extreme wildfire growth are associated with particularly high daily values of fire weather indices[36]. Though not directly comparable, it is still useful to place our results in the context of projected changes in the frequency of these days with particularly high values. Fig. S17 shows the domain mean frequency of annual occurrences of “very high” fire danger days by mid-century from 18 downscaled CMIP5 climate models, under the RCP8.5 emissions scenario[43] (https://climatetoolbox.org/).

The mean increase across models of 34% (red line) is similar to our calculated 45% increase in the occurrence of extreme daily growth (524 vs. 362, Fig. 2G) for mid-century under the SSP5-8.5 emissions scenario. None of the results discussed here are directly comparable to ours but they indicate similar magnitudes of change and thus they appear consistent with our results.

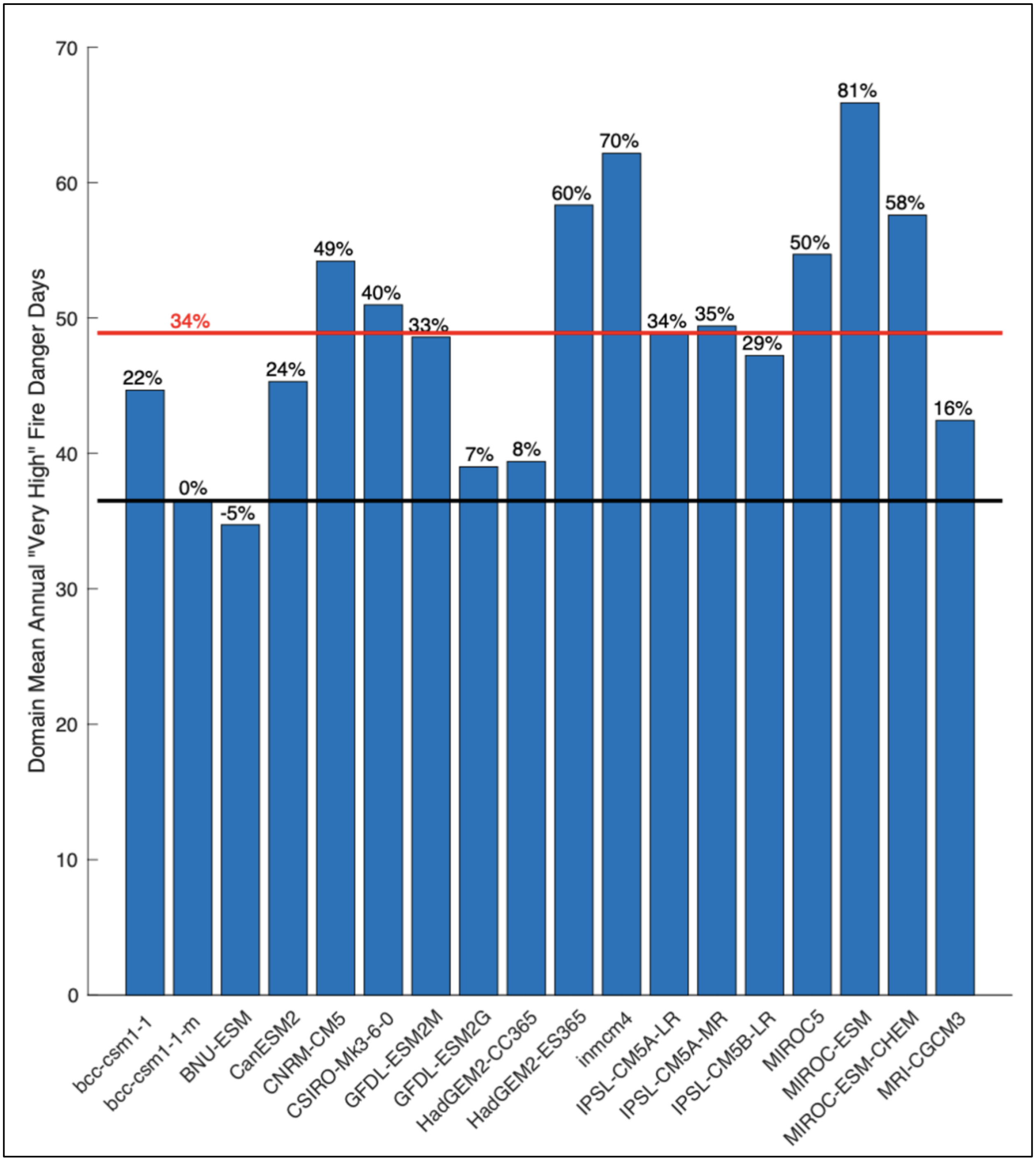

Fig. S17. Domain mean number of annual occurrences of "very high" fire danger days from a downscaling method applied to the output of 18 CMIP5 models. The black line represents the baseline of 36 days per year over the 1971-2010 time period. Each bar represents the mean value simulated over the period 2040-2069 under the RCP8.5 emissions scenario. The mean increase of 34% (red line) is similar to our calculated 45% increase in the occurrence of extreme daily growth (524 vs. 362, Fig. 2G) for mid-century under SSP5-8.5 emissions scenario. (https://climatetoolbox.org/).

S24. The Value of Allowing Predictors to Interact Via Machine Learning Models

We conducted an experiment to test whether there was substantial value added by allowing predictors to interact through the architecture of the random forest and neural network models as opposed to simply using a predefine mathematical formula or fireweather index. Specifically, we calculated ROC-AUC scores using three of the most commonly used fire weather indices: Sharples Fire Weather Index, Fosberg Fire Weather Index, and Hot-Dry-Windy as predictors. Since these predictors have a single value they require a logistic regression model to calculate probabilities of extreme growth. ROC-AUC scores were 0.586 for Sharples, 0.549 for Fosberg, and 0.608 for hot-dry-windy, all well below the performance of the top random forest and neural network models which average 0.89. This justifies the use of these more flexible machine learning methods that are not constrained by simple mathematical formulas.

S25. Lower Threshold Value for the Definition of Extreme Growth

We test the sensitivity of our results to a lower growth threshold for the definition of extreme growth which allows for a higher number of extreme events to be in the sample with the inherent tradeoff that the events being studied are less extreme. Our test consisted of lowering the threshold by half from >10,000 acres of growth in a single day to >5,000 acres. This had the effect of changing the number of extreme events in our dataset from 380 (2.1% of 17,910 total fire-days) to 906 (5.4% of 17,910 total fire-days).

We find that skill is comparable between the two sets of experiments. For the top 10% of models, the log-loss skill score decreased from 0.22 for predicting >10,000 acres growth to 0.21 for >5,000 acres growth, the Brier skill score increased from 0.064 to 0.104, the reliability diagram score stayed the same at 0.88, and the ROC-AUC decreased from 0.89 to 0.85.

Figs. S18 – S20 show the main results of the study recalculated while using the >5,000 acre growth threshold.

We find that anthropogenic warming causes a larger enhancement of the probability of the more extreme class of events. Specifically, the average probability ratio under the historical period is 1.22 for >5,000 acres of growth (Fig. S18A) compared to 1.33 for >10,000 acres of growth (Fig. 2A, Also see Table S4). This indicates that anthropogenic warming increases the probability of extreme growth nonlinearly where the more extreme the growth rate the more it's probability is enhanced by warming. This is the same characteristic that is observed for other phenomena like extreme rainfall[12] and temperature[44].

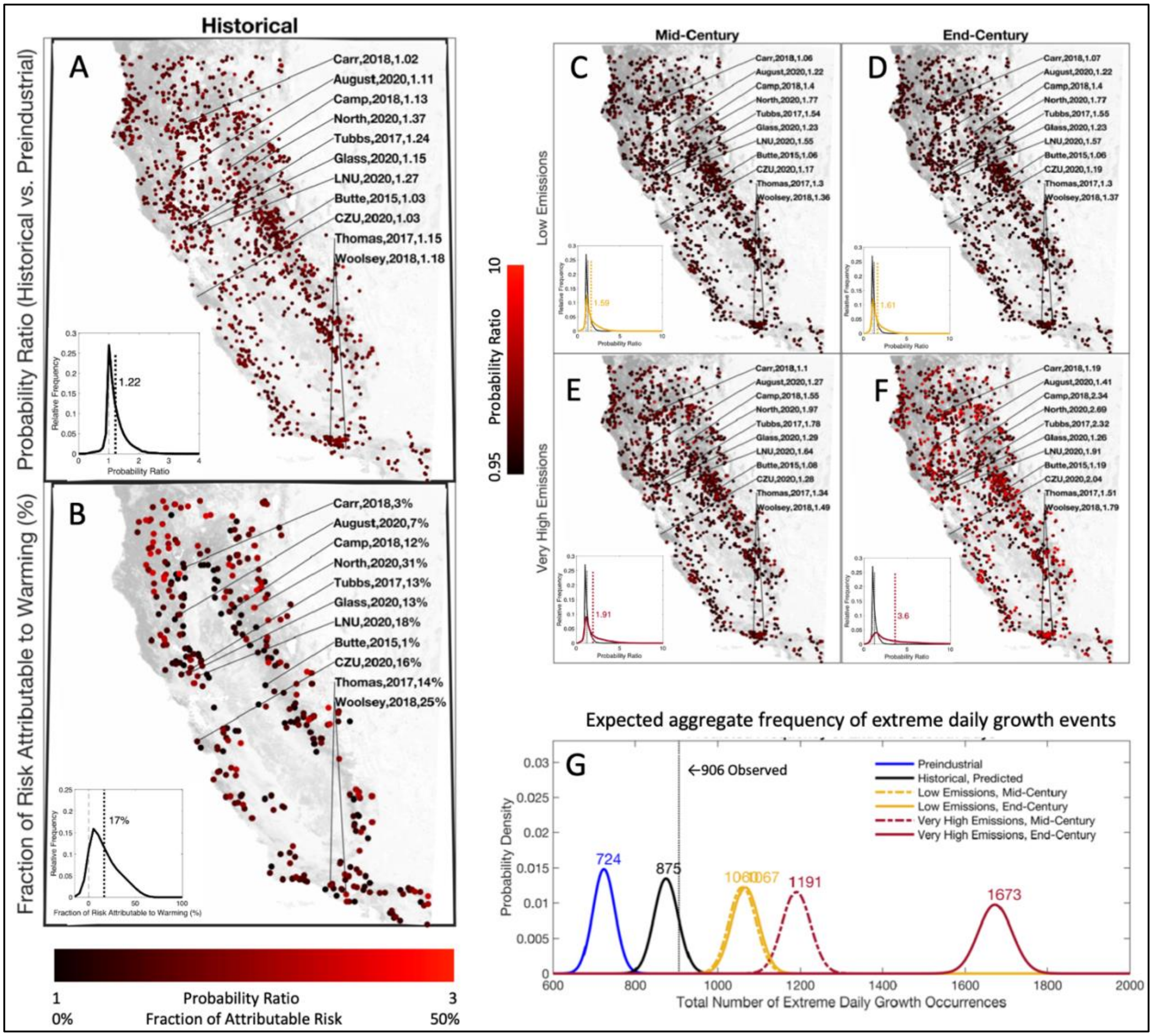

Fig. S18. Same as Fig. 2 but with extreme growth defined at >5,000 acres in a day rather than >10,000 acres in day.

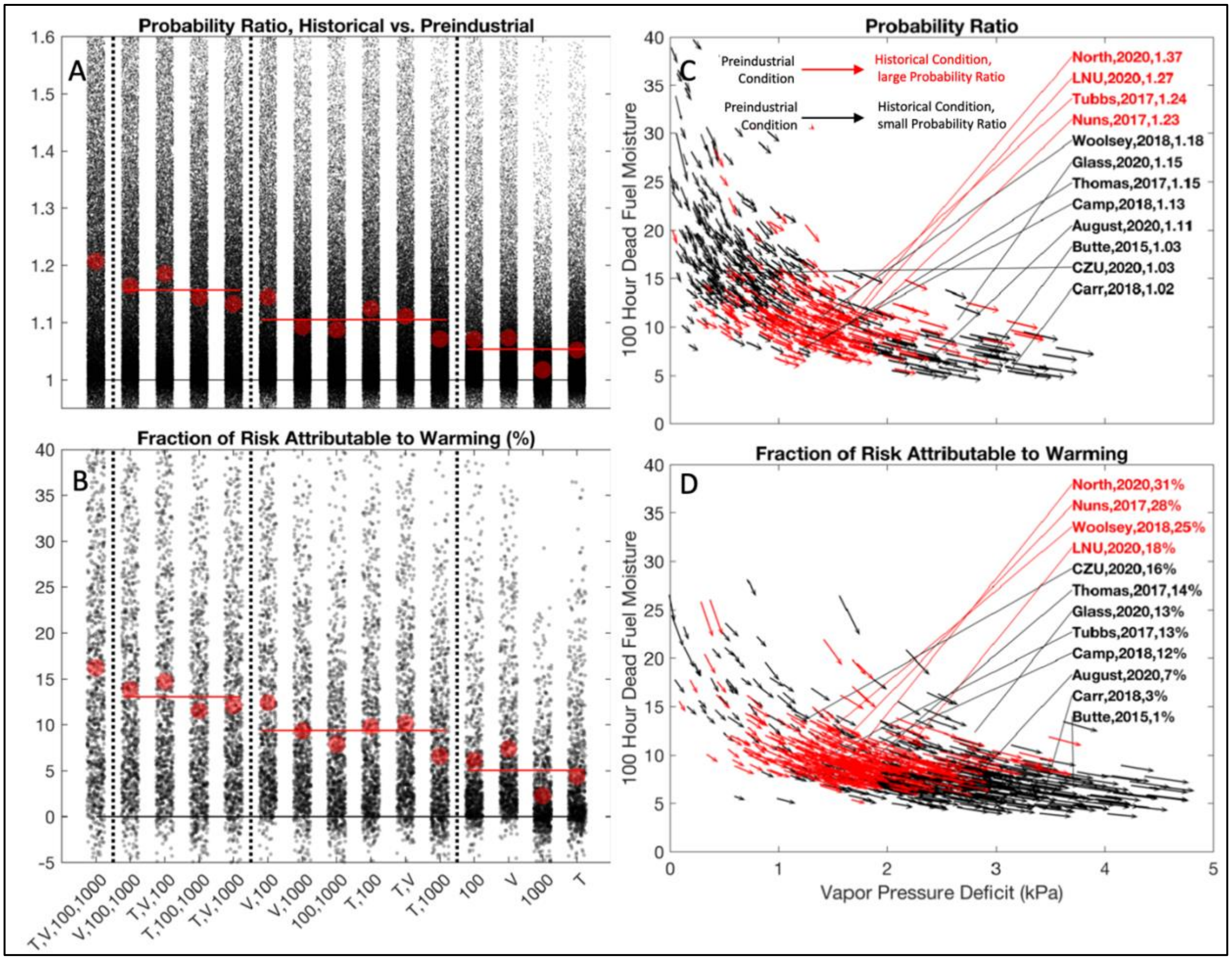

Fig. S19. Same as Fig. 3 but with extreme growth defined at >5,000 acres in a day rather than >10,000 acres in day.

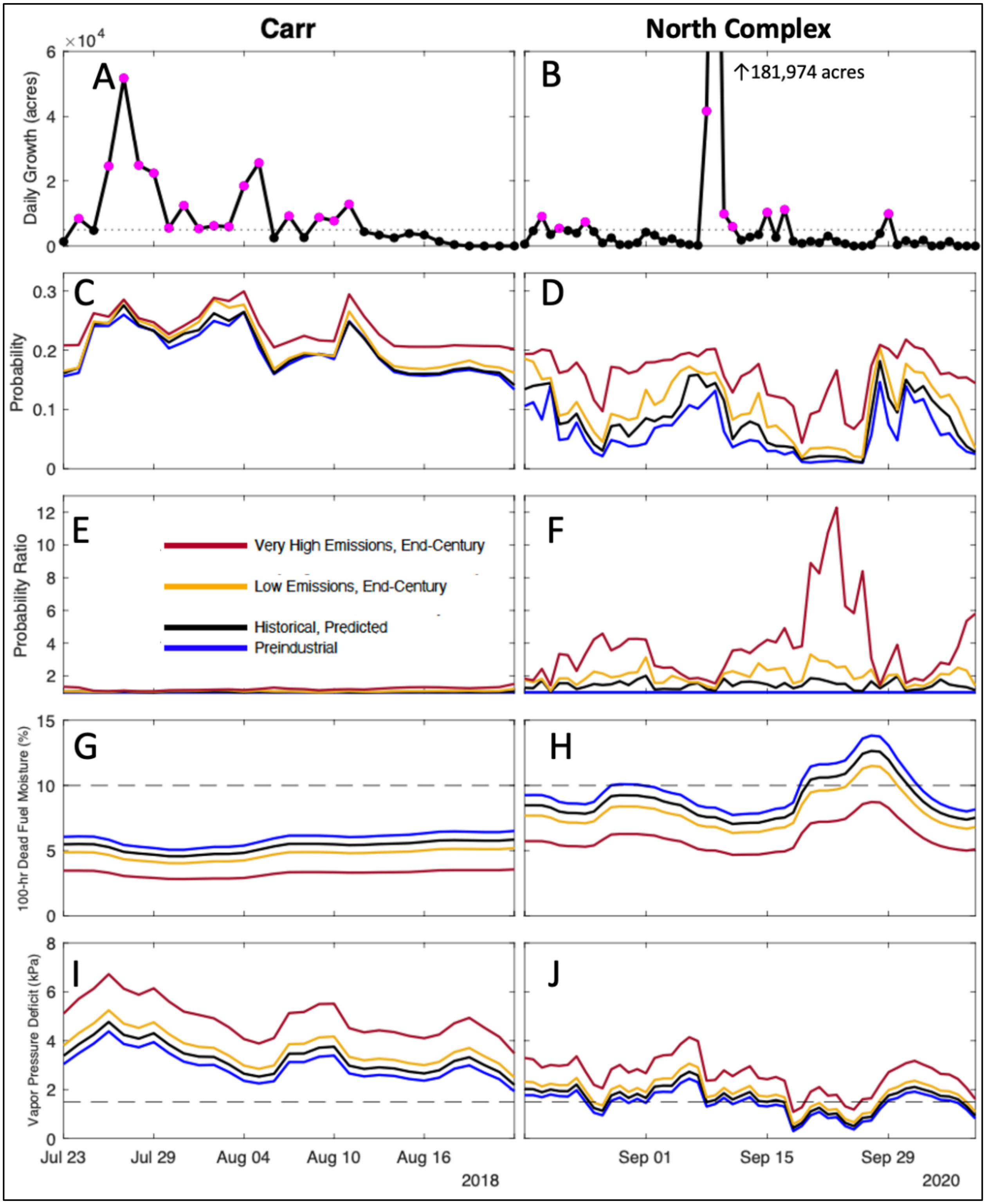

Fig. S20. Same as Fig. 4 but with extreme growth defined at >5,000 acres in a day rather than >10,000 acres in day.

S26. Temperature Change Calculated with Higher Resolution Regional Climate Models

In the main text, background climatological temperature changes are calculated from the mean of CMIP6 models (Table S1). We use CMIP6 models for this calculation because they are the standard tools used for projecting climatological temperature changes in the climate literature. A drawback of CMIP6 models is that they have relatively low spatial resolution. In order to test if this has any meaningful impact on our conclusions, we recalculated our results using the Coordinated Regional Downscaling Experiment (CORDEX, https://cordex.org/). CORDEX has 12.5 by 12.5 km spatial resolution over our domain. The higher resolution can be seen by comparing Fig. S21 to Fig. S2.

Figs. S22 – S24 show the main results of the study recalculated while using CORDEX for the calculation of climatological temperature change.

Overall, we find that our results are very similar between the cases when CORDEX is used compared to CMIP6. For example, the historical mean fraction of attributable risk for large growth days is 19% for CMIP6 but 22% for CORDEX. The mean end-century enhancement of the expected frequency of extreme growth for low emissions is +59% for CMIP6 but +65% for CORDEX. Finally the mean end-century enhancement of the expected frequency of extreme growth for very high emissions is +172% for CMIP6 but +164% for CORDEX (See Table S4).

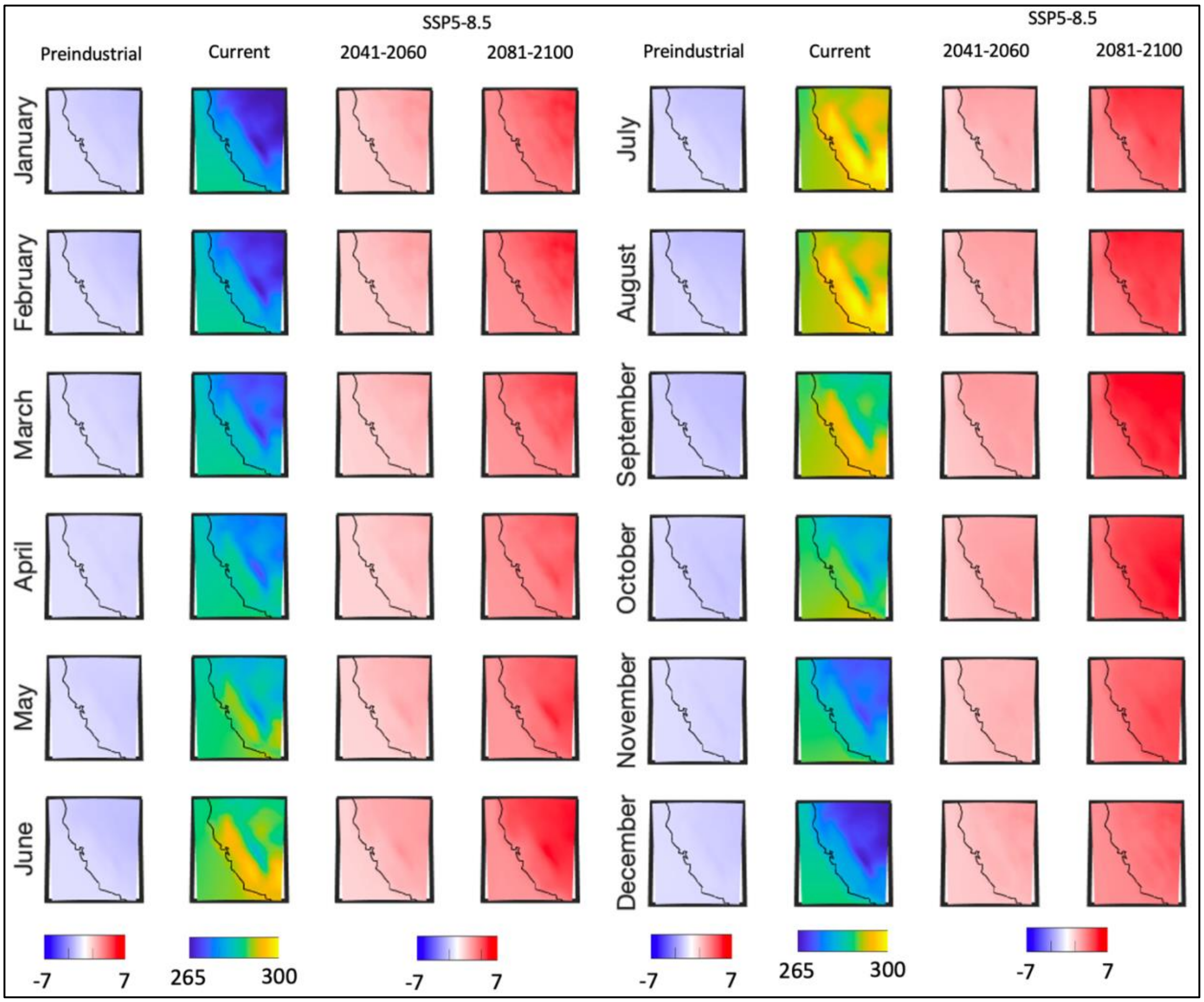

Fig. S21. Same as Fig. S2 but with high resolution CORDEX temperature changes.

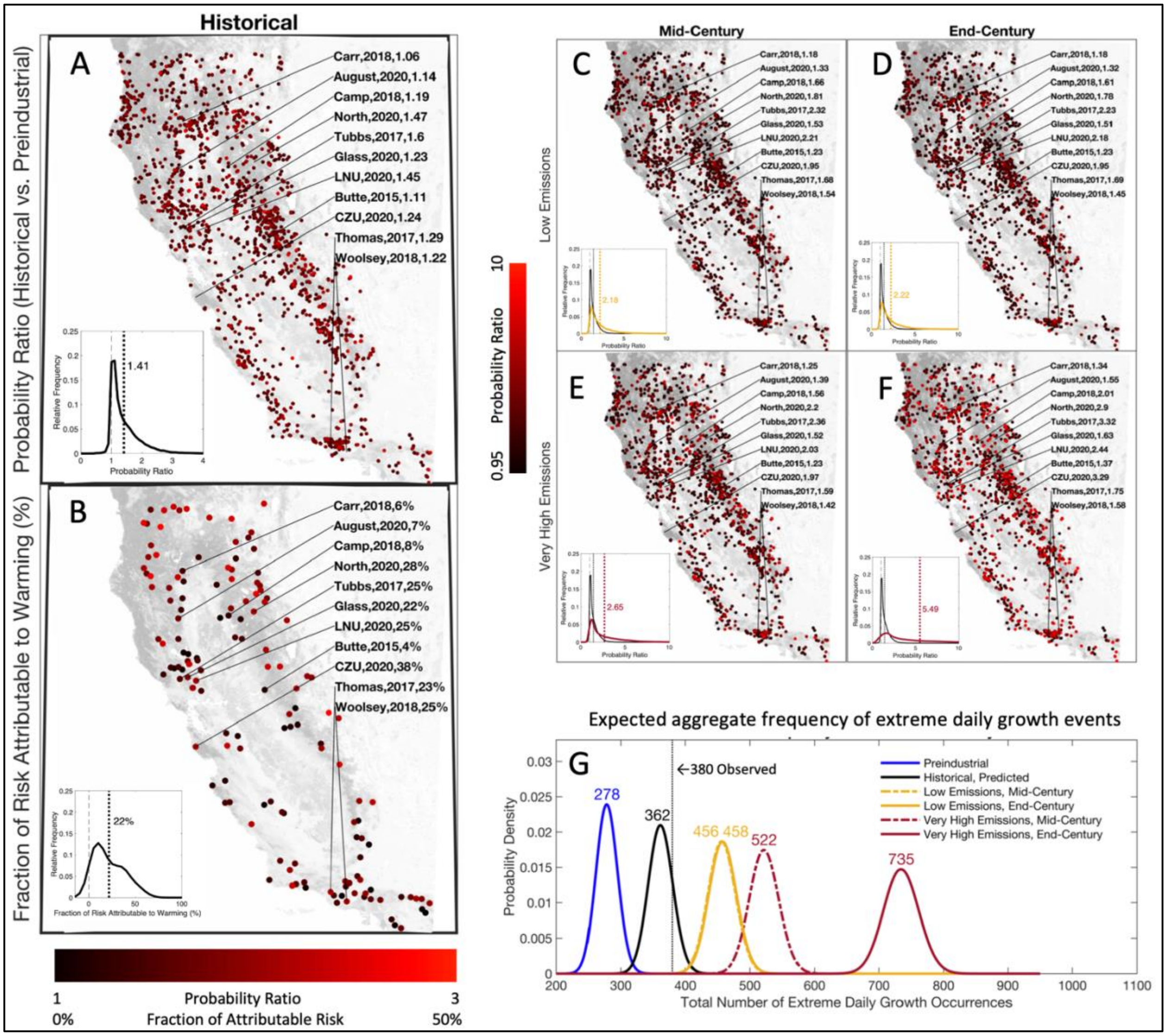

Fig. S22. Same as Fig. 2 but using CORDEX instead of CMIP6 to calculate the change in background temperature.

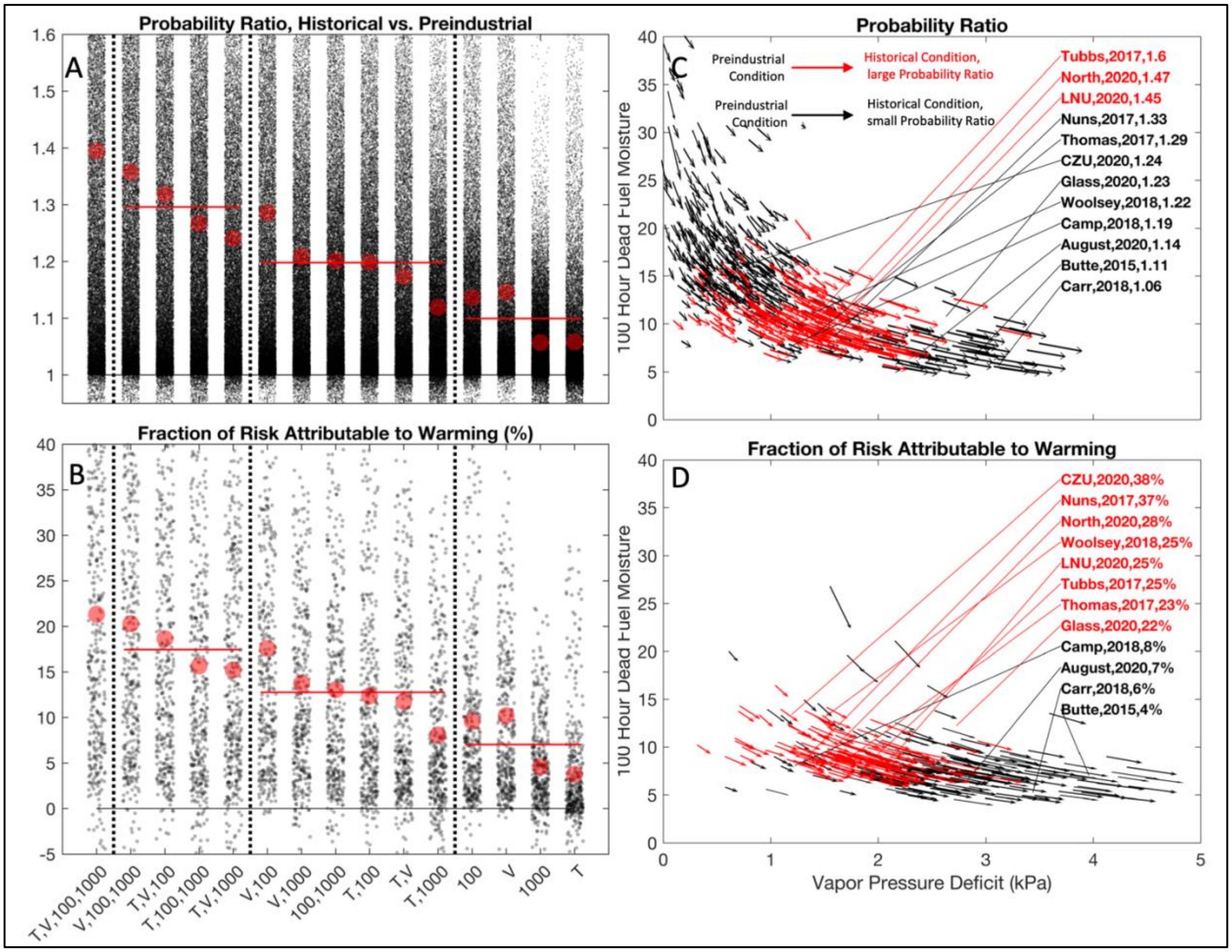

Fig. S23. Same as Fig. 3 but using CORDEX instead of CMIP6 to calculate the change in background temperature.

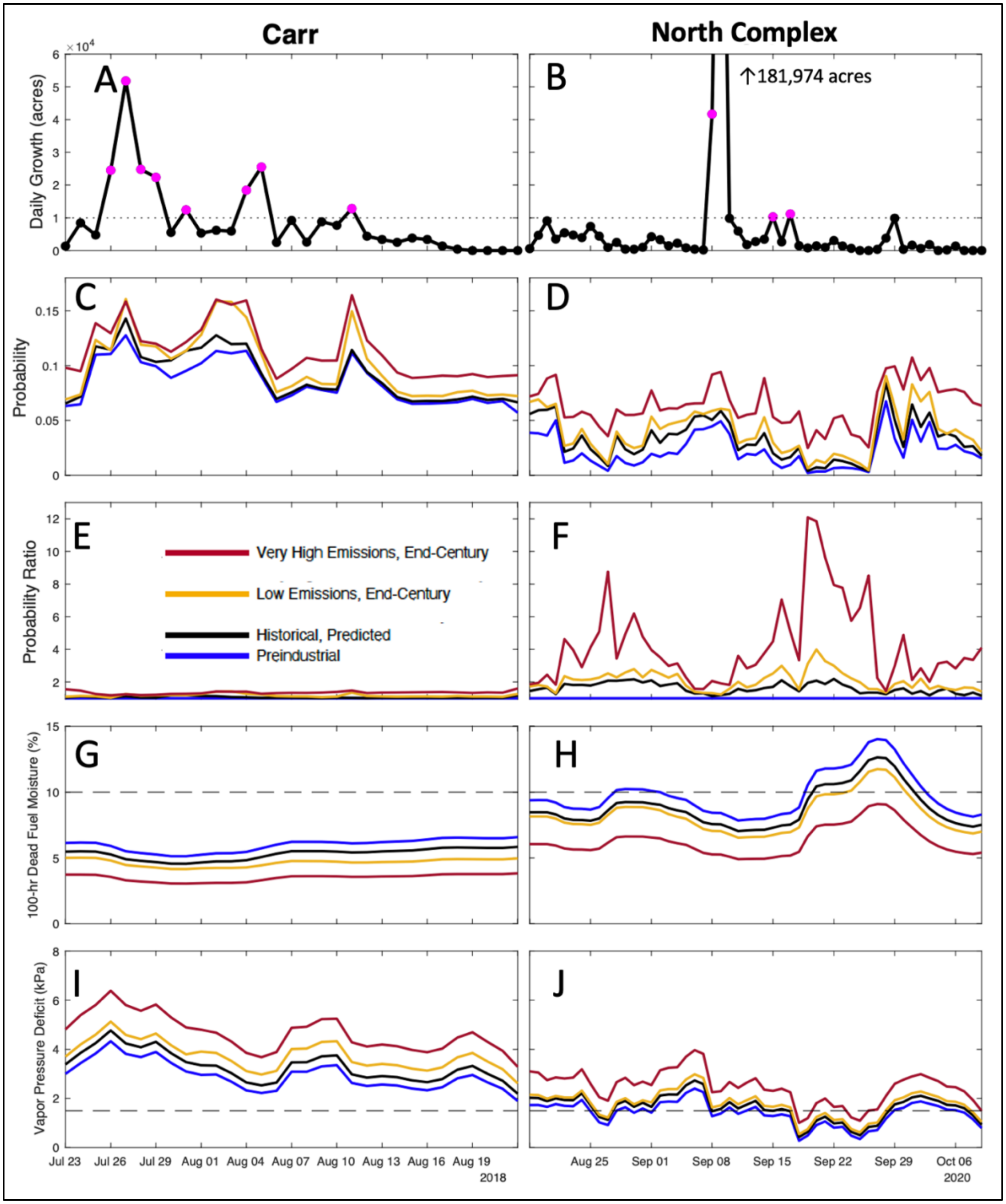

Fig. S24. Same as Fig. 4 but using CORDEX instead of CMIP6 to calculate the change in background temperature.

S27. Replace Vapor Pressure Deficit Predictor with 10-hour Dead Fuel Moisture Predictor

Our analysis was conducted at the daily timescale and thus in the main text we did not include fuel moisture variables with relaxation timescales that were sub-daily (e.g., 10-hour and 1-hour dead fuel moisture). Instead we chose to use daily mean vapor pressure deficit to represent daily variability in aridity because vapor pressure deficit has emerged as a leading predictor or fire proclivity in other studies[45]. However, it is reasonable to ask whether our results are sensitive to using a measure other than vapor pressure deficit to represent aridity of finer fuels.

Figs. S25-S27 show the main results of the study recalculated using daily mean 10-hour dead fuel moisture rather than daily mean vapor pressure deficit as the 'fast response' aridity predictor.

Overall, we find that our results are very similar between the cases when 10 hour dead fuel moisture is used compared to vapor pressure deficit. For example, the historical mean fraction of attributable risk for large growth days remains at exactly 19% when 10 hour dead fuel moisture is substituted for vapor pressure deficit. The mean end-century enhancement of the expected frequency of extreme growth for low emissions is +59% for vapor pressure deficit but + 58% for 10 hour dead fuel moisture. Finally the mean end-century enhancement of the expected frequency of extreme growth for very high emissions is +172% for vapor pressure deficit but +167% for 10 hour dead fuel moisture (See Table S4).

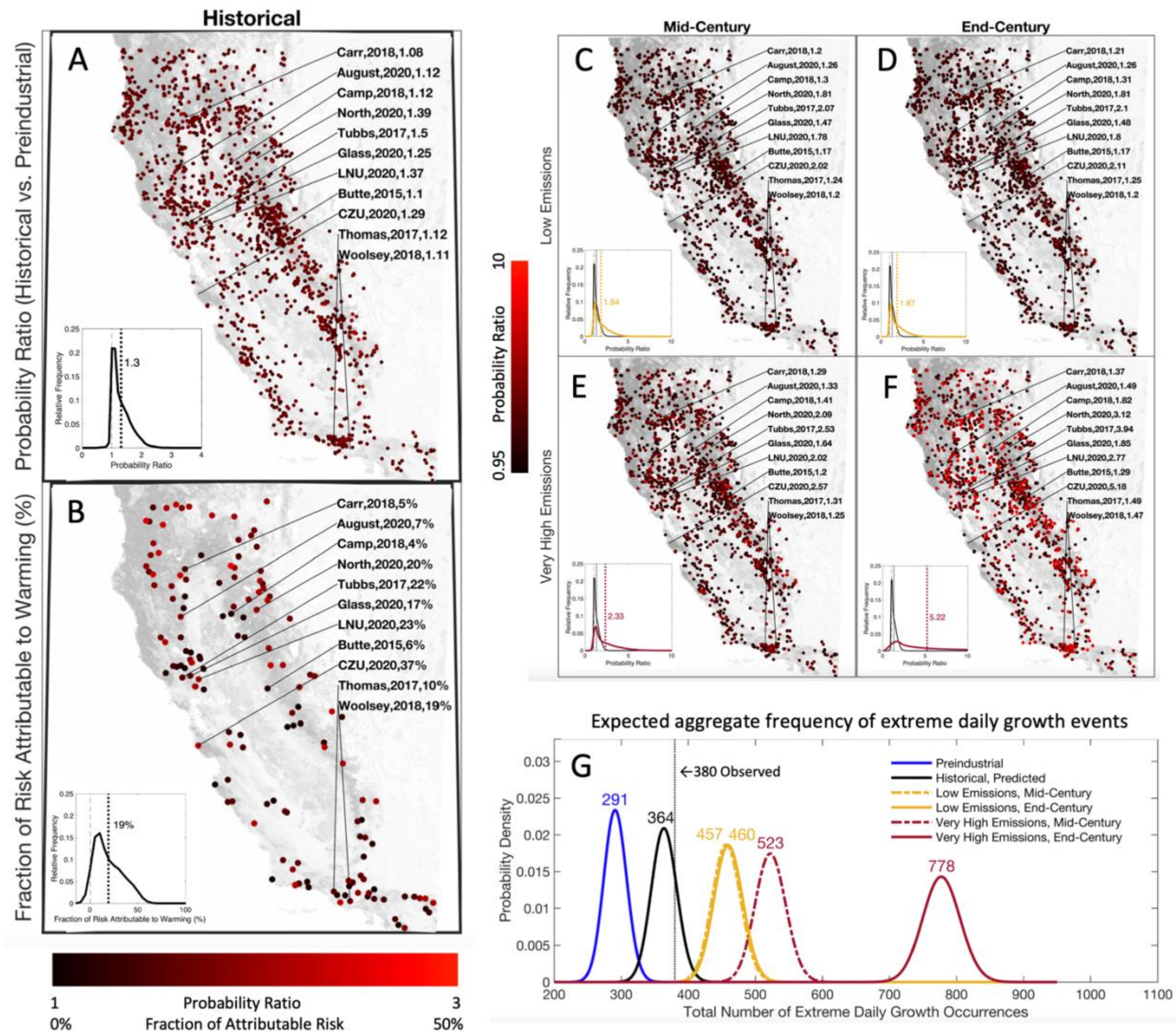

Fig. S25. Same as Fig. 2 but with the daily mean vapor pressure deficit predictor replaced with daily mean 10 hour dead fuel moisture.

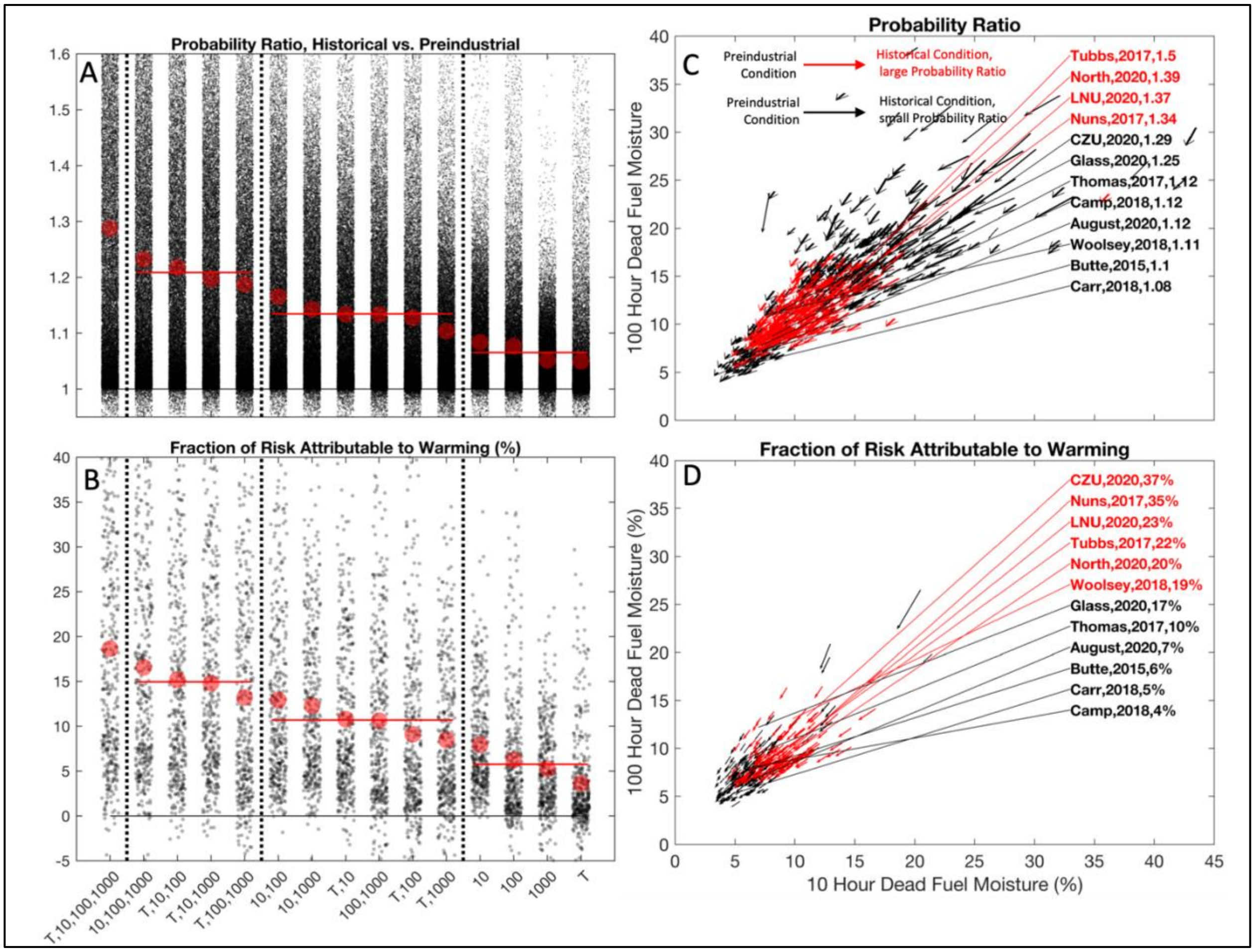

Fig. S26. Same as Fig. 3 but with the daily mean vapor pressure deficit predictor replaced with daily mean 10 hour dead fuel moisture.

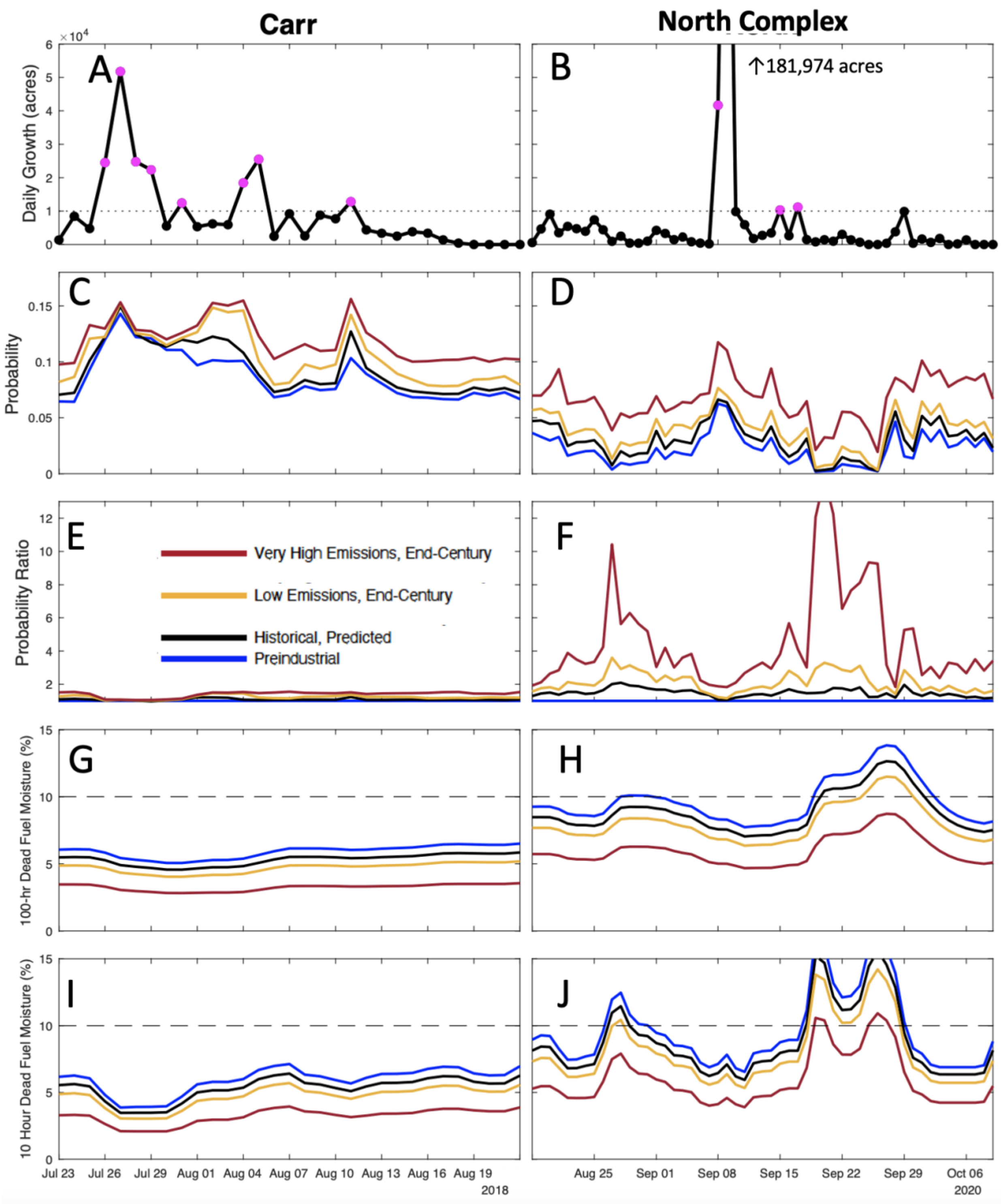

Fig. S27. Same as Fig. 4 but with the daily mean vapor pressure deficit predictor replaced with daily mean 10 hour dead fuel moisture.

S28. Summary of Results Under Sensitivity Tests

Table S4 summarizes the main results from Fig. 2G under the three sensitivity tests discussed above.

Table S4. Summary of main results for 3 sensitivity tests

		Expected Frequency of Extreme Daily Growth						Fractional Increase in Expected Frequency of Extreme Daily Growth	
	Mean FAR, Historical	Preindustrial	Historical, Predicted	Low Emissions, Mid-Century	Low Emissions, End-Century	Very High Emissions, Mid-Century	Very High Emissions, End-Century	Low Emissions, End-Century	Very High Emissions, End-Century
Main	19%	289	362	456	459	524	786	1.59	2.72
Swap CORDEX for CMIP6	22%	278	362	456	458	522	735	1.65	2.64
Swap 10hr DFM for VPD	19%	291	364	457	460	523	778	1.58	2.67
Swap 5,000 acres for 10,000 acres	17%	724	875	1060	1067	1191	1673	1.47	2.31

References

1 Moritz, M. A., Morais, M. E., Summerell, L. A., Carlson, J. M. & Doyle, J. Wildfires, complexity, and highly optimized tolerance. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17912-17917 (2005). https://doi.org:10.1073/pnas.0508985102

2 Wang, S. S. C., Qian, Y., Leung, L. R. & Zhang, Y. Identifying Key Drivers of Wildfires in the Contiguous US Using Machine Learning and Game Theory Interpretation. *Earth's Future* **9** (2021). https://doi.org:10.1029/2020ef001910

3 Huang, Y., Jin, Y., Schwartz, M. W. & Thorne, J. H. Intensified burn severity in California’s northern coastal mountains by drier climatic condition. *Environmental Research Letters* **15**, 104033 (2020). https://doi.org:10.1088/1748-9326/aba6af

4 Elia, M. *et al.* Estimating the probability of wildfire occurrence in Mediterranean landscapes using Artificial Neural Networks. *Environmental Impact Assessment Review* **85**, 106474 (2020). https://doi.org:10.1016/j.eiar.2020.106474

5 Satir, O., Berberoglu, S. & Donmez, C. Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem. *Geomatics, Natural Hazards and Risk* **7**, 1645-1658 (2016). https://doi.org:10.1080/19475705.2015.1084541

6 Philip, S. *et al.* A protocol for probabilistic extreme event attribution analyses. *Advances in Statistical Climatology, Meteorology and Oceanography* **6**, 177-203 (2020). https://doi.org:10.5194/ascmo-6-177-2020

7 Bindoff, N. L., P.A. Stott, K.M. AchutaRao, M.R. Allen, N. Gillett, D. Gutzler, K. Hansingo, G. Hegerl, Y. Hu, S. Jain, I.I. Mokhov, J. Overland, J. Perlwitz, R. Sebbari and X. Zhang. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.).

8 Pall, P. *et al.* Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature* **470**, 382-385 (2011). https://doi.org:10.1038/nature09762

9 Stone, D. A. & Allen, M. R. The End-to-End Attribution Problem: From Emissions to Impacts. *Climatic Change* **71**, 303-318 (2005). https://doi.org:10.1007/s10584-005-6778-2

10 Eyring, V. *et al.* Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **9**, 1937-1958 (2016). https://doi.org:10.5194/gmd-9-1937-2016

11 Patricola, C. M. & Wehner, M. F. Anthropogenic influences on major tropical cyclone events. *Nature* **563**, 339-346 (2018). https://doi.org:10.1038/s41586-018-0673-2

12 Seneviratne, S. I. *et al.* (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021).

13 Riahi, K. *et al.* RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic Change* **109**, 33-57 (2011). https://doi.org:10.1007/s10584-011-0149-y

14 Hausfather, Z. G. P. P. *Nature* **577**, 618-620 (2020). https://doi.org:10.1038/d41586-020-00177-3

15 Burgess, M. G., Ritchie, J., Shapland, J. & Pielke, R. IPCC baseline scenarios have over-projected CO2 emissions and economic growth. *Environmental Research Letters* **16**, 014016 (2021). https://doi.org:10.1088/1748-9326/abcdd2

16 Mandel, J. *et al.* Recent advances and applications of WRF–SFIRE. *Natural Hazards and Earth System Sciences* **14**, 2829-2845 (2014). https://doi.org:10.5194/nhess-14-2829-2014

17 Ficklin, D. L. & Novick, K. A. Historic and projected changes in vapor pressure deficit suggest a continental-scale drying of the United States atmosphere. *Journal of Geophysical Research: Atmospheres* **122**, 2061-2079 (2017). https://doi.org:10.1002/2016jd025855

18 Zhuang, Y., Fu, R., Santer, B. D., Dickinson, R. E. & Hall, A. Quantifying contributions of natural variability and anthropogenic forcings on increased fire weather risk over the western United States. *Proceedings of the National Academy of Sciences* **118**, e2111875118 (2021). https://doi.org:10.1073/pnas.2111875118

19 Jain, P., Castellanos-Acuna, D., Coogan, S. C. P., Abatzoglou, J. T. & Flannigan, M. D. Observed increases in extreme fire weather driven by atmospheric humidity and temperature. *Nature Climate Change* **12**, 63-70 (2022). https://doi.org:10.1038/s41558-021-01224-1

20 Dunn, R. J. H., Willett, K. M., Ciavarella, A. & Stott, P. A. Comparison of land surface humidity between observations and CMIP5 models. *Earth System Dynamics* **8**, 719-747 (2017). https://doi.org:10.5194/esd-8-719-2017

21 Benedetti, R. Scoring Rules for Forecast Verification. *Monthly Weather Review* **138**, 203-211 (2010). https://doi.org:10.1175/2009mwr2945.1

22 Brier, G. W. VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY. *Monthly Weather Review* **78**, 1-3 (1950). https://doi.org:10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2

23 Mandrekar, J. N. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology* **5**, 1315-1316 (2010). https://doi.org:10.1097/jto.0b013e3181ec173d

24 Rothermel, R. C. (U.S. Department of Agriculture, Forest Service, Intermountain Research Station, 1991).

25 Lareau, N. P. & Clements, C. B. Environmental controls on pyrocumulus and pyrocumulonimbus initiation and development. *Atmospheric Chemistry and Physics* **16**, 4005-4022 (2016). https://doi.org:10.5194/acp-16-4005-2016

26 Duane, A., Castellnou, M. & Brotons, L. Towards a comprehensive look at global drivers of novel extreme wildfire events. *Climatic Change* **165** (2021). https://doi.org:10.1007/s10584-021-03066-4

27 Mandel, J., Beezley, J. D. & Kochanski, A. K. Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. *Geoscientific Model Development* **4**, 591-610 (2011). https://doi.org:10.5194/gmd-4-591-2011

28 Khorshidi, M. S. *et al.* Increasing concurrence of wildfire drivers tripled megafire critical danger days in Southern California between1982 and 2018. *Environ. Res. Lett.* **15** (2020). https://doi.org:10.1088/1748-9326/abae9e

29 Thonicke, K. *et al.* The influence of vegetation, fire spread and fire behaviour on biomass burning and trace gas emissions: results from a process-based model. *Biogeosciences* **7**, 1991-2011 (2010). https://doi.org:10.5194/bg-7-1991-2010

30 Rabin, S. S. *et al.* The Fire Modeling Intercomparison Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions. *Geosci. Model Dev.* **10**, 1175-1197 (2017). https://doi.org:10.5194/gmd-10-1175-2017

31 Li, F. *et al.* Historical (1700–2012) global multi-model estimates of the fire emissions from the Fire Modeling Intercomparison Project (FireMIP). *Atmospheric Chemistry and Physics* **19**, 12545-12567 (2019). https://doi.org:10.5194/acp-19-12545-2019

32 Teckentrup, L. *et al.* Response of simulated burned area to historical changes in environmental and anthropogenic factors: a comparison of seven fire models. *Biogeosciences* **16**, 3883-3910 (2019). https://doi.org:10.5194/bg-16-3883-2019

33 Wang, S. S.-C., Qian, Y., Leung, L. R. & Zhang, Y. Interpreting machine learning prediction of fire emissions and comparison with FireMIP process-based models. *Atmospheric Chemistry and Physics* **22**, 3445-3468 (2022). https://doi.org:10.5194/acp-22-3445-2022

34 Coop, J. D. *et al.* Extreme fire spread events and area burned under recent and future climate in the western USA. *Global Ecology and Biogeography* (2022). https://doi.org:10.1111/geb.13496

35 Ranasinghe, R. *et al.* Climate Change Information for Regional Impact and for Risk Assessment. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. https://doi.org:10.1017/9781009157896.014

36 Williams, A. P. *et al.* Observed Impacts of Anthropogenic Climate Change on Wildfire in California. *Earth's Future* **7**, 892-910 (2019). https://doi.org:10.1029/2019ef001210

37 Abatzoglou, J. T. & Williams, A. P. Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences* **113**, 11770-11775 (2016). https://doi.org:10.1073/pnas.1607171113

38 Jolly, W. M. *et al.* Climate-induced variations in global wildfire danger from 1979 to 2013. *Nature Communications* **6**, 7537 (2015). https://doi.org:10.1038/ncomms8537

39 Westerling, A. L. (California's Fourth Climate Change Assessment, California Energy Commission., 2018).

40 Quilcaille, Y., Batibeniz, F., Ribeiro, A. F. S., Padrón, R. S. & Seneviratne, S. I. *Fire weather index data under historical and SSP projections in CMIP6 from 1850 to 2100* (Copernicus GmbH, 2022).

41 Abatzoglou, J. T., Williams, A. P. & Barbero, R. Global Emergence of Anthropogenic Climate Change in Fire Weather Indices. *Geophysical Research Letters* **46**, 326-336 (2019). https://doi.org:10.1029/2018gl080959

42 Michael, G. *et al.* Climate change is increasing the risk of extreme autumn wildfire conditions across California. *Environmental Research Letters* (2020).

43 Abatzoglou, J. T. Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology* **33**, 121-131 (2013). https://doi.org:10.1002/joc.3413

44 Brown, P. T. Reporting on global warming: A study in headlines. *Physics Today* (2016). https://doi.org:10.1063/PT.3.3310

45 Seager, R. *et al.* Climatology, Variability, and Trends in the U.S. Vapor Pressure Deficit, an Important Fire-Related Meteorological Quantity. *Journal of Applied Meteorology and Climatology* **54**, 1121-1141 (2015). https://doi.org:10.1175/jamc-d-14-0321.1