Peer Review File

Manuscript Title: Climate Warming Increases Risk of Extreme Daily Fire Growth in California

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

This manuscript quantifies the relationship between temperature and the risk of extreme daily wildfire growth in California and then uses these relationships to estimate future extreme daily wildfire growth. The topic is of interest and the authors have done a great deal of work but I can't recommend acceptance at this time.

The biggest problem I have with this work is the definition of extreme daily wildfire growth of 10,000 acres. I believe a percentile growth threshold would be a better approach. One can still require wildfires to be above a fire size threshold so that a 1 acre fire growing to 10 acres would be excluded. The problem with 10,000 is that in California you can have a fire of 1 million acres and a 10,000 acre increase is 1% (I realize this is at the extreme end but it is illustrative of the problem). At the other end you could have a 5,000 acre fire grow by 10,000 acres and it has tripled in size. Additionally, under the current system a 5,000 acre fire could increase by 9,000 acres that not qualify as an extreme fire event.

The second aspect that is a concern is the use of wildfire growth as the key variable. As the authors acknowledge there are numerous factors that play a confounding role in wildfire growth that are not directly accounted for in this study (L37-51). Vegetation type (fuel), ignitions (lightning and people), fire management activities (direct and indirect suppression, prescribed fire, policies such as fire bans and forest closures) and fire load.

Did you consider using other fuel moisture variables such as 1 hour and 10 hour fuel that can be important is some fuel types (grass and shrub etc.)

Referee #2 (Remarks to the Author):

This study assessed the impact of warming on probability of extreme daily wildfire growth and further quantified the attribution to the extreme wildfire behavior at the individual fire events. It took advantage of a large data set on satellite derived daily fire progression data in California during 2003 and 2020. Machine learning models were developed to build the non-linear relationship between occurrence of extreme fire behavior and weather and other predictors. Overall it is an interesting study addressing a critical question on the extreme fires.

My main concern is on the robustness of the empirical models due to the extremely unbalanced samples for the binary response variables (extreme vs non-extreme fire days: 380 vs. 18,000) and the very small size for the occurrence samples, especially considering the diverse landscape in California in terms of fuel types and topography.

It would be interesting to see what the response function to temperature looks like, based on the final models.

The statement about the critical thresholds of the aridity needs more solid evidence.

A bilinear spatial interpolation from 1 deg by 1 deg future climate projection data to fire-day locations was used for the projection of extreme fire risks while holding everything else constant. This interpolation may cause large uncertainty due to the complex topography, while there are quite a few higher resolution climate projection data available for California.

Methods for attribution needs to be clarified a bit more.

Referee #3 (Remarks to the Author):

A. Summary of the key results

The authors apply a machine-learning based algorithm trained on observations to project risks of extreme fire growth given an active fire under different climate conditions in California. The approach is of "storyline" type, i.e. using observed conditions as reference and adding on top of the observed climate a climate-change signal based on projected changes in temperature. The results are interesting and of relevance to the general public, as they suggest that if historical fires that occurred between 2003 and 2020 in California were submitted to changed climate conditions expected under a ca. 2°C (SSP1-2.6) or 4-5°C (SSP5-8.5) of global warming, these would experience a substantially increased expected frequency of extreme daily growth events (of 59% vs 172%, respectively).

Nonetheless, I have several concerns with the methodology applied in the study, and can thus not recommend this study for publication in Nature at the moment. It is however possible that a very substantially revised article could address these concerns and I thus would encourage the authors to consider resubmitting after very substantial revisions if these are possible. The main concerns are related to the following aspects:

1) The climate change scenario only includes temperature as input for the modified climate. However, changes in atmospheric humidity would also be highly relevant for predicting changes in VPD or fuel moisture.

2) It is not clear why the authors have decided to only use a machine-learning based algorithm, while physically-based models for fire modelling do exist (for instance the SpitFire model developed at PIK: Thonicke et al. 2010, Biogeosciences; https://www.pik-potsdam.de/en/institute/departments/earth-system-analysis/models/archiv-models-rd1/spitfire). I would have expected that the authors would have provided a comparison of the projections they derived with those derived with such a physical model. At the minimum, some comparison with alternative approaches, e.g. using fire indices (Abatzoglou et al. 2019, GRL), would be helpful to put their results in perspective.

3) This is a more minor point, but the scenario that the authors refer to as "in line with the Paris Agreement" is not actually within the limits set in the Paris agreement. The SSP1-2.6 scenario has approximately a global warming of 1.8° by the end of the century, with a very likely range of 1.3-2.4°C (IPCC 2021; https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM.pdf). This does not fit the limit of 1.5°C that world countries have set as aim to make efforts to limit global warming to within the Paris Agreement (and reiterated in 2021 in Glasgow), and is also arguably not compatible with a definition of "well below 2°C". In the literature, the scenario SSP1-1.9 is generally used as 1.5°C scenario in line with the Paris Agreement.

B. Novel and significance

The study is potentially novel in that it provides specific estimates of changes in fire risk in a highly exposed region. Nonetheless, the approach is possibly not robust enough as it only relies on a machine-learning algorithm and the study lacks comparisons with other widely applied approaches to assess fire risks in climate projections, in particular physically-based models and well-established fire indices.

C. Data and methodology

As mentioned, I have some concerns regarding the robustness of the applied methodology (see A).

D. Appropriate use of statistics and treatment of uncertainties

As mentioned, the uncertainty associated to the applied modelling approach is not sufficiently quantified and discussed (see A.).

E. Conclusions: robustness, validity, reliability

It is difficult to assess the robustness, validity and reliability of the results given the applied methodology (see above).

F. Suggested improvements
See under A (points 1-3).

G. References:
The references appear overall appropriate, but I miss a discussion of physically-based fire models (e.g. Thonicke et al. 2010; see A). Note that the included reference to the extremes chapter (Chapter 11) of IPCC AR6 WG1 (Seneviratne et al. 2021; https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter11.pdf) only seems to be a partial reference, with no page numbers or doi indication. In addition, also chapter 12 of the same report (Ranasinghe et al. 2021, https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter12.pdf) includes relevant material on fire projections and would need to be cited.

H. Clarity and context

Overall suitable but not enough context on applied methodology and on relevance of chosen projection scenario with respect to Paris Agreement.

Mentioned references:

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

IPCC 2021: Summary for Policymakers. 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 [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 3−32, doi:10.1017/9781009157896.001.

Ranasinghe, R., A.C. Ruane, R. Vautard, N. Arnell, E. Coppola, F.A. Cruz, S. Dessai, A.S. Islam, M. Rahimi, D. Ruiz Carrascal, J. Sillmann, M.B. Sylla, C. Tebaldi, W. Wang, and R. Zaaboul, 2021:

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 [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1767–1926, doi:10.1017/9781009157896.014.

Seneviratne, S.I., X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S.M. Vicente-Serrano, M. Wehner, and B. Zhou, 2021: Weather and Climate Extreme Events in a Changing Climate. 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 [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1513–1766, doi:10.1017/9781009157896.013.

Thonicke, K., Spessa, A., Prentice, I. C., Harrison, S. P., Dong, L., and Carmona-Moreno, C. 2010: 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, doi:10.5194/bg-7-1991-2010, 2010

Author Rebuttals to Initial Comments:

Reviewer 1

Comments to the author (if any):

This manuscript quantifies the relationship between temperature and the risk of extreme daily wildfire growth in California and then uses these relationships to estimate future extreme daily wildfire growth. The topic is of interest and the authors have done a great deal of work but I can't recommend acceptance at this time.

The biggest problem I have with this work is the definition of extreme daily wildfire growth of 10,000 acres. I believe a percentile growth threshold would be a better approach. One can still require wildfires to be above a fire size threshold so that a 1 acre fire growing to 10 acres would be excluded. The problem with 10,000 is that in California you can have a fire of 1 million acres and a 10,000 acre increase is 1% (I realize this is at the extreme end but it is illustrative of the problem). At the other end you could have a 5,000 acre fire grow by 10,000 acres and it has tripled in size. Additionally, under the current system a 5,000 acre fire could increase by 9,000 acres that not qualify as an extreme fire event.

Response: We focus on absolute growth of >10,000 acres per day (~98th percentile of fire-days in our dataset) because absolute growth is more directly related to firefighting challenges and threats to life and property than relative growth.

Nevertheless, when we began this research, we shared the reviewer's intuition that a statistical model might be better at predicting relative growth than it would be at predicting absolute growth for exactly the reasons the reviewer articulates. We initially tested the ability of models to predict various fractional growth values but found very little skill.

Upon reflection, we believe that this is because the cumulative area burned at any given time (i.e., the "size" of the fire) does not actually provide information on the size of the *active* fire front. For example, a fire that has cumulatively burned 1,000 acres may have the same size *active* fire front as a fire that has cumulatively burned 100,000 acres.

Below we demonstrate the lack of predictive skill for relative growth. Specifically, we test the sensitivity of our results to replacing 10,000 acres of daily growth with a quadrupling of fire size from the previous

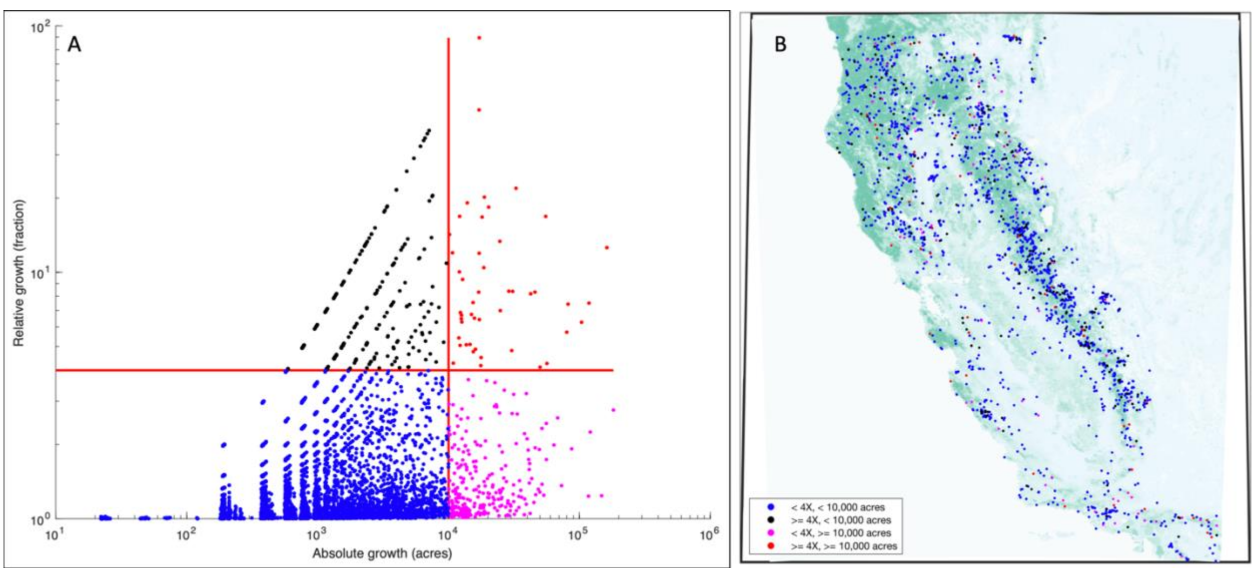
day. A quadrupling was chosen because it represented a similar level of extremity to the 10,000 acre threshold. Specifically, once our fire dataset was limited to fires with multi-day growth (necessary in order to calculate relative growth), 3% of the fire-days demonstrated a quadrupling in size from the previous day (recall that 2% of our fire-days displayed 10,000 acres growth).

We find that the statistical models are much less skillful in their predictions of quadruplings than in their predictions of 10,000 acres of growth.

For the top 10% of model configurations, the log-loss skill score decreased from 0.22 for predicting absolute growth to 0.02 for relative growth (a decrease of 91%), the Brier skill score decreased from 0.064 to 0.002 (a decrease of 97%), the reliability diagram score decreased from 0.88 to 0.48 (a decrease of 45%), and the ROC-AUC decreased from 0.89 to 0.62 (a decrease of 30%).

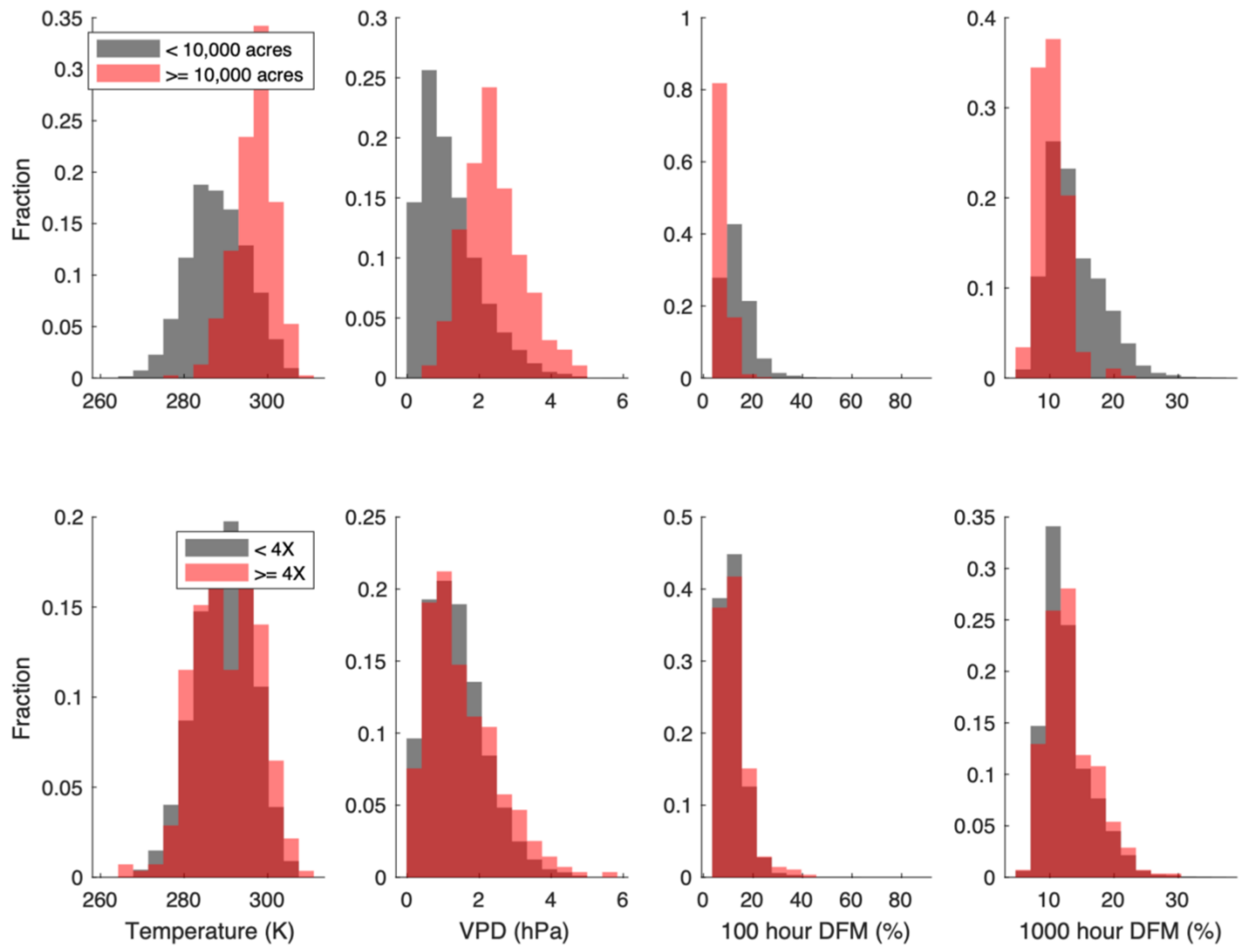
Understanding exactly why quadrupling days are so much more difficult to predict (from the 11 predictors used) than 10,000-acre growth days is beyond the scope of this study, but Response Fig. 1 and Response Fig. 2 provide some initial clues.

Response Fig. 1 indicates that there is not a tremendous amount of overlap between 10,000-acre growth days and quadrupling days. There are many fire-days in which fires quadruple in size but do not grow 10,000 acres (black dots in Response Fig. 1) and many fire days in which fires grow 10,000 acres but do not quadruple in size (pink dots in Response Fig. 1) The r^2 value between absolute growth and relative growth is only 0.03.



Response Fig. 1 | All fire-days in the dataset with at least 1 antecedent fire-day plotted in absolute and relative growth space (left) and in latitude-longitude space (right). Fire-days that show at least quadrupling in size but not more than 10,000 acres in growth are shown in black, and fire-days that show at least 10,000 acres in growth but not at least a quadrupling are shown in pink.

Response Fig. 2 illustrates one of the proximate reasons why quadruplings are so much more difficult to predict than 10,000 acres of growth. For 10,000 acres of growth, the distributions of the aridity-related predictors show significant differentiation (lack of overlap), indicating that these predictors will be useful for predicting large absolute growth (top row of Response Fig. 2). On the other hand, for quadruplings, the distributions of the predictor values show a lack of differentiation (substantial overlap), indicating that these predictors will not be useful in the prediction of relative growth (bottom row of Response Fig. 2).



Response Fig. 2 | Distributions for the four temperature/fuel aridity predictors used in this study, divided into small growth (grey) and large growth (red). The top row is for the 10,000 acres absolute growth threshold, and the bottom row is for the quadrupling relative growth threshold. The overlap between the two distributions in the bottom row shows that these predictors provide little information for predicting relative growth, but the differentiation between the distributions in the top row shows that these predictors can help discriminate between large and small absolute growth.

Overall, the lack of skill in predicting relative growth indicates that we cannot use our predictors and models to perform attribution tests on relative growth, and we cannot use them to project changes in relative growth into the future. Perhaps future work could investigate this issue more thoroughly.

The second aspect that is a concern is the use of wildfire growth as the key variable. As the authors acknowledge there are numerous factors that play a confounding role in wildfire growth that are not directly accounted for in this study (L37-51). Vegetation type (fuel), ignitions (lightning and people), fire management activities (direct and indirect suppression, prescribed fire, policies such as fire bans and forest closures) and fire load.

Response: We chose to focus on wildfire growth for the practical reason that it is relevant to consequences: large growth rates are associated with challenges in fighting fires, and thus these large growth days pose the biggest threat to life and property (articulated on lines 68-71 in the revised manuscript). Also, the most meaningful wildfire characteristics that we could substitute for growth would suffer from the same confounding factors.

Nevertheless, our study demonstrates that large growth days of 10,000 acres or more *are predictable* using our predictors *despite* our model having no information on the mentioned confounding factors. 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 at the upper end of what is typically considered "excellent" (0.8-0.9) and close to "outstanding" (>0.9).

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.

In the revised manuscript, we have added a discussion of this in Section S17.

Did you consider using other fuel moisture variables such as 1 hour and 10 hour fuel that can be important is some fuel types (grass and shrub etc.)

Response: Our analysis was conducted at the daily timescale, and thus 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 been highlighted as a particularly useful predictor of fire proclivity in other studies. However, it is reasonable to ask whether our results are sensitive to using a measure other than vapor pressure deficit to represent the aridity of finer fuels.

Thus, in the revised manuscript, we have included a sensitivity test where the daily mean vapor pressure deficit predictor is replaced with daily mean 10-hour fuel moisture (Section S27). We find that this substitution has a minimal impact on our calculations (c.f. Figs 2-4 to Fig. S25-S27). For example, here is how the numbers in the abstract would be updated under these conditions:

> "…the fraction of the risk of extreme daily growth attributable to anthropogenic warming to date averages ~~19%~~ 19% [did not change] but varies substantially depending on whether background warming pushed fires over critical aridity thresholds. When the historical fires from 2003 to 2020 are subjected to projected end-of-century temperatures, the expected frequency of extreme daily growth events increases by ~~59%~~ 58% under a low emissions scenario compared to an increase of ~~172%~~ 167% under a very high emissions scenario."

See Table S4 in the revised manuscript for a summary of the main results of the study under three sensitivity tests.

Reviewer 2

This study assessed the impact of warming on probability of extreme daily wildfire growth and further quantified the attribution to the extreme wildfire behavior at the individual fire events. It took advantage of a large data set on satellite derived daily fire progression data in California during 2003 and 2020. Machine learning models were developed to build the non-linear relationship between occurrence of extreme fire behavior and weather and other predictors. Overall it is an interesting study addressing a critical question on the extreme fires.

My main concern is on the robustness of the empirical models due to the extremely unbalanced samples for the binary response variables (extreme vs non-extreme fire days: 380 vs. 18,000) and the very small size for the occurrence samples, especially considering the diverse landscape in California in terms of fuel types and topography.

Response: We acknowledge the reviewer's concern, which is an inherent tradeoff in the study of extreme events—that the more extreme the event, the fewer observations you have of it. We choose 10,000 acres of daily growth because we made the assessment that this threshold balanced our goals of studying extreme events and having enough events to make robust statistical inferences.

All of our scoring metrics are calculated on out-of-training-sample data, and good skill is shown. Specifically, the models we use to assess warming's influence on the risk of extreme growth (the top 10% of model configurations on out-of-training-sample predictions) have a Receiver Operator Characteristic Area Under the Curve (ROC-AUC) of 0.89 (Fig. S3), which is at the upper end of what is typically considered "excellent" (0.8-0.9) and close to "outstanding" (>0.9).

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.

Nevertheless, in the revised manuscript, we have added additional analysis using a less extreme threshold and, thus, a less unbalanced sample. Specifically, we recalculate our results using daily growth of 5,000 acres or more (rather than 10,000 acres or more). This less extreme class of events increases our number of positive cases 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 quite similar between the two sets of analyses. For the top 10% of model configurations, the log-loss skill score decreased from 0.22 for predicting >10,000 acres of growth to 0.21 for predicting >5,000 acres of 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. The fact that the skill scores are similar between these two sets of analyses indicates that the unbalanced sample of 380 to 17,910 is not particularly problematic.

We also find that all results are qualitatively similar (c.f. Figs. S18 – Fig. S20 to Figs 2-4) but that anthropogenic warming causes a larger enhancement of the probability of the more extreme class of events. For example, here is how the numbers in the abstract would be updated using the less extreme >5,000 acres growth threshold:

> "…the fraction of the risk of extreme daily growth attributable to anthropogenic warming to date averages ~~19%~~ 17% but varies substantially depending on whether background warming pushed fires over critical aridity thresholds. When the historical fires from 2003 to 2020 are subjected to projected end-of-century temperatures, the expected frequency of extreme daily growth events increases by ~~59%~~ 47% under a low emissions scenario compared to an increase of ~~172%~~ 131% under a very high emissions scenario."

This indicates that anthropogenic warming increases the probability of extreme growth nonlinearly, where the more extreme the growth, the more its probability is enhanced. This is the same characteristic that is observed for other phenomena like extreme rainfall (e.g., Fig 11.15 in IPCC AR6).

This is discussed in the revised manuscript in Section S25.

See Table S4 in the revised manuscript for a summary of the main results of the study under three sensitivity tests.

It would be interesting to see what the response function to temperature looks like, based on the final models.

The statement about the critical thresholds of the aridity needs more solid evidence.

Response: We thank the reviewer for the suggestion to include response functions, as this adds interesting analysis and it provides additional evidence for nonlinearity and thresholds in the data.

In the revised manuscript, Section S21 discusses response functions for temperature and the three aridity variables (shown in Figure S13). 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 confirms 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. This is consistent with the critical thresholds inferred from Fig. 2C, Fig. 2D, and Fig. 4 though they are not directly comparable because the values of the other predictors in Fig. 2 are not necessarily at their mean value (as they are for the calculation of response functions). These values are very close to the critical thresholds identified in a more formal manner (using different data and methodologies) for large growth in Southern California (Khorshidi et al., 2020). That study identified 1.5 kPa of vapor pressure deficit, 9.4% for 100 hour dead fuel moisture, 12.5% for 1000 hour dead fuel moisture as thresholds.

A bilinear spatial interpolation from 1 deg by 1 deg future climate projection data to fire-day locations was used for the projection of extreme fire risks while holding everything else constant. This interpolation may cause large uncertainty due to the complex topography, while there are quite a few higher resolution climate projection data available for California.

Response: We thank the reviewer for pointing this out.

In the revised manuscript, we test the sensitivity of our results to the use of model output from the Coordinated Regional Downscaling Experiment (CORDEX, https://cordex.org/). CORDEX uses CMIP6 global climate models to drive regional climate models. CORDEX's spatial resolution for North America is 12.5 by 12.5 km, which is significantly higher than the CMIP6 mean (Table S1).

In the revised manuscript, section S26 shows the sensitivity of our result to the use of CORDEX instead of CMIP6. Fig S21 shows the background climatological temperature change as calculated by CORDEX and Figs. S22-S24 show our main figures 2-4 recalculated using CORDEX. Results are quantitatively very similar. For example, the main findings cited in the abstract would be

> "…the fraction of the risk of extreme daily growth attributable to anthropogenic warming to date averages ~~19%~~ 22% but varies substantially depending on whether background warming pushed fires over critical aridity thresholds. When the historical fires from 2003 to 2020 are subjected to projected end-of-century temperatures, the expected frequency of extreme daily growth events increases by ~~59%~~ 65% under a low emissions scenario compared to an increase of ~~172%~~ 164% under a very high emissions scenario."

Comparing Fig. 2G to S22G probably gives the cleanest comparison between the two.

See Table S4 in the revised manuscript for a summary of the main results of the study under three sensitivity tests.

We believe it is valuable to have this sensitivity test as part of the supplement, but we have chosen to keep the CMIP6-based results in the main text. The reason for this is that using CORDEX presents a practical issue that complicates our procedure: CORDEX does not have a preindustrial experiment (its base experiment runs from 1950-2014). Thus, we needed to use raw CMIP6 output to calculate preindustrial temperature departures from the current climate. While we feel that this is a valid way to address the issue, it does add an additional step and complicates our methodology. Since using CMIP6 output is more standard than using CORDEX, since using CORDEX adds additional methodological complications, and since the resulting difference between CMIP6 and CORDEX is small, we prefer to feature the CMIP6 results in the main text and discuss the CORDEX results as a sensitivity test in the supplement.

Methods for attribution needs to be clarified a bit more.

Response: In the revised manuscript, we have added additional clarifications of the methods for attribution on lines 770-775.

Reviewer 3

A. Summary of the key results

The authors apply a machine-learning based algorithm trained on observations to project risks of extreme fire growth given an active fire under different climate conditions in California. The approach is

of "storyline" type, i.e. using observed conditions as reference and adding on top of the observed climate a climate-change signal based on projected changes in temperature. The results are interesting and of relevance to the general public, as they suggest that if historical fires that occurred between 2003 and 2020 in California were submitted to changed climate conditions expected under a ca. 2°C (SSP1-2.6) or 4-5°C (SSP5-8.5) of global warming, these would experience a substantially increased expected frequency of extreme daily growth events (of 59% vs 172%, respectively).

Nonetheless, I have several concerns with the methodology applied in the study, and can thus not recommend this study for publication in Nature at the moment. It is however possible that a very substantially revised article could address these concerns and I thus would encourage the authors to consider resubmitting after very substantial revisions if these are possible. The main concerns are related to the following aspects:

1) The climate change scenario only includes temperature as input for the modified climate. However, changes in atmospheric humidity would also be highly relevant for predicting changes in VPD or fuel moisture.

Response:

We agree that climatic variables other than temperature are important for projecting changes in wildfire risk. In addition to absolute atmospheric humidity, other important variables include changes in precipitation, wind patterns, vegetation, snowpack, ignitions, antecedent fire activity, etc. Not to mention factors like changes in human population distribution, fuel breaks, land use, ignition patterns, firefighting tactics, forest management strategies, and long-term buildup of fuels.

Accounting for changes in all of these variables and their potential interactions simultaneously is very difficult. This is precisely why we chose to use a methodology that addresses the much cleaner but more narrow question of what the influence of *warming alone* is on the risk of extreme daily wildfire growth.

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. There is no consensus on even the expected direction of the change of many of the other relevant variables.

We do not believe that absolute humidity is an exception to this, and thus it is just as justifiable to hold it constant as any of the other variables mentioned above. Support for this can be summarized with the following two points.

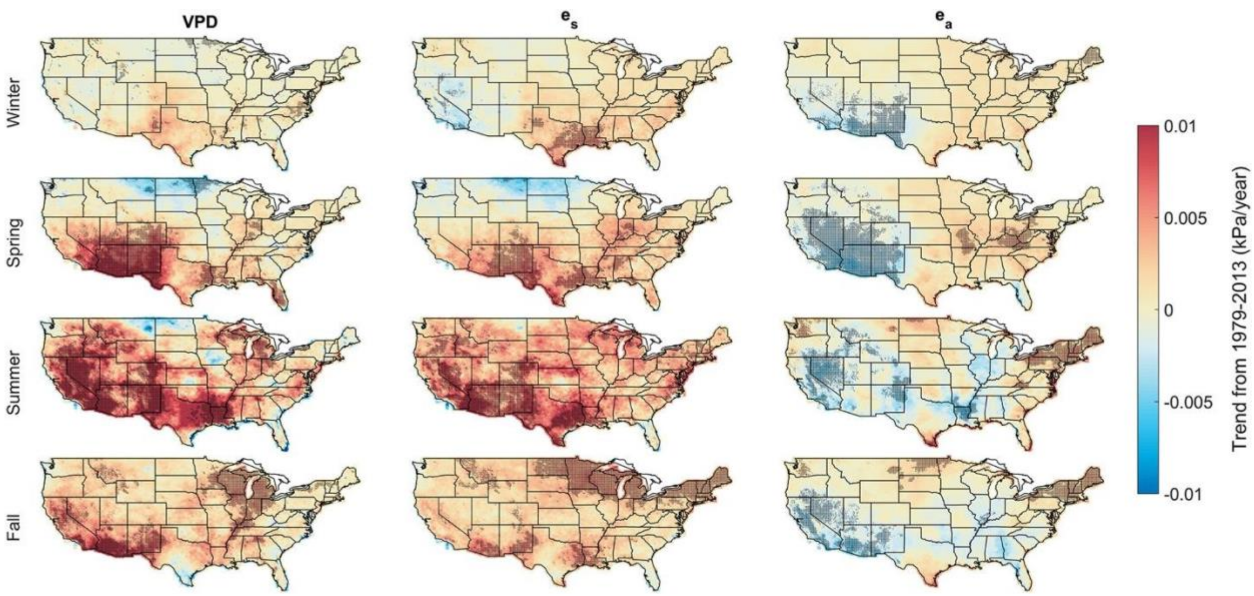
1) Previous work has demonstrated that long-term changes in relative humidity and vapor pressure deficit are not heavily influenced by changes in absolute humidity but rather are dominated by changes in saturation vapor pressure (or saturation specific humidity, etc.) which in turn is due to temperature change. For example, Ficklin and Novick (2017) find that:

> "For the summer, the general circulation model ensemble median showed a 51% projected increase (quartile range of 39 and 64%) in summer vapor pressure deficit for the U.S., reflecting temperature-driven increases in saturation vapor pressure but decreases or minimal changes in relative humidity that promotes negligible changes in vapor pressure".

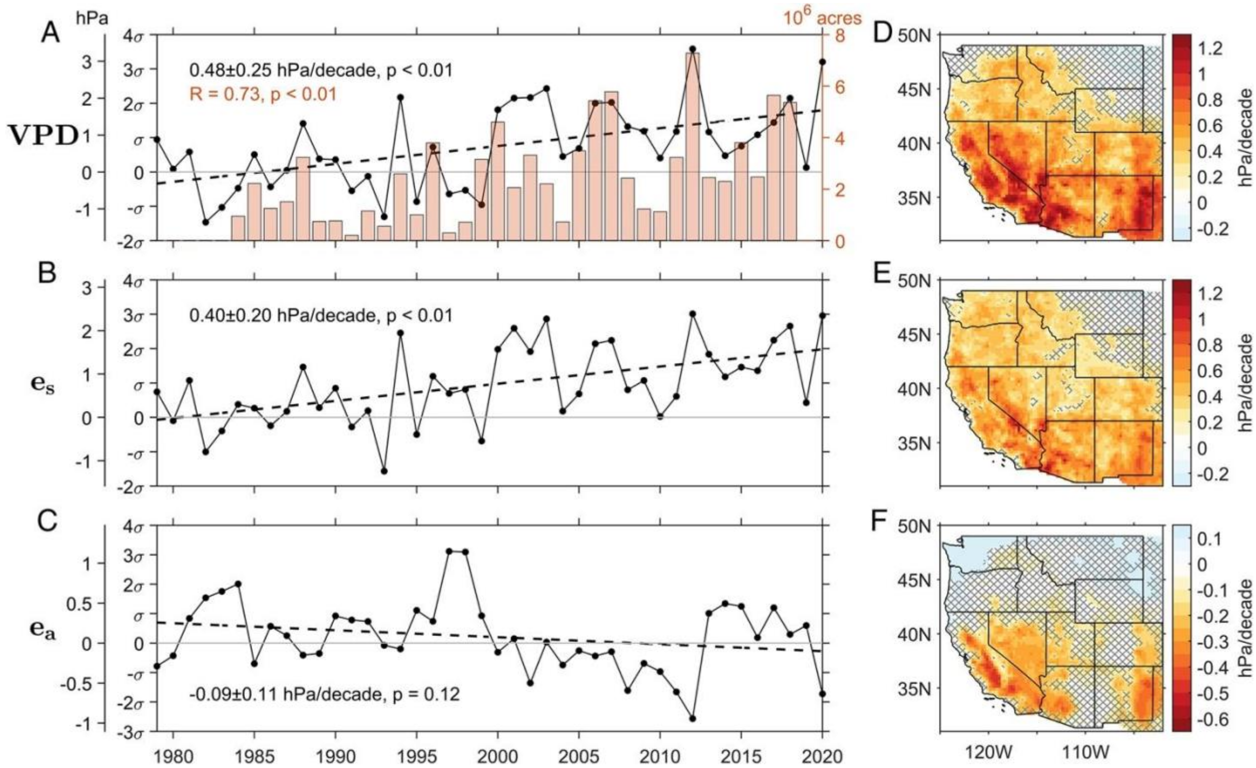
2) We do not have high confidence in even the direction of future changes in absolute humidity over our region. This is because global climate models struggle to reproduce both the observed general relationship between absolute humidity and temperature as well as the long-term trend in absolute humidity observed over the US West.

More specifically, a global analysis revealed that the relationship between temperature and vapor pressure in global climate models is more steeply positive than in observations in all regions, and this over-correlation may be due to missing processes in the models (Dunn et al. 2017). This means that global climate models will have a tendency to project increases in absolute humidity over our region but that we should not have high confidence in these projections.

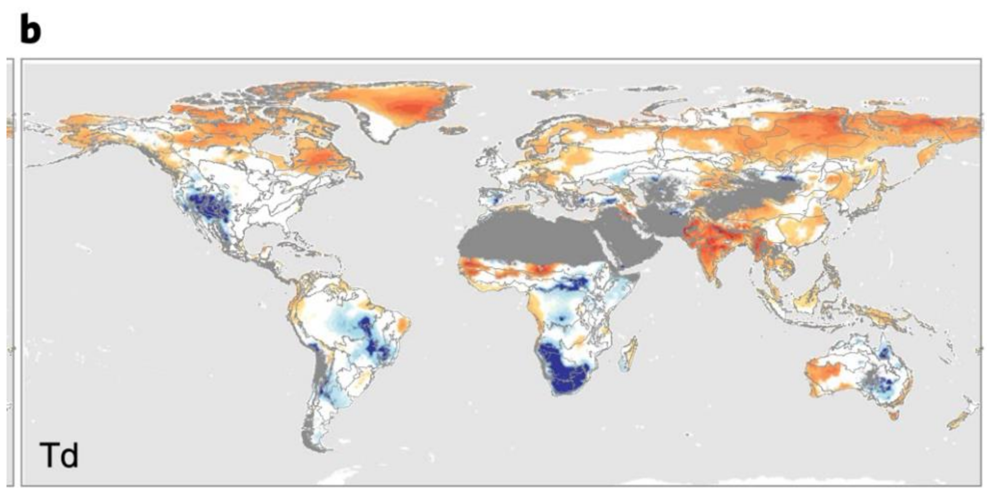
Indeed, historical observations show a *decrease* in absolute humidity (vapor pressure) over our region since 1979 (Response Fig. 3, Response Response Fig. 4, and Response Fig. 5, all from different publications). This could be due to forced changes in circulation or land-atmosphere feedbacks not properly captured by the models. Or it could be due to an unforced response temporarily overwhelming the forced response. We do not know.



Response Fig. 3 | Observed trends by season from 1979 to 2013 in vapor pressure deficit (left column), saturation vapor pressure (middle column), and vapor pressure (right column). From Ficklin and Novick (2017)



Response Fig. 4 | Observed trends in vapor pressure deficit (top row), saturation vapor pressure (middle row), and vapor pressure (bottom row). From Zhuang et al. (2021).



Response Fig. 5 | Observed trends in absolute humidity (here represented by dew point) from 1979-2020 over fire season. Red is positive, and blue is negative. From Jain et al. (2022)

Overall, we have confidence that changes in relative humidity and vapor pressure deficit are dominated by changes in temperature (and thus saturation vapor pressure), but it is unclear how we should expect absolute humidity (vapor pressure) to change in the long term over our region. It follows then that absolute humidity is similar to the other variables that we hold constant, and there is no reason to give it special treatment by allowing it to change. In other words, if we allowed absolute humidity to vary, it

would not be clear which direction we should trend it (positive change or negative change), and it would not be clear why we chose to vary absolute humidity but not other variables like precipitation.

Having said that, the reviewer is absolutely correct to point out that holding absolute humidity, as well as every other variable constant is *the* major caveat in our study. It is critical that the reader understand this caveat, so we have adjusted the title to reflect this. We changed the title from:

Climate-Driven Risk of Extreme Wildfire in California

to

Climate warming increases risk of extreme wildfire behavior in California

We have also taken care to emphasize this caveat prominently throughout the revised manuscript (e.g., lines 76-141, lines 217-220, lines 340-352 [the concluding paragraph], and lines 782-834).

2) It is not clear why the authors have decided to only use a machine-learning based algorithm, while physically-based models for fire modelling do exist (for instance the SpitFire model developed at PIK: Thonicke et al. 2010, Biogeosciences;

https://www.pik-potsdam.de/en/institute/departments/earth-system-analysis/models/archiv-models-rd1/spitfire). I would have expected that the authors would have provided a comparison of the projections they derived with those derived with such a physical model. At the minimum, some comparison with alternative approaches, e.g. using fire indices (Abatzoglou et al. 2019, GRL), would be helpful to put their results in perspective.

Response: We thank the reviewer for the encouragement to explicitly justify the use of our methods in the context of existing process-based models and fire weather indices. This has helped us clarify the novel contribution of this study.

In the revised manuscript, the 2nd through 6th paragraphs communicate that our study is designed specifically to answer a research question that is difficult to address with existing process-based models or with fire weather index analysis.

Despite this, the reviewer is correct to point out that it is valuable to compare our results to the output from process-based models and to fire weather index analyses to make sure that there are no obvious inconsistencies. Thus, in the revised manuscript, we have added a section comparing our results to each of these (Section S22 for process-based models and Section S23 for fire weather indices).

In Section S22, we show that prominent process-based models have insufficient resolution and have too much uncertainty to address our research question. However, we also show that there is no reason to believe that the output from these models is *inconsistent* with our findings.

In Section S23, we find that the percent change in extreme fire weather days is similar to the percent change in extreme growth days calculated by our method under the same emissions scenario.

Overall, comparing our results to calculations from process-based models and to changes in fire weather indices emphasizes that our study fills a research gap (because neither of these methods can answer our

research question easily) while also indicating that our results are consistent with those produced from these alternative approaches.

3) This is a more minor point, but the scenario that the authors refer to as "in line with the Paris Agreement" is not actually within the limits set in the Paris agreement. The SSP1-2.6 scenario has approximately a global warming of 1.8° by the end of the century, with a very likely range of 1.3-2.4°C (IPCC 2021; https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM.pdf). This does not fit the limit of 1.5°C that world countries have set as aim to make efforts to limit global warming to within the Paris Agreement (and reiterated in 2021 in Glasgow), and is also arguably not compatible with a definition of "well below 2°C". In the literature, the scenario SSP1-1.9 is generally used as 1.5°C scenario in line with the Paris Agreement.

Response: We thank the reviewer for pointing this out. In the revised manuscript, we no longer refer to the SSP1-2.6 scenario as being "in line with the Paris Agreement"

B. Novel and significance

The study is potentially novel in that it provides specific estimates of changes in fire risk in a highly exposed region. Nonetheless, the approach is possibly not robust enough as it only relies on a machine-learning algorithm and the study lacks comparisons with other widely applied approaches to assess fire risks in climate projections, in particular physically-based models and well-established fire indices.

Response: We appreciate the reviewer's push for us to better place our findings in the context of physically-based models and fire weather indices, which we have done in the new sections S22 and S23.

C. Data and methodology

As mentioned, I have some concerns regarding the robustness of the applied methodology (see A).

Response: See our response to A above.

D. Appropriate use of statistics and treatment of uncertainties

As mentioned, the uncertainty associated to the applied modelling approach is not sufficiently quantified and discussed (see A.).

E. Conclusions: robustness, validity, reliability

It is difficult to assess the robustness, validity and reliability of the results given the applied methodology (see above).

Response: See our response to A above.

Also, in the revised manuscript, we have added several tests of robustness requested by other reviewers (Section S28).

In terms of the explicit quantification of the uncertainty of our results, we demonstrate that our machine learning models produce reliable results (in the technical sense) by showing reliability diagrams

on out-of-training-sample test data. The reliability diagram (Fig. S3C) shows that on out-of-training-sample predictions, when the models predict, e.g., about a 10% chance of extreme growth, extreme growth does, in fact, in about 10% of cases. This is also seen in the 'train on cool, test on warm' experiment (Section S18 and Fig. S3D), which we think is a critical demonstration of the robustness of our results.

F. Suggested improvements
See under A (points 1-3).

Response: See above.

G. References:

The references appear overall appropriate, but I miss a discussion of physically-based fire models (e.g. Thonicke et al. 2010; see A).

Note that the included reference to the extremes chapter (Chapter 11) of IPCC AR6 WG1 (Seneviratne et al. 2021; https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter11.pdf) only seems to be a partial reference, with no page numbers or doi indication.

In addition, also chapter 12 of the same report (Ranasinghe et al. 2021, https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter12.pdf) includes relevant material on fire projections and would need to be cited.

Response: We thank the reviewer for pointing this out. In the revised manuscript, we have updated the IPCC Chapter 11 reference and added a reference to IPCC Chapter 12 as well as Thonicke et al. 2010.

H. Clarity and context

Overall suitable but not enough context on applied methodology and on relevance of chosen projection scenario with respect to Paris Agreement.

Response: Thank you, see above.

Mentioned references:

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

IPCC 2021: Summary for Policymakers. 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 [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 3−32, doi:10.1017/9781009157896.001.

Ranasinghe, R., A.C. Ruane, R. Vautard, N. Arnell, E. Coppola, F.A. Cruz, S. Dessai, A.S. Islam, M. Rahimi, D. Ruiz Carrascal, J. Sillmann, M.B. Sylla, C. Tebaldi, W. Wang, and R. Zaaboul, 2021: 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 [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1767–1926, doi:10.1017/9781009157896.014.

Seneviratne, S.I., X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S.M. Vicente-Serrano, M. Wehner, and B. Zhou, 2021: Weather and Climate Extreme Events in a Changing Climate. 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 [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1513–1766, doi:10.1017/9781009157896.013.

Thonicke, K., Spessa, A., Prentice, I. C., Harrison, S. P., Dong, L., and Carmona-Moreno, C. 2010: 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, doi:10.5194/bg-7-1991-2010, 2010

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

This manuscript quantifies the relationship between temperature and three associated variables influenced by temperature (vapor pressure deficit and two fuel moisture values) to the risk of extreme daily wildfire growth in California. These relationships are used to estimate future extreme daily wildfire growth. The topic is of interest and the authors have done an enormous amount of work but I can't recommend acceptance at this time.

The biggest problem I have with this work is the definition of extreme daily wildfire growth of 10,000 acres. In the manuscript it states L62-64

'Extreme wildfire behavior can be defined in several ways23, but here we define it as an exceptionally high rate of spread, and in particular, growth of 10,000 acres (about 2/3rds the size of Manhattan) or more in a single day.'

High rate of spread is not necessarily equal to fire growth of 10,000 or more acres. You can have a fire with a rapid rate of spread but the entire fire area is less than 10,000 acres. Conversely, you could have a large fire with a long active fire front with a rate of spread that is not all that rapid and still meet the 10,000 threshold. If days of rapid relative growth (a doubling or more) had a high overlap with the 10,000 acres I would then accept this variable. In the response to the reviews the authors state

Response Fig. 1 indicates that there is not a tremendous amount of overlap between 10,000-acre growth days and quadrupling days

I think this infers that this study is a reflection of growth characteristics of large fires in California not necessarily those fires with exceptionally high rate of spread. One could use a modelling approach to directly estimate the rate of spread from the satellite data – something akin to the Parks (2014) approach. The other option would be to use a doubling or tripling etc. of fire area. This option often happens in the early stages of an extreme fire but this can be when significant impacts occur (e.g., Camp Fire). Yes, I see that quadrupling fires are more difficult to predict but you could try tripling or doubling to see if you get better results. Also, the results for Fire Weather Indices were not great because I think 10,000 acres growth is not a best variable for rapid rates of spread (L1192-1199).

One could reframe the study to explore days where fire growth is 10,000 acres or more and how these large growth fire days could change in the future.

Parks, S.A., 2014. Mapping day-of-burning with coarse-resolution satellite fire-detection data. Int. J. Wildland Fire 23, 215–223.

Referee #2 (Remarks to the Author):

This study aims to quantify the influences of temperature on the extreme daily fire growth using machine learning approaches based on the observed daily burned area, and to make attribution assessment. Overall it is an interesting statistical study and the findings further support previous findings on extreme fire risk attribution to anthropogenic warming. The revision has addressed

most of the reviewers' comments and improved the quanlity of the manuscript. The trend analysis on extreme fire growth, extreme weather, and the association between then would make the conclusion more compelling.

A few other comments:

* Line 63 - "here we define it as an exceptionally high rate of spread, and in particular, growth of 10,000 acres". The data used for this study is daily area burned, which is different from fire rate of spread (velocity), although it is correlated for the most part.

* I would suggest including the number of fire-days samples (all and extreme) in the main text, as the sample size is critical for the machine learning approaches.

* Line 73 - It is worth mentioning that fuels including both fuel types, structure, and amount were held constant in addition to fuel moisture.

* For the extreme fire risk machine learning modeling with random forest, how the probability of extreme fire was calculated for the response variable?

Referee #3 (Remarks to the Author):

Recommendation: Accept subject to minor revisions.

The authors have satisfactorily addressed most of my comments, including by revising their title, and by providing more analyses in the supplementary information.

Regarding the fact that absolute humidity is kept constant in the experiments, I still consider this a major caveat (as also recognized by the authors in their rebuttal: "the reviewer is absolutely correct to point out that holding absolute humidity, as well as every other variable constant is the major caveat in our study"). I am still not fully convinced that this caveat is sufficiently addressed in the revised manuscript:
- The authors write "We have also taken care to emphasize this caveat prominently throughout the revised manuscript (e.g., lines 76-141, lines 217-220, lines 340-352 [the concluding paragraph], and lines 782-834". However, in the concluding paragraph (which is on lines 263-275, and not 340-352) the authors do not specifically refer to specific humidity. In particular, humidity is not mentioned in the following text (lines 271-275): "Despite this, we believe that our calculations probably result in conservative estimates of changes in risk because the balance of evidence suggests that it is more likely than not that important variables that we hold constant (precipitation36, wind11, tree mortality37, fire season length15,38,39, and lifetimes of fires2) are changing in a way that would exacerbate the enhanced risk caused by warming rather than ameliorate it.". However, the authors show in their rebuttal (Fig. 3 of the rebuttal) that absolute humidity has decreased in the region, which would also suggest that this factor would contribute to increasing fire risk. At the minimum, absolute humidity should be thus explicitly mentioned in the highlighted sentence. But I would also recommend that the authors add more text on the potential role of humidity in the long-term trends elsewhere in the manuscript.

Author Rebuttals to First Revision:

Reviewer 1

Comments to the author (if any):

This manuscript quantifies the relationship between temperature and three associated variables influenced by temperature (vapor pressure deficit and two fuel moisture values) to the risk of extreme daily wildfire growth in California. These relationships are used to estimate future extreme daily wildfire growth. The topic is of interest and the authors have done an enormous amount of work but I can't recommend acceptance at this time.

The biggest problem I have with this work is the definition of extreme daily wildfire growth of 10,000 acres. In the manuscript it states L62-64

> *'Extreme wildfire behavior can be defined in several ways23, but here we define it as an exceptionally high rate of spread, and in particular, growth of 10,000 acres (about 2/3rds the size of Manhattan) or more in a single day.'*

High rate of spread is not necessarily equal to fire growth of 10,000 or more acres. You can have a fire with a rapid rate of spread but the entire fire area is less than 10,000 acres. Conversely, you could have a large fire with a long active fire front with a rate of spread that is not all that rapid and still meet the 10,000 threshold. If days of rapid relative growth (a doubling or more) had a high overlap with the 10,000 acres I would then accept this variable. In the response to the reviews the authors state

Response Fig. 1 indicates that there is not a tremendous amount of overlap between 10,000-acre growth days and quadrupling days

I think this infers that this study is a reflection of growth characteristics of large fires in California not necessarily those fires with exceptionally high rate of spread. One could use a modelling approach to directly estimate the rate of spread from the satellite data – something akin to the Parks (2014) approach. The other option would be to use a doubling or tripling etc. of fire area. This option often happens in the early stages of an extreme fire but this can be when significant impacts occur (e.g., Camp Fire). Yes, I see that quadrupling fires are more difficult to predict but you could try tripling or doubling to see if you get better results. Also, the results for Fire Weather Indices were not great because I think 10,000 acres growth is not a best variable for rapid rates of spread (L1192-1199).

> Parks, S.A., 2014. Mapping day-of-burning with coarse-resolution satellite fire-detection data. Int. J. Wildland Fire 23, 215–223.

One could reframe the study to explore days where fire growth is 10,000 acres or more and how these large growth fire days could change in the future.

Response: The reviewer is correct in pointing out that the phrase 'high rate of spread' in the above quoted text is erroneous and we have corrected it.

Specifically, our predictand is 'extreme daily growth' (area per unit time) not 'rate of spread' (length per unit time). We focus on absolute extreme daily growth because it is more directly related to firefighting challenges and threats to life and property than the rate of spread (or relative growth).

Thus, our study was already framed around "days where fire growth is 10,000 acres or more and how these large growth fire days could change in the future". The phrase 'extreme daily growth' appeared 45 times in the manuscript and the phrase 'rate of spread' appeared only the single time mentioned above.

We have corrected this single erroneous use of the phrase 'high rate of spread'. The text in question now reads:

> *"Extreme wildfire behavior can be defined in several ways, but here we define it as an exceptionally high rate of growth in area-burned: 10,000 acres (about 2/3rds the size of Manhattan) or more in a single day."*

This correction brings our definition in line with the framing of the rest of the study and creates consistency throughout.

Reviewer 2

This study aims to quantify the influences of temperature on the extreme daily fire growth using machine learning approaches based on the observed daily burned area, and to make attribution assessment. Overall it is an interesting statistical study and the findings further support previous findings on extreme fire risk attribution to anthropogenic warming. The revision has addressed most of the reviewers' comments and improved the quanlity of the manuscript.

Response: We appreciate this sentiment.

The trend analysis on extreme fire growth, extreme weather, and the association between then would make the conclusion more compelling.

Response: Our dataset spans only 17 years which is not sufficiently long to draw robust conclusions on trends in either fire behavior or conditions conducive to extreme fire behavior. Thus we have intentionally not made trend analysis a focus of our study. Trend analysis can be conducted on datasets that span longer timeframes and we do have a second manuscript under review at another journal that does just that. However, we feel that work is sufficiently differentiated from this work that it warrants a stand-alone publication and trying to incorporate it here would be adding too much extraneous analysis to the supplement which is already very long.

A few other comments:

* Line 63 - "here we define it as an exceptionally high rate of spread, and in particular, growth of 10,000 acres". The data used for this study is daily area burned, which is different from fire rate of spread (velocity), although it is correlated for the most part.

Response: We thank the reviewer for pointing this out. It has been corrected. That text now reads:

> *"Extreme wildfire behavior can be defined in several ways, but here we define it as an exceptionally high rate of growth in area burned: 10,000 acres (about 2/3rds the size of Manhattan) or more in a single day."*

* I would suggest including the number of fire-days samples (all and extreme) in the main text, as the sample size is critical for the machine learning approaches.

Response: We have added this information to the main text in line 105.

* Line 73 - It is worth mentioning that fuels including both fuel types, structure, and amount were held constant in addition to fuel moisture.

Response: We have added this information to the main text in line 81.

* For the extreme fire risk machine learning modeling with random forest, how the probability of extreme fire was calculated for the response variable?

Response: In the revised manuscript, we describe this in the Supplementary Information starting on line 349:

> 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).

Reviewer 3

Recommendation: Accept subject to minor revisions.

The authors have satisfactorily addressed most of my comments, including by revising their title, and by providing more analyses in the supplementary information.

Response: We appreciate the approval.

Regarding the fact that absolute humidity is kept constant in the experiments, I still consider this a major caveat (as also recognized by the authors in their rebuttal: "the reviewer is absolutely correct to point out that holding absolute humidity, as well as every other variable constant is the major caveat in our study").

I am still not fully convinced that this caveat is sufficiently addressed in the revised manuscript:
- The authors write "We have also taken care to emphasize this caveat prominently throughout the revised manuscript (e.g., lines 76-141, lines 217-220, lines 340-352 [the concluding paragraph], and lines 782-834".

However, in the concluding paragraph (which is on lines 263-275, and not 340-352) the authors do not specifically refer to specific humidity.

In particular, humidity is not mentioned in the following text (lines 271-275):

"Despite this, we believe that our calculations probably result in conservative estimates of changes in risk because the balance of evidence suggests that it is more likely than not that important variables that we hold constant (precipitation36, wind11, tree mortality37, fire season length15,38,39, and lifetimes of fires2) are changing in a way that would exacerbate the enhanced risk caused by warming rather than ameliorate it.".

However, the authors show in their rebuttal (Fig. 3 of the rebuttal) that absolute humidity has decreased in the region, which would also suggest that this factor would contribute to increasing fire risk. At the minimum, absolute humidity should be thus explicitly mentioned in the highlighted sentence. But I

would also recommend that the authors add more text on the potential role of humidity in the long-term trends elsewhere in the manuscript.

Response: We have taken both of these recommendations. In the revised manuscript, absolute humidity is now mentioned in the sentence in question (line 182) and there is now more text on the potential role of long term changes in absolute humidity in the supplement on lines 319-331.

Reviewer Reports on the Second Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

Once again, I appreciate all the work the authors have done on this study. I have just a few suggestions.

There are references to extreme wildfire behavior including the title. I would suggest - extreme daily fire growth (or area burned). Fire behavior has a number of characteristics including intensity and rate of spread and I would argue that on some of the larger fires that fire behavior does not have to be extreme in order to reach 10,000 acres in a day. However, if the authors feel strongly about keeping extreme fire behavior (I do note they have defined extreme fire behavior in the paper as days with 10,000 or more acres burned) then some acknowledgement that on larger fires with long active perimeters the rate of spread/fire intensity could be lower than smaller fires to achieve 10,000 acres (or something to that effect).

Lastly, area burned is influenced by many variables and I think that fire management effort is one that deserves some mention. Fire management effectiveness varies due to a number of factors, but fire load/resource availability is a key factor that could confound your results.

Author Rebuttals to Second Revision:

Response to Reviewer Comments RE: ms 2022-07-11535B

Key

- Black font = reviewer/editor comment
- Green font = response to reviewer comment

Reviewer 1

Referees' comments:

Referee #1 (Remarks to the Author):

Once again, I appreciate all the work the authors have done on this study. I have just a few suggestions.

There are references to extreme wildfire behavior including the title. I would suggest - extreme daily fire growth (or area burned). Fire behavior has a number of characteristics including intensity and rate of spread and I would argue that on some of the larger fires that fire behavior does not have to be extreme in order to reach 10,000 acres in a day. However, if the authors feel strongly about keeping extreme fire behavior (I do note they have defined extreme fire behavior in the paper as days with 10,000 or more acres burned) then some acknowledgement that on larger fires with long active perimeters the rate of spread/fire intensity could be lower than smaller fires to achieve 10,000 acres (or something to that effect).

Response: We have taken the reviewer's (and editor's) advice and changed the title of the paper to:

"Climate Warming Increases Risk of extreme **daily fire growth** in California"

In accordance with this change, we have altered a small amount of text on line 77-79 to convey that we focus on *one aspect* of extreme wildfire behavior. This in contrast to the previous version where we *defined* extreme wildfire behavior *as* extreme daily growth.

Lastly, area burned is influenced by many variables and I think that fire management effort is one that deserves some mention. Fire management effectiveness varies due to a number of factors, but fire load/resource availability is a key factor that could confound your results.

Response: Forest management and fuel loads are stated as a key caveat on lines 133-136 and in light of the reviewer's advice we have added further mention of this in the concluding paragraph on line 197-200.