Response to Reviews

We thank the editor and the reviewers for their constructive comments to improve the manuscript. Their comments are reproduced below with our responses in blue. The corresponding changes in the manuscript are highlighted in blue.

Editor

While the manuscript has been improved, there are still a few major concerns need to be addressed. Please revise the manuscript according to the two additional points raised by referee #3 (attached). Also, it seems that one the major concerns about the necessity to use 58 variables (since many of which are highly correlated) from previous round has not been fully addressed. The authors may consider to use some statistical ways to test for multicollinearity, e.g., computing the variance inflation factor (or VIF), rather than choosing 0.7 as a threshold for correlation.

The issue of multicollinearity should indeed be considered more quantitatively. We stated in the original manuscript that the random forest as a machine learning tool is less unaffected by the issue of multicollinearity than traditional regression methods because the random forest randomly selects predictors used for each tree, in which the probability of sampling strongly correlated variables in a particular tree is largely avoided (Siroky, 2009). To prove this for our model, we calculate VIF for our random forest model by a bootstrapping of seven predictors (the number of predictors used in each tree) for 5000 times. We randomly select seven predictors out of all 58 potential predictors and compute the VIFs, and we repeat this sampling 5000 times for a VIF distribution. Each sampling yields seven VIFs values, and hence for 5000 sampling we obtain 35000 VIFs which forms a distribution. Figure R1 shows the distribution of VIFs for all the selected predictors. The distribution has a median of 1.67 for the winter-spring and a median of 1.62 for the summer fire season. The distribution has about 96% of the VIF values smaller than 10 for both fire seasons, demonstrating the minimized multicollinearity in tree models. We thus contend that all 58 potential predictors should be kept as model inputs and we should let the random forest algorithm choose the best predictors for itself. We have included the above discussion into the manuscript (line 503-511). Figure R1 is added in the supplementary.

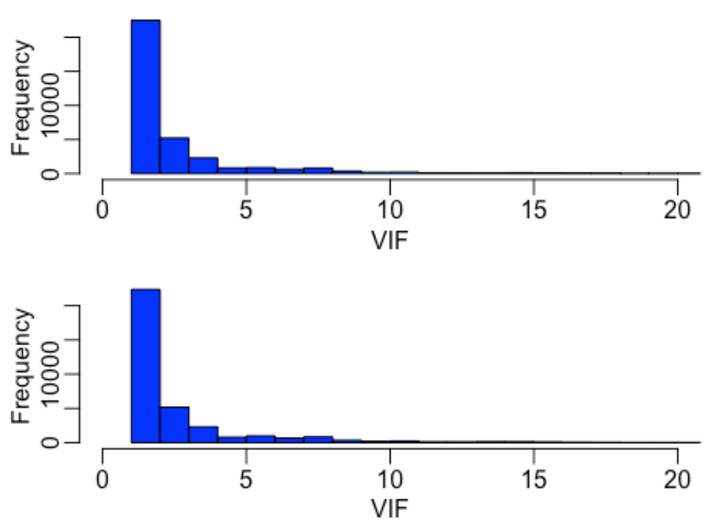

Figure R1. Distributions of VIF calculated based on randomly selected seven variables of 5000 times sampling for winter-spring (top) and summer fire season (bottom) (This figure is now Fig. S12. in the revised manuscript)

In addition, the South Central US has been chosen as a study area where the risk of wildfires has been predicted to be the highest in 2031-2050. But the proposed model seems to fail to predict BA during the years with abnormal fire activity especially during spring fire season (figure 5). If it is because "random forest or quantile regression forest cannot predict burned area greater than it observes before", how would that influence the performance of the proposed model for future predictions assuming that fire activity will increase in the next several decades. Please comment on it and discuss it in the revised manuscript.
Our model is able to predict future burned area for the following two reasons. First, the predicted burned area across the whole domain for the future scenario can be larger than it has observed before. The limitation that the maximum observed burned area cannot be exceeded is applicable *only at the grid level* and this upper limit is taken from all available grids of the whole training period, which can be referred to as the global upper limit per grid. The global upper limit is 514 km^2 per grid for the winter-spring fire season, and 238 km^2 per grid for the summer fire season. Under the effect of climate change, the total burned area summed across the domain can greatly exceed the present-day total burned area. Figure R2 shows an example for a randomly selected grid box. The model can predict the largest burned area on Feb 2008 and this is consistent with the observed burned area. This demonstrates that any single grid can predict burned area larger than the grid maximum by learning from other grids, and that therefore a much larger total burned area for the domain can be predicted by our model under future climate change.

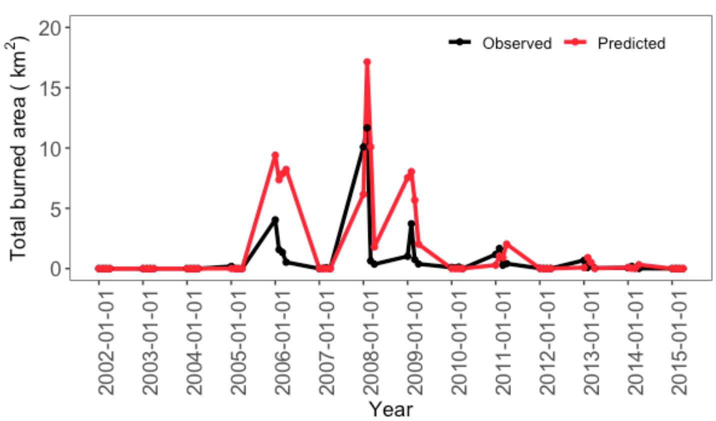

Figure. R2. Timeseries of observed (black line) and predicted total burned area (red line) for the selected grid (Lon: -98.75, Lat: 29.25) for the winter-spring fire season. (This figure is now Fig. S16. in the revised manuscript)

Second, the global upper limit is a sufficiently large value and thus the burned area per grid in the future would hardly exceed the global upper limit per grid. To further demonstrate the global upper limit per grid would be rarely exceeded, we show in Figure R3 the distribution of gridded burned area for year 2011, an extremely severe fire year for the study domain, in comparison to the distribution of all other years for 2002-2015. It can be seen that the majority of the burned areas for the extreme year are still within the range of the observed burned area in 2002-2015. Only two grids with burned areas exceed the global upper limit from 2002-2015 (excluding 2011). The total burned area of those exceedance grids only accounts for 20% of total burned area for 2011, which is within the stated uncertainty range of our prediction model. The above discussions have been included in the manuscript (line 622-636).

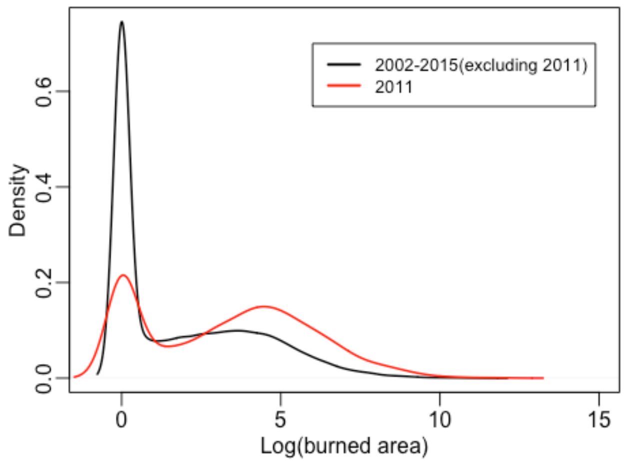

Figure. R3. Distribution of burned area of all the grids for the study period excluding 2011 (black line) and of the grids for the extreme year 2011 (red line) combined both seasons. (This figure is now Fig. S17. in the revised manuscript)

Reviewer #3

1. Per previous suggestion, the authors give the parameter information of the XGBoost. The different parameter configurations in the XGBoost and Random Forests are used for winter-spring and summer. Perhaps, we want to a uniform robust machine learning model that can achieve high accuracy both in the different seasons.
We understand the reviewer's perspective about a unified robust model configuration. However, we have used two different sets of predictor variables for the two fire seasons to characterize different important factors and processes, because the length and characteristics of the two pre-fire seasons are fundamentally different. In this regard, using a single set of parameter configuration for two different input predictor variables could not give us two fully optimized prediction models. Two parameter configurations that are tailor-made for two separate input predictor variables are needed to fully optimize the two prediction models. Using one unified parameter configuration for both seasons can technically be achieved easily, but it is not the best approach from the perspective of fine-tuning machine learning models. We have included the above explanations in the manuscript (line 241-244).

2. The authors analyze how RH anomaly and temperature anomaly affect the prediction. But the temperature anomaly is just ranked 10th in the summer season. I think the authors need to analyze how the top at least 3 variables affect the prediction so that we can learn something from the machine learning model, not just the accuracy. On the other hand, the author said "The physical reason behind their importance is that higher temperature coupled with lower relative humidity in

the summer can cause drier fuel and this condition is favorable for fires to start, spread, and burn more intensely". But, the machine learning importance cannot provide the influence of change of variable values. The authors should further prove that.

(1) The analyses of relationship between RH anomaly, temperature anomaly, and burned area demonstrate different controls of burned area in the two fire seasons. To better understand how the changes of top variables affect burned area, the partial dependence plots can be applied to the built model and show the marginal effect of a variable on the prediction performance (Friedman, 2001), as suggested by the reviewer. As we only included the results of partial dependence plots of the top two variables for the winter-spring fire season in the manuscript, the results of other top ranked variables and for the summer fire season are similar and more discussions are provided here. Figure R4 shows the partial dependence plots for the model and the top four variables (RH anomaly, SPEI_mean4m, apcp_avg, and temp_sd) for the winter-spring fire season. For RH anomaly, the fitted logarithmic burned area is getting larger if the RH anomaly is smaller than 2% (Figure R4a). The change likely indicates the sensitivity of burned area to the fire-season moisture. Similar pattern is also shown in the partial dependence plot of the mean SPEI of the preceding 4 months (Figure R4b). Larger fitted burned area is observed to be associated with the preceding SPEI smaller than zero, suggesting that burned area in this season is highly dependent on the pre-fire-season drought conditions, which is consistent with the findings of prior studies (Scott and Burgan., 2005; Riley et al., 2013; Turco et al., 2017). As for the average precipitation of 1979-2000, the fitted burned area increases as the average precipitation increases (Figure R4c). This implies that larger fires occur in the areas where the average precipitation was more in the past. For standard deviation of temperature during 1979-2000, the fitted burned area declines dramatically when the standard deviation of temperature is larger than 9K, suggesting burned area may be larger with relatively less variation of temperature in the winter-spring fire season (Figure R4d).

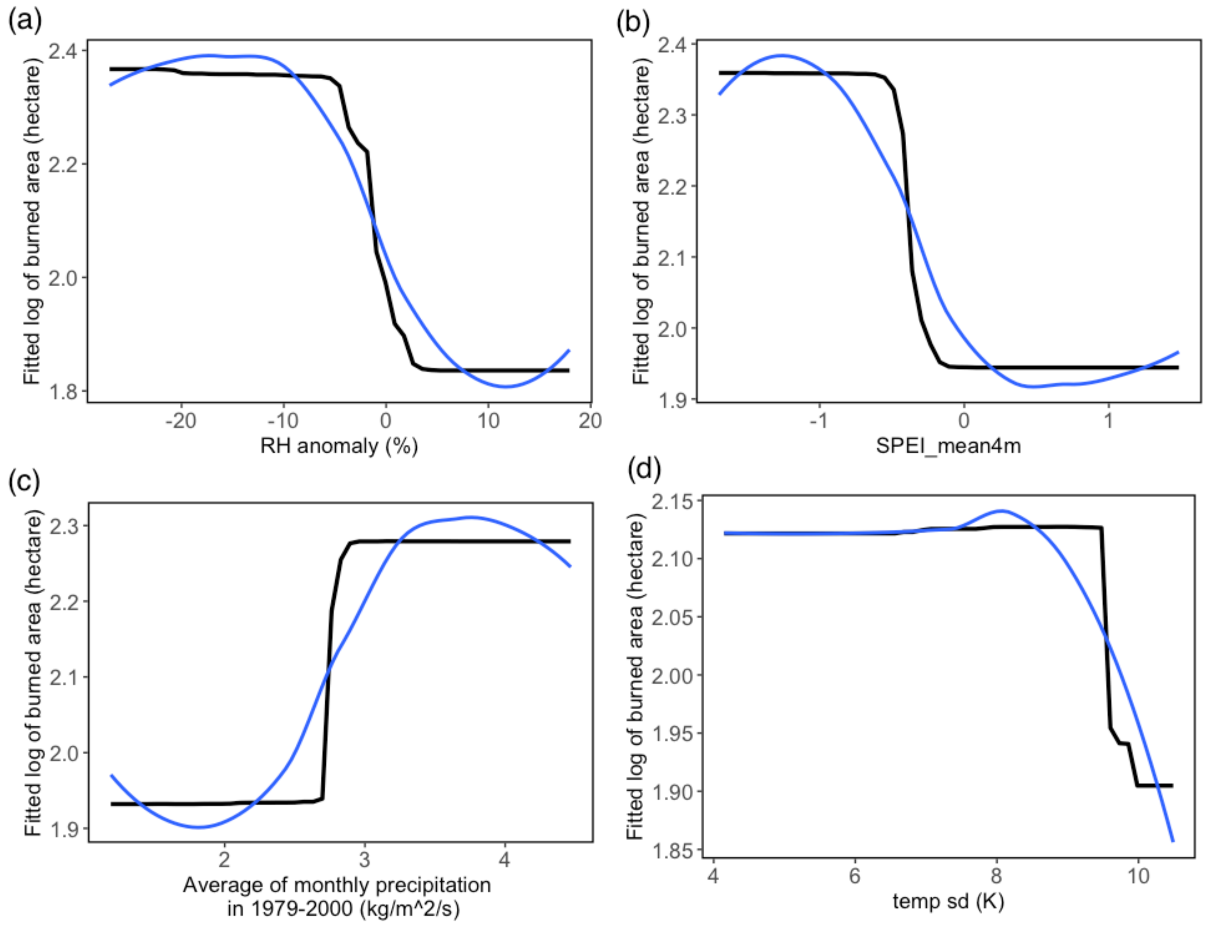

Figure. R4. Partial dependence plots for the burned area model and (a) RH anomaly, (b) the mean SPEI of the preceding 4 months, (c) the average precipitation of 1979-2000, (d) the standard deviation of temperature of 1979-2000 for the winter-spring fire season. The blue line is the LOESS smooth line. (This figure is now Fig. S9. in the revised manuscript)

For the summer fire season, the large burned area is associated with low values of RH anomaly, minimum RH anomaly, the mean SPEI of the preceding 2 months, and long-term (1979–2000) standard deviation of temperature (Figure R5). The fitted logarithmic burned area increases rapidly as the RH anomaly decreases toward zero and the increase in burned area reaches a maximum at RH anomaly of –14% (Fig. R5a). Compared to the partial dependence plot for RH anomaly, the fitted burned area increases more rapidly with decreasing minimum RH anomaly (Fig. R5c). At below zero, the sensitivity of log(burned area) to the minimum RH anomaly is 0.04 $\%^{-1}$ (Fig. R5c), while the corresponding sensitivity to RH anomaly is only 0.02 $\%^{-1}$ (Fig. R5a). The stronger sensitivity of burned area to minimum RH anomaly indicates the stronger effects of extremely low humidity conditions on fire growth as compared with the mean RH conditions. For the standard deviation of temperature during 1979-2000, larger burned area is observed with smaller standard deviation of temperature in the past. This suggests burned area would become larger for the grids with less variation of temperature (persistent high temperature) in the summer. As for the mean SPEI of the preceding 2 months, we see an increase of fitted burned area at zero, with the largest increase at –1.8, which supports the importance of fuel drying process in the summer fire season.

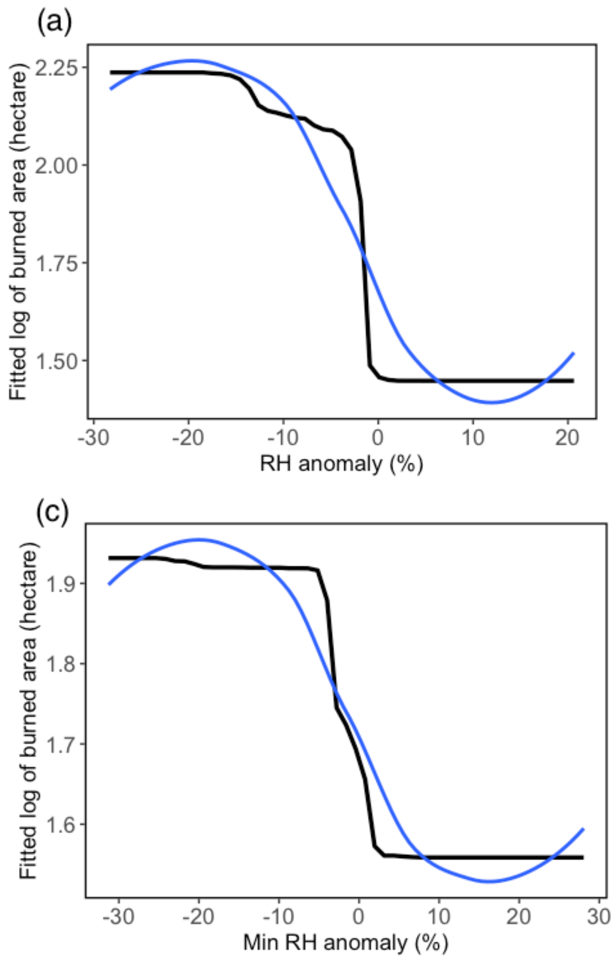

Figure. R5. Partial dependence plots for the burned area model and (a) RH anomaly, (b) long-term (1979-2000) standard deviation of temperature, (c) minimum RH anomaly, and (d) the mean SPEI of the preceding 2 months for the summer season. The blue line is the LOESS smooth line. (This figure is now Fig. S10. In the revised manuscript)

For both fire seasons, RH anomaly, mean SPEI of preceding months, and standard deviation of temperature for 1979-2000 are selected as the top 4 predictors, highlighting the importance of the common variables of the two seasons but with different thresholds and magnitudes in their effects on burned area. We have included the information and the above-mentioned examples in the revised manuscript (line 447-459 and 472-486).

(2) The statement of "This highlights the importance of the stronger combined effects of RH and temperature anomalies on burned area during summer, when higher temperature coupled with lower relative humidity can cause drier fuel and create favorable conditions for fires to start, spread, and burn more intensely" is mainly based on differences in correlation between RH anomaly and temp anomaly for the two fire seasons. Additionally, in the variable importance analyses, RH anomaly is selected for both seasons, while temperature anomaly is only shown for the summer fire season. To further prove the statement, here we plot out the relationship between RH anomaly, temperature anomaly, and burned area. We perform a regression for the RH anomaly (y) and temperature anomaly (x), and fires with different sizes labeled with different colors. The slope of

the line is the change in RH anomaly over the change in temperature anomaly, which represents the dependence of RH anomaly on temperature anomaly. The slopes are -3.7 and -0.89 for the summer and winter-spring fire season, respectively, showing that a strong dependence of RH anomaly on temperature anomaly in the summer (Figure R6). In addition, large burned area (75th percentile, black dots in Figure R6) mainly occur in the condition of low RH anomaly and high temperature anomaly (bottom-right corner), in particular for the summer fire season. The conclusion from this plot supports our statement that "higher temperature coupled with lower relative humidity can cause drier fuel and create favorable conditions for fires to start, spread, and burn more intensely". We have revised the corresponding paragraphs and included the above discussions and Fig R6 into the supplementary (line 431-436).

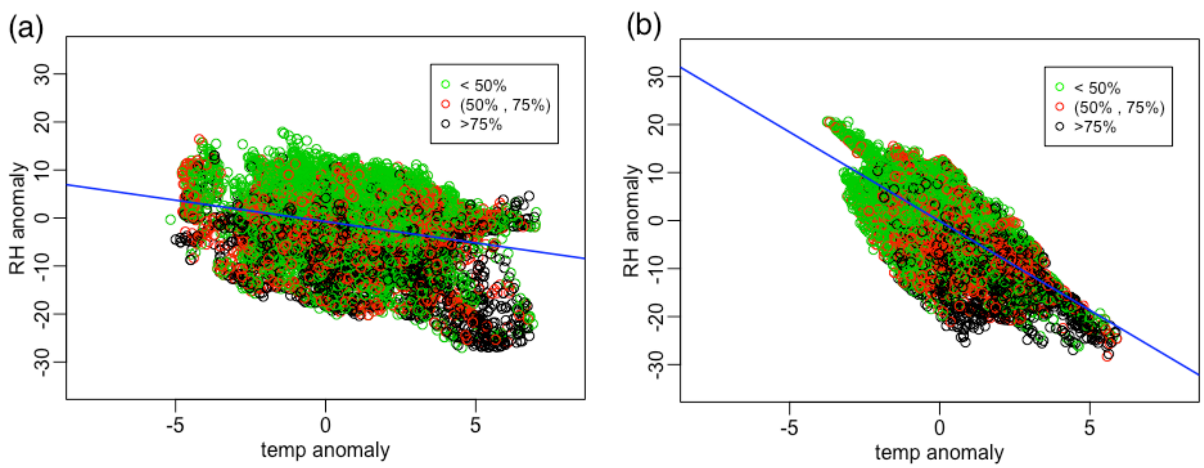

Figure R6. Scatter plot of RH anomaly versus temperature anomaly for (a) winter-spring and (b) summer fire season. The blue line is the fitted regression line. The color represents different sizes of fire burned area (Green: smaller than 50th percentile; Red: larger than 50th percentile but smaller than 75th percentile; Black: larger than 75th percentile). (This figure is now Fig. S8. in the revised manuscript)

References

Friedman, J. H.: Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, 29(5), 1189–1232, 2001.

Riley, K. L., Abatzoglou, J. T., Grenfell, I. C., Klene, A. E. and Heinsch, F. A.: The relationship of large fire occurrence with drought and fire danger indices in the western USA, 1984–2008: the role of temporal scale, International Journal of Wildland Fire, 22(7), 894, doi:10.1071/WF12149, 2013.

Scott, J. H. and Burgan, R. E.: Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model, Gen. Tech. Rep. RMRS-GTR-153. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 72 p., 153, doi:10.2737/RMRS-GTR-153, 2005.

Siroky, D. S.: Navigating Random Forests and related advances in algorithmic modeling, Statist. Surv., 3, 147–163, doi:10.1214/07-SS033, 2009.

Turco, M., Hardenberg, J. von, AghaKouchak, A., Llasat, M. C., Provenzale, A. and Trigo, R. M.: On the key role of droughts in the dynamics of summer fires in Mediterranean Europe, Scientific Reports, 7(81), doi:10.1038/s41598-017-00116-9, 2017.